KIT

Karlsruher Institut für Technologie

Differential cross section measurements in the $H \to \tau\tau$ decay channel with CMS data of proton-proton collisions at the Large Hadron Collider at CERN

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

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

DISSERTATION

von

M.Sc. Sebastian Wozniewski

Tag der mündlichen Prüfung: 4. Dezember 2020

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

KIT

Karlsruhe Institute of Technology

Differential cross section measurements in the $H \to \tau\tau$ decay channel with CMS data of proton-proton collisions at the Large Hadron Collider at CERN

For obtaining the academic degree

DOKTOR DER NATURWISSENSCHAFTEN

at the Department of Physics
of the Karlsruhe Institute of Technology (KIT)

DOCTORAL THESIS

of

M.Sc. Sebastian Wozniewski

Date of the oral exam: 4th December 2020

Reviewer: Priv.-Doz. Dr. Roger Wolf
Second Reviewer: Prof. Dr. Günter Quast

Contents

Introduction

It was a big step for the understanding of the fundamental laws of nature when a particle compatible with the expected Higgs boson was discovered at the Large Hadron Collider (LHC) in 2012. Before that, it was not clear how the heavy electroweak gauge bosons obtain their mass. The Brout-Englert-Higgs-Mechanism had provided a plausible explanation but was not experimentally verified. Finally, the discovery of the Higgs boson, which had been predicted by the mechanism, provided strong evidence.

Since that time, researchers have been analysing more and more data taken by the two involved experiments at the LHC, ATLAS and the Compact Muon Solenoid (CMS), in order to measure the properties of the Higgs boson and the related physics sector. Many institutes affiliated with the experiments and the European Organization for Nuclear Research (CERN), which the LHC belongs to, are involved in this research. One of these is the Institute for Experimental Particle Physics (ETP) in Karlsruhe, where the studies of this thesis have been carried out.

The Higgs field does not only serve to explain the masses of gauge bosons. Via additional couplings it is also used to introduce mass to fermions. These couplings, which are predicted to be proportional to the mass of the fermion, can be measured in order to verify this assumption. Therefore, the decay channel of the Higgs boson to a pair of τ leptons is of particular interest. The τ lepton is the heaviest known lepton and since its final states provide a clearer signature at a hadron collider than a decay into hadrons, this decay channel is the most sensitive for measuring the coupling of the Higgs boson to fermions. Nevertheless, the complexity of analyzing this channel arising from the multiple decay channels of the τ-pair that are taken into account in order to reach the full sensitivity must not be underestimated.

The discovery of the Higgs boson in the di-τ decay channel was announced with a significance of five standard deviations [1] at an early stage of LHC Run-2 after the data-taking in 2016. This confirmed the coupling of the Higgs boson to fermions aside from more indirect observations related to the production mode. The $H \rightarrow \tau\tau$ channel is being analysed in various ways, measuring CP properties of the production or the decay, searching for further Higgs bosons motivated by supersymmetry and potentially enhanced couplings to down-type fermions, and measuring the cross section more precisely and more differentially.

The topic of this thesis is the differential measurement of cross sections and their ratios to the prediction by theory in the $H \to \tau\tau$ channel. Given the dataset of the complete Run-2 of the LHC, which is about four times larger than the dataset the discovery of the Higgs boson in this channel was achieved with, and with improved analysis techniques, the measurement of the signal in about twelve partitions of the kinematic phase space is feasible. This is done within the Simplified Template Cross Sections (STXS) scheme, which is shared among analyses of the various decay channels of the observed Higgs boson and the two experiments ATLAS and CMS.

Aside from the extraction of the signal from the large background, which remains a necessity, the separation of the single signal components becomes a challenging task where much sensitivity can be gained with the right measures. In the analysis presented in this thesis, multi-classification based on neural networks is used to classify and sort the observed proton-proton collisions in order to obtain optimal sensitivity for the differential measurement. It makes it possible to compare many hypotheses about the type of a given collision event at the same time. In the context of multiple signal components and several background processes, this is a very powerful tool.

Before getting into the measurement, an overview of the theory of the Brout-Englert-Higgs-Mechanism and the experimental investigation of the production and decay of the Higgs boson at the LHC is provided in chapter 2. Subsequently, the reconstruction of events at the CMS experiment is explained in chapter 3. Chapter 4 is related to the selection of events for this analysis and how these events are modelled in the statistical framework. The machine learning based classification of the selected events and creation of quantities that are used to discriminate between the involved signal and background processes are discussed in chapter 5. Finally, the results of the measurement are presented in chapter 6 after a brief introduction to the statistical methods used to retrieve them.

Higgs boson under investigation

The presented analysis is part of a broad field of physics investigations related to the Brout-Englert-Higgs-Mechanism. This chapter provides an introduction to this research, starting with a summary of the underlying theory, subsequently giving an overview of related coupling measurements conducted at the LHC and discussing a scheme for differential measurements that is shared among the various analyses and an important ingredient of this thesis.

2.1 Brout-Englert-Higgs-Mechanism

The Standard Model of Particle Physics (SM) describes three of the four fundamental interactions between all known fundamental particles: the strong, weak and electromagnetic interactions. It is formulated as a renormalisable quantum field theory with inner symmetries corresponding to the three interactions:

$$\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1) \tag{2.1}$$

The interactions are derived by imposing local gauge invariance under these symmetries. The strong interaction and its charges correspond to the SU(3) symmetry group. The electromagnetic and weak interactions are described as a unified theory with an underlying $\mathrm{SU}(2) \times \mathrm{U}(1)$ symmetry, where the fields of the photon and the weak bosons correspond to superpositions of the gauge fields. The latter was introduced by S. Glashow [2], A. Salam [3] and S. Weinberg [4], who were awarded the Nobel prize for this in 1979. The Brout-Englert-Higgs-Mechanism is an essential part of this theory. It was formulated in 1964, independently by P. Higgs [5][6] and by F. Englert and R. Brout [7]. This was awarded with the Nobel price in 2013 after the discovery of the Higgs boson as well.

The theory of electroweak interactions without this mechanism encounters difficulties when including masses of particles. The $W^\pm$ and Z bosons are known to carry significant mass of $80.4\,\mathrm{GeV}$ and $91.2\,\mathrm{GeV}$ respectively [8], but simple mass terms in the Lagrangian density of the SM like

$$m_W^2 W_\mu^\dagger W^\mu + \frac{1}{2} m_Z^2 Z_\mu Z^\mu \tag{2.2}$$

would break the symmetry and renormalisability of the theory [9][10]. Similarly, mass terms for fermions are problematic due to the different treatment of right- and left-handed

components that describes the parity violating behaviour of the weak interaction. The expression

$$-\bar{\psi}m_f\psi = -m_f(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \tag{2.3}$$

is not invariant under global gauge transformations because ψ_R and ψ_L have different isospin representations and different hypercharges [11].

The Brout-Englert-Higgs-Mechanism avoids such difficulties via the postulate that the $\mathrm{SU(2)} \times \mathrm{U(1)}$ symmetry is spontaneously broken, such that the Lagrangian density is still invariant under the corresponding transformations but the ground state of the vacuum is not. It was initially installed to solve the problem of describing the masses of the weak gauge bosons, but can also be used to take the mass of fermions into account. A new field Φ is introduced as an isospin doublet of two complex scalar fields. It is assigned the hypercharge $Y = 1/2$ such that according to the Gell-Mann-Nishijima relation[1] one component of the field carries electrical charge $Q = 1$ and the other one is electrically neutral. It can therefore be written as

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{2.4}$$

It is introduced to the Lagrangian density together with a potential that leads to the spontaneous symmetry breaking and can be understood as an expansion of a more complex potential:

$$\mathcal{L}_{\mathrm{HEB}} = [D_\mu\Phi]^\dagger[D^\mu\Phi] + \mu^2\Phi^\dagger\Phi - \lambda[\Phi^\dagger\Phi]^2 \tag{2.5}$$

with two parameters $\lambda > 0$, μ and the covariant derivatives of the $\mathrm{SU(2)} \times \mathrm{U(1)}$ symmetry

$$D_\mu = \partial_\mu + ig\frac{\tau^i}{2}W_\mu^i + ig'YB_\mu, \tag{2.6}$$

where τ_i are the generators of the $\mathrm{SU(2)}$ group represented by the Pauli matrices and g, g' the coupling constants. The potential apparently preserves the $\mathrm{SU(2)} \times \mathrm{U(1)}$ symmetry. With $\mu^2 > 0$, the minimum of the potential is located at $|\Phi| = \sqrt{\mu^2/2\lambda}$ leading to a non-zero vacuum ground state of this field. Note that the ground state is degenerate due to the symmetry of the potential. This is illustrated in figure 2.1 for a single component complex field.

In the unitary gauge, the ground state that is realised in nature takes the form

$$\Phi_0 = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}, \ v := \sqrt{\frac{\mu^2}{\lambda}}. \tag{2.7}$$

Expressing the Lagrangian density in terms of real scalar fields relative to the vacuum ground state reveals the benefit of this approach. Note that the degrees of freedom represented by the fields η_i in the following can be eliminated via gauge transformations and

---

[1]The convention $Q = I_3 + Y$ with half-integer hypercharge for leptons is used here.
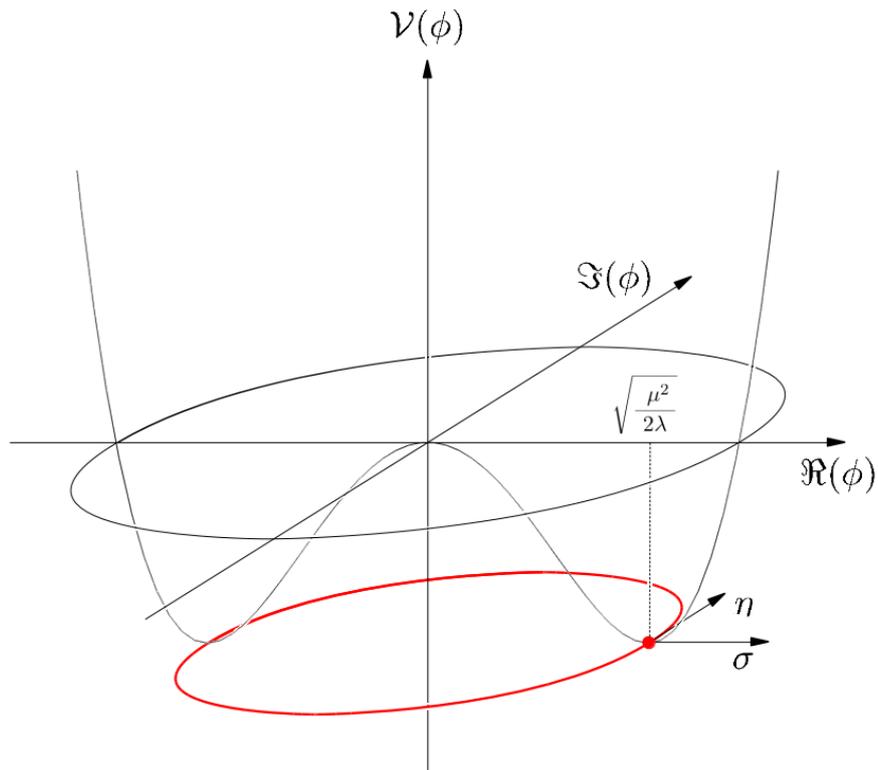
Figure 2.1: Spontaneous symmetry breaking for a U(1) symmetric Higgs potential (from [12]): The field ϕ has degenerate ground states shown as a red circle. The red dot marks the ground state the system has spontaneously fallen into. The real fields σ and η are defined relatively to the ground state.

are not considered physical. Therefore, terms that include η_i are going to be summarised as $\mathcal{L}_\eta$.

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \eta_1 + i\eta_2 \\ v + \sigma + i\eta_3 \end{pmatrix} \tag{2.8}$$

$$\begin{aligned} \implies \mathcal{L}_{\text{HEB}} =& \frac{1}{2} \left[D_\mu \begin{pmatrix} 0 \\ v+\sigma \end{pmatrix} \right]^\dagger \left[D^\mu \begin{pmatrix} 0 \\ v+\sigma \end{pmatrix} \right] - \mu^2\sigma^2 - \lambda v\sigma^3 - \frac{\lambda}{4}\sigma^4 \\ &+ \mathcal{L}_\eta + \text{const.} \end{aligned}$$

$$\begin{aligned} =& \frac{(v+\sigma)^2}{2} \begin{pmatrix} 0 & 1 \end{pmatrix} \left(i\frac{g}{2}\tau^a W_\mu^a + i\frac{g'}{2}B_\mu \right) \left(i\frac{g}{2}\tau^b W^{b\mu} + i\frac{g'}{2}B^\mu \right) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &+ \frac{1}{2}(\partial_\mu\sigma)^2 - \mu^2\sigma^2 - \lambda v\sigma^3 - \frac{\lambda}{4}\sigma^4 \\ &+ \mathcal{L}_\eta + \text{const.} \end{aligned} \tag{2.9}$$

Using that $\{\tau^a, \tau^b\} = 2\delta_{ab}$ and that τ^1 and τ^2 do not have diagonal elements it follows

$$\begin{aligned} \mathcal{L}_{\text{HEB}} =& \frac{(v+\sigma)^2}{8} \left[\sum_{a=1}^{3} ((gW_\mu^a)^2) - 2gg'W_\mu^3 B^\mu + (g'B_\mu)^2 \right] \\ &+ \frac{1}{2}(\partial_\mu\sigma)^2 - \mu^2\sigma^2 - \lambda v\sigma^3 - \frac{\lambda}{4}\sigma^4 \\ &+ \mathcal{L}_\eta + \text{const.} \end{aligned} \tag{2.10}$$

$$\begin{aligned} \iff \mathcal{L}_{\text{HEB}} =& \frac{(v+\sigma)^2}{8} \left[g(W_\mu^1)^2 + g(W_\mu^2)^2 + (gW_\mu^3 - g'B_\mu)^2 \right] \\ &+ \frac{1}{2}(\partial_\mu\sigma)^2 - \mu^2\sigma^2 - \lambda v\sigma^3 - \frac{\lambda}{4}\sigma^4 \\ &+ \mathcal{L}_\eta + \text{const.} \end{aligned} \tag{2.11}$$

Applying the definitions of the physical gauge fields [11]

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \\ Z_\mu^0 &= \frac{1}{\sqrt{g^2 + g'^2}}(gW_\mu^3 - g'B_\mu) \\ A_\mu &= \frac{1}{\sqrt{g^2 + g'^2}}(g'W_\mu^3 + gB_\mu) \end{aligned} \tag{2.12}$$

this results in

$$\begin{aligned} \mathcal{L}_{\text{HEB}} =& m_W^2 W_\mu^- W^{+\mu} + \frac{1}{2}m_Z^2 Z_\mu^0 Z^{0\mu} \\ &+ \frac{1}{2}(\partial_\mu\sigma)^2 - \mu^2\sigma^2 - \lambda v\sigma^3 - \frac{\lambda}{4}\sigma^4 \\ &+ \frac{2m_W^2}{v}\sigma W_\mu^- W^{+\mu} + \frac{m_W^2}{v^2}\sigma^2 W_\mu^- W^{+\mu} + \frac{m_Z^2}{v}\sigma Z_\mu^0 Z^{0\mu} + \frac{m_Z^2}{2v^2}\sigma^2 Z_\mu^0 Z^{0\mu} \\ &+ \mathcal{L}_\eta + \text{const.} \end{aligned} \tag{2.13}$$

where we set $m_W = gv/2$ and $m_Z = \sqrt{g^2 + g'^2}v/2$. Checking the first line of equation 2.13, there are indeed mass terms such that m_W and m_Z are the masses of the massive gauge bosons. They are proportional to the vacuum expectation value of Φ. There is no mass term for the electromagnetic vector potential A_μ, which is good since photons are known to be massless. It can be retraced in the above calculation that this is due to the choice for the hypercharge of Φ or equivalently the choice that the electrically neutral component of Φ was chosen to yield the non-zero value in the ground state. The second line contains a kinematic term for the field σ, furthermore a mass term and self-coupling terms. This gave rise to the postulate of the massive Higgs boson and in particular yields the chance to test this theory experimentally. The third line contains coupling terms between the Higgs boson and the massive electroweak gauge bosons.

Up to this point, the Brout-Englert-Higgs-Mechanism does not have any impact on the fermion fields, in particular it does not lead to mass terms for fermions. These are introduced independently as Yukawa couplings [11]:

$$\mathcal{L}_{\text{Yukawa}} = -\lambda_d \bar{\Psi}_L \Phi d_R + \text{h.c.}, \ \ \Psi_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$$

$$\Longrightarrow \mathcal{L}_{\text{Yukawa}} = -\frac{\lambda_d}{\sqrt{2}} \bar{d}_L(v + \sigma)d_R + \mathcal{L}'_\eta + \text{h.c.} \tag{2.14}$$

$$= -m_d \bar{d}_L d_R - \frac{m_d}{v} \bar{d}_L \sigma d_R + \mathcal{L}'_\eta + \text{h.c.}$$

Here again, the SU(2) components were contracted and the coupling re-parametrised with $m_d = \lambda_d v/\sqrt{2}$. The result yields a mass term for the isospin down-type fermion field and a coupling to the Higgs boson proportional to the fermion mass and anti-proportional to the vacuum expectation value of Φ. Such couplings can be introduced for each particle family while the mass enters as a free parameter. Since this only covers isospin down-type fermions, which is fine for leptons with massless neutrinos, but not for quarks, terms for up-type quarks require a separate treatment. With the given ground state, Φ cannot be used to create mass terms for up-type fermions. Instead, those are introduced via the charge conjugate of Φ [11][13]:

$$\mathcal{L}_{\text{Yukawa},u} = -\lambda_u \bar{\Psi}_L i\tau_2 \Phi^* u_R + \text{h.c.}$$

$$= -m_u \bar{u}_L u_R - \frac{m_u}{v} \bar{u}_L \sigma u_R + \mathcal{L}''_\eta + \text{h.c.} \tag{2.15}$$

With the shown steps, all necessary mass terms for the known massive particles in the SM are obtained without breaking gauge invariance or renormalisability of the theory. In addition, a new massive particle shows up, the Higgs boson. And furthermore, couplings between known particles and the Higgs boson are predicted, where the salient feature is the dependence of these couplings on the particle masses. Therefore, an experimental verification of this theory goes beyond the pure discovery of the Higgs boson and also includes measurements of the couplings to the different particle types and families.
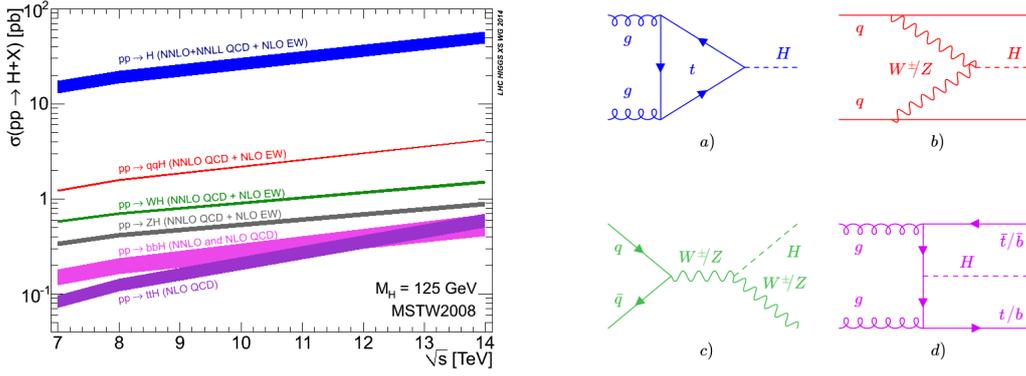
Figure 2.2: Production modes of the Higgs boson at the LHC. Left: Expected production cross sections depending on the centre of mass energy $\sqrt{s}$ (from [14]). Right: Related leading order Feynman diagrams, a) gluon fusion (pp → H), b) vector boson fusion (pp → qqH), c) Higgs strahlung (pp → WH / ZH), d) heavy quark associated production (pp → ttH / bbH).

2.2 Experimental analysis at the LHC

To be precise, it should be formulated that in 2012 a Higgs-boson-like particle was discovered. While it was clear that a new sector of particle physics was found, it was still open whether the discovered particle fulfils all properties of the expected Higgs boson. This question has been addressed over the past years with increasing luminosity and experimental sensitivity to more and more production and decay processes. On one hand, it needs to be confirmed that all couplings are as expected. In particular due to the non-uniform way of introducing masses for gauge bosons and fermions, this is a well motivated question. On the other hand, the observed particle might not be the only Higgs boson and just part of a larger "Higgs sector". Apart from direct searches for additional Higgs bosons, e.g. in the context of the Minimal Supersymmetric Extension to the Standard Model (MSSM), such an extended sector would also influence properties like branching fractions of the observed Higgs boson, which is another motivation for detailed analyses of as many related processes as possible.

This section gives an overview of the production and decay modes of the Higgs boson as well as related analyses at the LHC. It restricts to analyses of the coupling between the observed Higgs boson and gauge bosons or fermions, which this thesis is part of. However, it should be noted that there are also other analyses conducted to explore the Higgs sector, dealing with the self-coupling or CP properties of the known Higgs boson or direct searches for additional Higgs bosons.

2.2.1 Production in proton proton collisions

At the LHC, Higgs bosons are produced in proton proton collisions, where four leading production processes are distinguished: gluon fusion (ggH), vector boson fusion (VBF),

Higgs strahlung (VH)[2] and heavy quark associated production. The cross sections predicted by theory and the leading order Feynman diagrams are illustrated in figure 2.2. The ggH process has the largest cross section, driven by strong couplings, the gluon PDF and the heavy top quark which dominates the quark loop due to its large mass and hence coupling to the Higgs boson. But its experimental signature is similar to the one of the production and decay of a Z boson, which constitutes an important background in the measurements. Therefore and depending on the actual analysis, the other production processes can still be more relevant. VBF is the production process with the second largest cross section. It involves couplings to the heavy electroweak gauge bosons. The partons which emit the heavy gauge bosons hadronise and initiate two particle jets, which can be used in some analyses to select collisions with this process more efficiently. VH has the third largest cross section. In terms of vertices, it resembles VBF, but the annihilation of the initial quarks leads to a smaller multiplicity and cross section. The produced vector boson decays leptonically or hadronically. In the latter case, two jets are observed like for VBF and at higher order, these two processes are not distinguishable. Therefore, they are often treated as one process, called qqH. This is also done in the further parts of this thesis. The remaining VH with a leptonically decaying vector boson is referred to as VH$_{lep}$. Next largest cross sections are predicted for the Higgs boson production in association with heavy quarks of the third generation, i.e. top or bottom quarks. The involved vertices are the same as for ggH, but there is no closed quark loop leading to the additional heavy flavour jets in the final state. Compared to ggH, the cross section of the bottom quark associated production (bbH) is two orders of magnitude smaller mainly due to the smaller mass of the bottom quark compared to the top quark resulting in a smaller coupling to the Higgs boson. This does not impact the top quark associated production (ttH), but the two top quarks in the final state increase the required energy. At the given centre of mass energy of the proton proton collisions of 13 TeV this results in a suppressed cross section almost as small as for bbH.

It is worth to emphasise that ggH and the heavy quark associated production involve the coupling of the Higgs boson to fermions, whereas VBF and VH involve the coupling to massive gauge bosons. Recalling the different realisations of these couplings in theory, detailed in section 2.1, this grouping is of particular interest and part of many measurements.

The sensitivity of analyses to these production modes strongly depends on the considered decay channel of the produced Higgs boson, which is detailed in the following section. At ATLAS and CMS, analyses are typically dedicated to one of these decay channels, like this thesis focusses on the di-τ-channel. Such analyses are part of the following summary. They focus on the production modes which they are sensitive to while minor production modes may be treated as background and potentially investigated further in separate analysis approaches. Note that the cross section of heavy quark associated production is too small to achieve reasonable results with some of the channel-specific analyses at the moment. It is still worth mentioning that a dedicated ttH analysis is conducted instead,

---

[2]The Higgs boson is emitted by a heavy gauge boson. In certain contexts it is useful to distinguish between the types of the gauge bosons (WH/ZH).
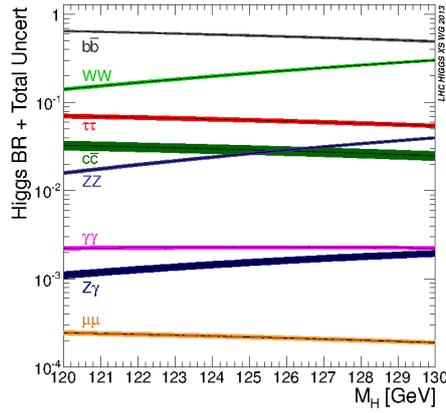
Figure 2.3: Expected branching ratios of the Higgs boson for the analysed final states. The decays to $\gamma\gamma$ or $Z\gamma$ are mediated via loops. (adapted from [14])

involving the three decay channels to τ leptons and heavy gauge bosons resulting in a multi-lepton final state. It allows to measure the signal strength at a precision of about $25\,\%$ [15][16].

2.2.2 Decay channels and analysis branches

The Higgs boson decays almost instantaneously after its production such that only the decay products can be detected. Within the experimental resolution they originate from the same vertex where the proton proton collision takes place. An overview of the experimentally accessible decay channels and the major related analyses is provided in the following. The related theoretical branching fractions of the Higgs boson are depicted in figure 2.3.

Bosonic decays

- $H \to \gamma\gamma$ [17][18]:
  The Higgs boson does not couple to photons because they are massless. However, a decay into two photons is mediated by a loop of predominantly W bosons and top quarks (large mass and electric charge). Therefore, it is not a purely bosonic decay in the sense of involved Higgs boson couplings. Despite the small branching fraction of about $0.23\,\%$ the measurement of this decay channel is one of the most sensitive ones due to the clear signature of two energetic photons forming a close resonance over a flat background in the spectrum of their invariant mass. The width of this resonance is driven by the detector resolution and is at the order of $1\,\text{GeV}$. Another benefit in this channel is that the signature of the Higgs boson decay does not overlap with additional products of the Higgs boson production modes such that the sensitivity to these does not suffer from any misidentification or selection issues. It is even sensitive enough to provide a standalone measurement of ttH production. The inclusive signal strength (production cross section times branching fraction

divided by the SM expectation) is measured at a precision of 10 % of the observed value by both ATLAS and CMS with Run-2 data. In both cases it is in agreement with the SM expectation. The same is true for the highly granular differential measurements where the signal is split into about 25 sub-processes.

- H $\rightarrow$ WW* $\rightarrow$ $e\nu_e\mu\nu_\mu$ [19][20][21]:
  The Higgs boson can decay via two W bosons into leptons and/or hadronic jets. For the experimental analysis, the final state with one electron and one muon[3] is chosen because the other decays are dominated by backgrounds from quark/gluon induced jets or decays of a Z boson into a pair of leptons of the same flavour. Thus, only about 3 % of the relatively high branching fraction of this channel is effectively usable [8][22]. With this selection, most of the background results from pair production of W$^+$W$^-$, t$\bar{\text{t}}$, $\tau^+\tau^-$, which can decay with a qualitatively equal signature in the detector. Requirements particularly involving the Missing Transverse Energy ($\not{E}_\text{T}$)[4] and the invariant mass as well as momenta of the produced leptons help to discriminate these backgrounds. Vetoing additional leptons helps to get a cleaner selection without ambiguity in the choice of the leptons. This increases the overall sensitivity but suppresses the VH$_{lep}$ process.
  The inclusive signal strength measurement with Run-2 data reaches an accuracy of 12 % of the observed value. Both inclusive and differential measurements are in agreement with the SM expectation.

- H $\rightarrow$ ZZ* $\rightarrow$ 4l [23][24]:
  Together with H $\rightarrow$ $\gamma\gamma$, this is the most sensitive channel. Also here, the branching fraction is small, but the advantage is that the final state can be fully reconstructed and the resolution of the lepton momenta is good enough to achieve a sharp resonance above the backgrounds. Pairs of oppositely charged electrons or muons are selected in order to reconstruct the Z bosons, resulting in three sub-channels, i.e. $e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$, $\mu^+\mu^-\mu^+\mu^-$. With additional kinematic discriminants it is even possible to include additional leptons from VH$_{lep}$ or ttH processes, which allows for a good separation between the production modes, together with jet selection criteria.
  The inclusive signal strength is measured with Run-2 data with a precision of 11 % of the observed value by both ATLAS and CMS respectively. Differential measurements are performed with up to 19 splits of the signal processes. All results are compatible with the SM prediction.

---

[3]These are not necessarily prompt.

[4]Neutrinos cannot be detected by the LHC experiments. However, the complete coverage of the detector in the transverse direction is used to calculate the undetected energy from the transverse momenta of the visible particles, which should sum up to zero if there were no neutrinos. Apart from detector effects and pileup, this provides a good estimate for the total transverse momentum of all neutrinos in an event. More details are given in section 3.2.4.

- *Zγ* [25]:
  This is one of the less sensitive channels due to the small branching fraction. The analysis selects events where the *Z* boson decays to electrons or muons because of the clear signature and because the *Z* boson mass can be completely and sharply reconstructed.
  It was recently measured by ATLAS with an inclusive signal strength of 2.0 ± 1.0 and a significance of two standard deviations.

Hadronic decays

- H → b$\bar{\text{b}}$ [26][27][28]:
  This is the channel with the largest branching fraction and important for testing the coupling of the Higgs boson to fermions. However, its sensitivity suffers from the hadronisation and quick decay of the bottom quark, resulting in a particle jet that is difficult to distinguish from other particle jets of different origin and that has a worse energy resolution compared to leptons or photons. There are algorithms to identify such bottom quark initiated jets, referred to as *b*-tagging. They exploit the jet substructure, in particular secondary vertices from the displaced bottom quark decays. However, their resolution is limited (e.g. 70 % efficiency for *b*-initiated jets with 0.3 % misidentification rate for light flavour jets [26]). Analysts make use of the leptons in the VH$_{lep}$ process to achieve a pure selection, which makes this channel particularly sensitive to this production mode despite the small cross section, even more than to ggH or VBF. This is in contrast to channels that are dealing with leptons in the final state, which may even veto additional leptons and lose sensitivity to VH$_{lep}$.
  Latest results by ATLAS with Run-2 data determine the VH$_{lep}$ process with a 16 % constraint and differentially in five bins with constraints between 30 % and 85 %, all compatible with the expectation according to the SM.

- H → c$\bar{\text{c}}$ [29][30]:
  This channel would be the next candidate to measure the coupling of the Higgs boson to quarks and it is of interest for accessing the second fermion generation. The analysis strategy is very similar to H → b$\bar{\text{b}}$. But the branching fraction is one order of magnitude smaller and even more problematic is the identification of the charm quark initiated jets. The impact parameters or secondary vertices are usually resolvable by the detectors but smaller than for bottom quark initiated jets, which makes charm quark initiated jets a category embraced by light flavour jets and bottom quark initiated jets in terms of jet flavour tagging. This makes it difficult to get a clean selection of signal events. The analysis of this channel was recently carried out as a search with exclusion limits of about 70 times the expected signal strength at 95 % CL.

Table 2.1: Branching fractions $\mathcal{B}$ of di-tau decays to electrons, muons and hadrons [22] (involved neutrinos are omitted in the notation)

decay channel	$\mathcal{B}\,(\%)$
$\tau\tau \to ee$	3.2
$\tau\tau \to e\mu$	6.2
$\tau\tau \to \mu\mu$	3.0
$\tau\tau \to e\tau_h$	23.1
$\tau\tau \to \mu\tau_h$	22.5
$\tau\tau \to \tau_h\tau_h$	42.0

Leptonic decays

- H $\to \tau\tau$ [this thesis][31][32][33]:
  Decays into leptons provide pure access to test the predicted Yukawa couplings. In contrast to the hadronic decays, they are free of hadronisation. The $\tau\tau$ channel has a relatively large branching fraction of $6\,\%$ such that it is the most sensitive leptonic channel. This is despite the fact that tau leptons decay before reaching the detector, which adds some complexity to the analysis. It is conducted in the various sub-channels of the di-τ decay. A τ lepton decays into an electron, muon or hadronically, denoted as τ_h. For a τ pair, this opens six possible final states. These are listed in table 2.1 together with the respective branching fractions. As for the H $\to$ WW* analysis, the sub-channels with two electrons or two muons are too much dominated by $Z \to ll$ background for a competitive measurement of the Higgs boson couplings. The remaining four sub-channels are taken into account for the analysis.
  The $\tau\tau$ channel is sensitive to ggH and VBF. Meanwhile it achieves an inclusive signal strength constraint of $14\,\%$ of the observed value. With the multivariate analysis presented in this thesis, it is also possible to achieve a differential measurement with twelve sub-signals at constraints ranging between $30\,\%$ and $150\,\%$, most of them around $60\,\%$.

- H $\to \mu\mu$ [34][35]:
  Due to the challenging event identification in the H $\to$ c$\bar{\text{c}}$ channel, the H $\to \mu\mu$ channel is the most sensitive one for probing the coupling of the Higgs boson to fermions of the second generation despite the smaller branching fraction. The branching fraction is only $0.02\,\%$ which is the major obstacle of this channel. But the final state is fully reconstructible and the energy of the muons can be measured rather precisely resulting in a $2\,\text{GeV}$ sharp resonance in the invariant mass spectrum. Due to this good resolution, even the major background with two muons from an off-shell Z boson decay can be controlled.
  The latest search performed by CMS on Run-2 data observed this decay channel at a significance of three standard deviations, establishing the first evidence for this process.

2.3 Differential measurements and the STXS framework

After the LHC Run-2, most of the analyses listed in the previous section performed differential measurements not only disentangling the production modes but also different kinematic partitions of the phase space. At the Les Houches workshop in 2015, an initiative has been launched to coordinate such differential measurements not only between the analysis branches but also between the experiments ATLAS and CMS as well as the theory community [36]. A scheme called Simplified Template Cross Sections (STXS) was drafted, which has been further developed and adapted since then, e.g. at the 2019 Les Houches workshop [37].

This effort is motivated by several aspects. One objective is to have a common scheme of exclusive phase space regions, called STXS bins, such that measurements of different analysis branches or the two experiments can be directly compared and combined. While the analysis branches are aligned with the decay channels of the Higgs boson, such a combination makes it possible to comprehensively understand both decay channels and production modes, in particular the involved couplings. Another motivation behind this initiative is to find a scheme that fits both theoretical and experimental requirements. In particular, one objective is to reduce theory dependence as much as possible. While kinematic templates according to the SM prediction will be used in order to allow analyses to choose smaller fiducial volumes and optimise their sensitivity, the bins of the scheme are chosen such that they align with quantities where variations due to theoretical uncertainties would be largest. The considered uncertainties may be inherent to the SM prediction or related to possible effects beyond the SM, e.g. as predicted by the MSSM. This way, related variations are preserved in the measurements and the latter can be reinterpreted later on. This also implies that the STXS scheme is designated for measuring cross sections rather than signal strengths. Besides this theoretical aspect, the scheme has to preserve experimental feasibility and obtain sensitive bins. Therefore, it is partially suited to typically observable event topologies of the production modes.

The first level that is treated differentially, referred to as STXS stage-0, is the splitting of production modes. Due to the different physical processes and involved couplings, current predictions may deviate differently from the observation and it is therefore of primary interest to measure the production modes separately.

The next level, referred to as STXS stage-1 [5], subdivides each production mode further into several kinematic bins. The schemes for ggH and qqH are sketched in figures 2.4 and 2.5. There are schemes for the other production modes as well, but due the lack of sensitivity they are not relevant for this thesis and will not be further discussed here.

In general, the quantities that are used to define the bins are based on truth information retrieved from the simulation of the scattering process without any reconstruction effects. The most relevant quantities of the STXS scheme are the number of particle jets and the transverse momentum of the Higgs boson. The jet definition is aligned with the clustering algorithm used for the reconstruction of experimental data, which is detailed

---

[5]The stage-1 scheme has already evolved over several iterations and the current state is labelled as stage-1.2. This is the only official version that is dealt with in this thesis and for simplicity it will be shortly referred to as stage-1 in following.

Figure 2.4: STXS scheme for ggH (from [38]): The ggH process is split along the transverse momentum of the Higgs boson p_T^H. The lowest bin, which comprises the majority of events, is further split according to the number of emitted particle jets with a transverse momentum larger than 30 GeV. Most resulting groups are further split into bins of p_T^H. From the events with two or more jets, the ones with an invariant mass of the two jets with largest transverse momentum (m_{jj}) larger than 350 GeV are treated separately and binned in m_{jj}. In this phase space region, VBF events appear at a relatively high purity and the ggH binning is aligned with the qqH scheme. Further splits involving the transverse momenta of the Higgs boson combined with jets are foreseen but not yet accessible at reasonable sensitivity by any analysis.

Figure 2.5: STXS scheme for qqH (from [38]): The qqH process is first split according to the number of emitted particle jets with a transverse momentum larger than 30 GeV. Most events of this process have at least two of such jets. The invariant mass of the two jets with largest transverse momentum (m_{jj}) is used to further subdivide the phase space. Most events from the VH process with the W or Z boson decaying into two jets accumulate in the range $60 \,\text{GeV} < m_{jj} < 120 \,\text{GeV}$ according to the masses of the W and Z bosons, respectively. The VBF process dominates the region of $m_{jj} > 350 \,\text{GeV}$ as the scattered quarks from the protons tend to keep their initial back-to-back topology. This phase space is further split along m_{jj} and into two bins of the transverse momentum of the Higgs boson p_T^H in order to look for possible new physics effects in the strongly boosted regime. Further splits involving the transverse momenta of the Higgs boson combined with the two jets are foreseen but not yet accessible at reasonable sensitivity by any analysis.

in section 3.2.2. Only jets with a transverse momentum of at least $30\,\mathrm{GeV}$ are taken into account in order to be close to what is typically required in the measurements to reduce pileup and energy uncertainty effects. By defining bins depending on the number of jets, the dependencies on theoretical uncertainties in the prediction of associated jets and their energy are reduced.

The transverse momentum of the Higgs boson p_T^H is also associated with theoretical uncertainties. For example, its spectrum varies between the calculations at different expansion orders of the strong coupling constant. Furthermore, the experimental sensitivity depends on p_T^H and this dependency is not the same for all decay channels. The measurement of $\mathrm{H} \to \tau\tau$ is more sensitive at higher p_T^H while the measurements in bosonic decay channels are more sensitive at lower p_T^H [36]. At $p_\mathrm{T}^\mathrm{H} > 200\,\mathrm{GeV}$, entering energy scales larger than the top quark mass and starting to resolve the top quark loop in the ggH process, additional theoretical effects need to be considered. On one hand, the approximation of infinite top quark mass that is used in calculations of the ggH process does not apply, resulting in larger theoretical uncertainties [36][39][40]. On the other hand, contributions from heavier particles that might contribute to the loop but still in a point-like way, as predicted for example in the MSSM, could become visible with a resolved top quark loop.

The STXS schemes define the fiducial volumes and provide guidance how to arrange event selections and categorisations accordingly. They are however not meant to replace the conventional differential cross section measurements where a main requirement is to keep the event selection as close as possible to the fiducial volume in order to minimise the extrapolation of the measurement. Within the STXS framework, closer selections are preferred if helpful to optimise the sensitivity and as long as no major dependencies on theoretical uncertainties are introduced. This particularly allows the usage of multivariate analysis techniques, which is done for the analysis of the $\mathrm{H} \to \tau\tau$ channel presented in this thesis.

Experimental conditions at the CMS Experiment
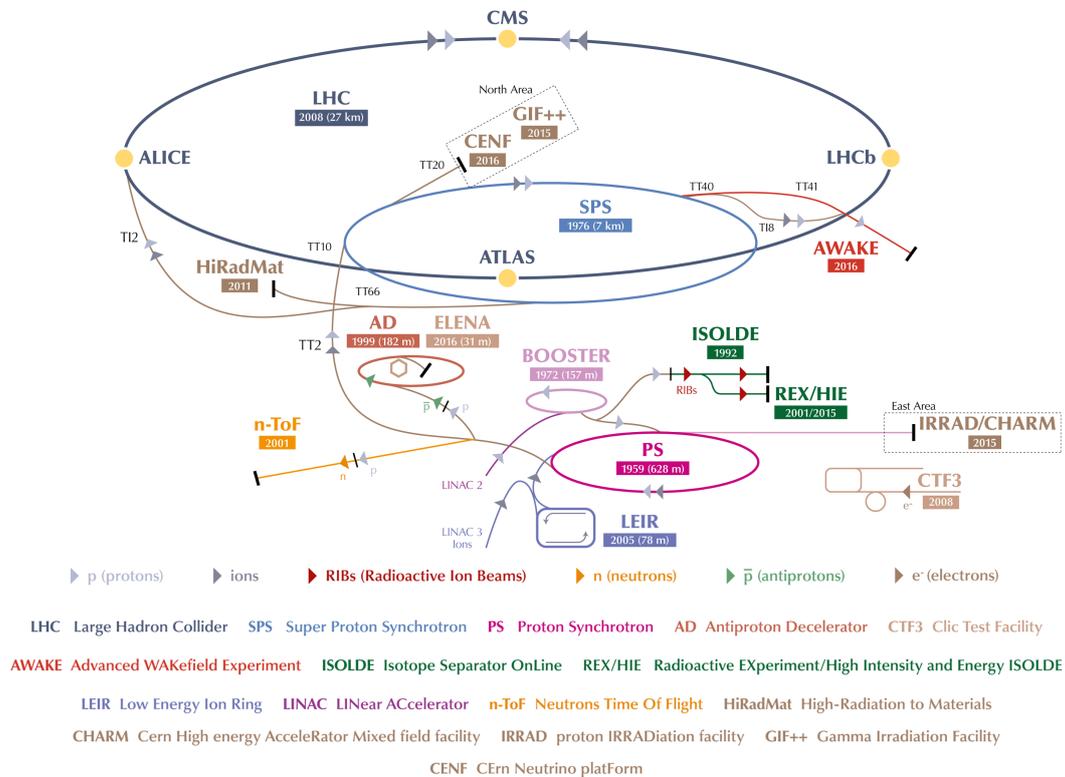
Figure 3.1: The CERN accelerator complex and experiments (from [41])

Almost five decades passed between the proposal of the Brout-Englert-Higgs-Mechanism and the discovery of the Higgs boson. This was due to the experimental challenge of providing collisions at high energy and rate as well as processing the recorded amount of data. The LHC [42] was built to achieve this goal and generates proton-proton collisions at a centre-of-mass energy of 13 TeV and a peak luminosity of $2 \cdot 10^{34} \, \mathrm{cm^{-2} s^{-1}}$ [43]. It is the largest synchrotron that has been built so far and is part of the accelerator complex (figure 3.1) located at CERN, Geneva.

The CMS experiment [44] is one of the two multi-purpose experiments at the LHC (along with ATLAS), both of which count the investigation of the Higgs boson among their major research objectives. From 2016 to 2018, the CMS experiment collected an integrated luminosity of proton-proton collisions of about $140\,\mathrm{fb}^{-1}$. This data enters the latest measurements and in particular the analysis presented in this thesis. The analysis of the collisions requires a complex system of detector components and reconstruction algorithms, which are summarised in this chapter.

3.1 Detector geometry and components

The CMS detector is built in cylindrical symmetry around the beam axis and the nominal collision point. It measures $30\,\mathrm{m}$ parallel to the beam axis and has a diameter of $15\,\mathrm{m}$. A right handed Cartesian coordinate system is defined such that the x-axis points horizontally to the centre of the LHC, the y-axis vertically upward and the z-axis along the beam axis. This is sketched in figure 3.2. In practice, an alternative spherical-like coordinate system is used to align with the symmetry of the detector. The azimuthal angle φ and the pseudorapidity η, which replaces the longitudinal angle θ, are defined as:

$$\varphi = \arctan(x, y)$$
$$\eta = -\ln\tan\left(\frac{\theta}{2}\right) \ \text{ with } \theta = \arccos\left(z/r\right) \tag{3.1}$$

For particles coming from the nominal collision point, these two quantities together with the transverse momentum p_T can also be used to describe the state in momentum space. For φ, η and the radius r, the spatial metric is given by

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2 \cdot \sin^2\theta \cdot [\mathrm{d}\varphi^2 + \mathrm{d}\eta^2]. \tag{3.2}$$

Translated to differences between particle momenta the metric reads

$$\mathrm{d}p^2 = \mathrm{d}|\vec{p}|^2 + p_\mathrm{T}^2 \cdot [\mathrm{d}\varphi^2 + \mathrm{d}\eta^2]. \tag{3.3}$$

A commonly used measure for angular distances is

$$\Delta R = \sqrt{\Delta\varphi^2 + \Delta\eta^2}. \tag{3.4}$$

The metric shows that ΔR is locally isotropic but the actual distance indicated by a given ΔR shrinks with increasing $|\eta|$, i.e. in the very forward or longitudinal direction of the detector. The quantity ΔR can also be understood as the difference between two momenta without the radial component and relative to p_T, which makes it particularly useful for defining cone sizes of particle clusters.

A central element of the detector is the superconducting solenoid. It generates a magnetic field in the longitudinal direction, which bends the trajectories of charged particles and can therefore be used to infer the transverse component of their momentum. Inside the solenoid, the magnetic field measures $3.8\,\mathrm{T}$. In the outer detector components, where
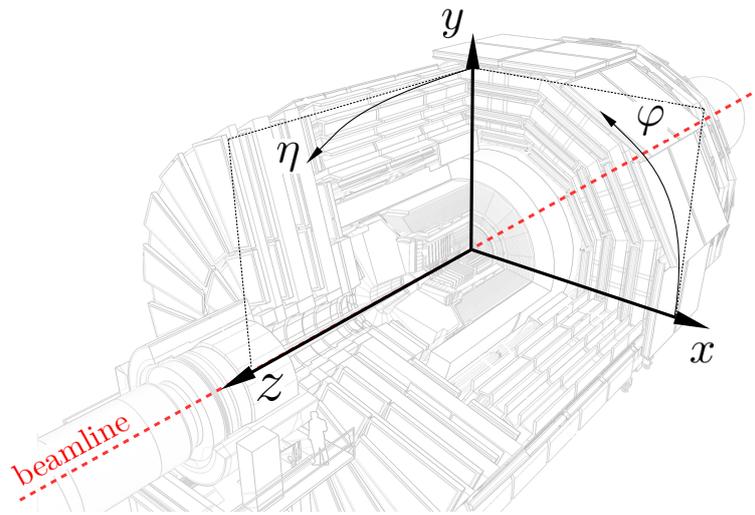
Figure 3.2: Coordinate system of the CMS detector
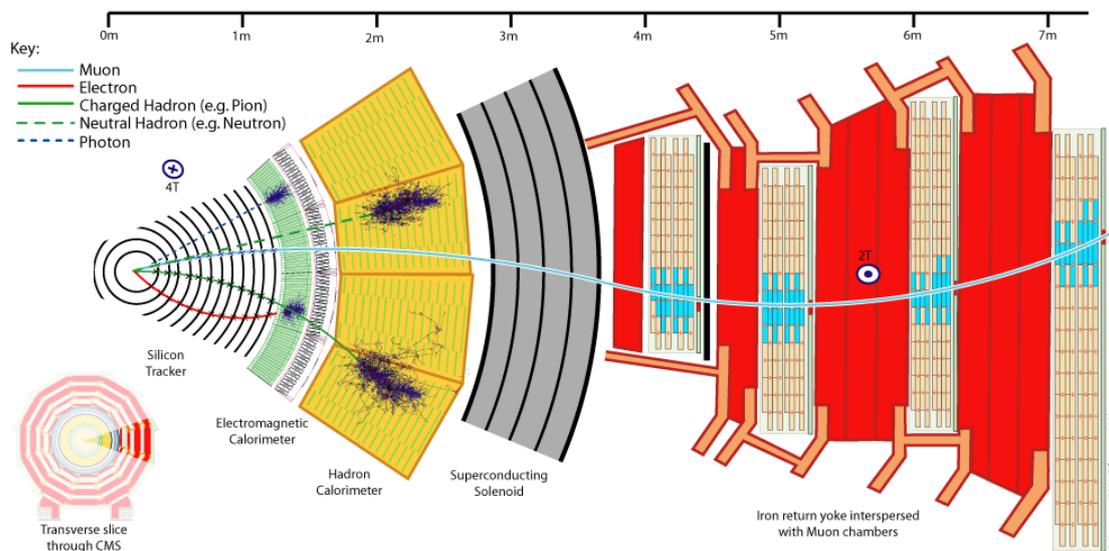
Figure 3.3: Transverse slice through the CMS detector. The technical components and typical particle signatures are shown. (from [45])
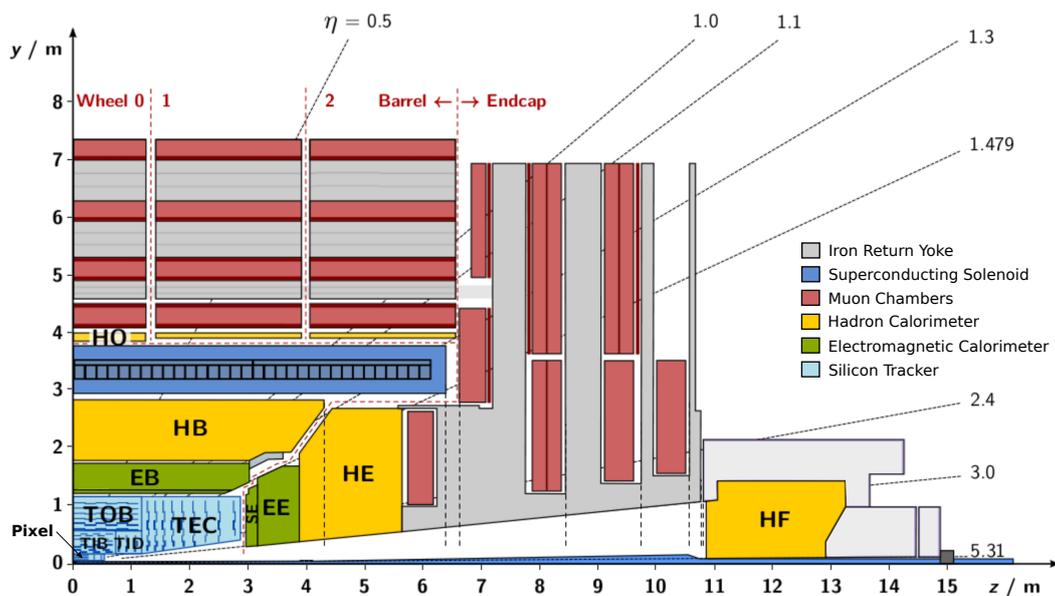
Figure 3.4: Longitudinal slice through the CMS detector, which consists of several parts that can be moved apart from each other for maintenance. There is a central barrel that contains the solenoid (dark blue) and its outer part is subdivided into five wheels. At both ends of the detector, the so-called endcaps are located. The silicon tracker is fully located inside the barrel. It is comprised of the following components: pixel, inner and outer barrel (TIB/TOB), inner disks (TID), tracker endcaps (TEC). The electromagnetic calorimeter is located in both the barrel (EB) and the endcap (EE). In the endcaps, preshower modules (SE) are attached to it. The hadron calorimeter has two components in the barrel, where one is located within the solenoid (HB) and one outside referred to as outer calorimeter (HO). Furthermore, the endcaps contain a part of the hadron calorimeter (HE). An additional forward calorimeter (HF) extends the coverage up to $|\eta| = 5.2$. The muon chambers are embedded in the iron yoke and extend over both barrel and endcaps. (adapted from [46])

the magnetic field is captured by an iron yoke, it measures $2\,\mathrm{T}$. Around the solenoid, several layers of components for particle detection are located. This is sketched in figures 3.3 and 3.4. The transverse projection shown in figure 3.3 also shows typical particle signatures in these layers, illustrating how different particle types can be distinguished. Figure 3.4 shows a longitudinal slice through the detector with a more detailed view of the detector components, their location and coverage in η.

- **Silicon Tracker:** This is the innermost component of the detector. It is designed to detect charged particles and measure their trajectories in the magnetic field in order to infer the transverse momentum. The energy loss of the particles while passing the silicon tracker is negligible compared to their total kinetic energy. Uncharged particles pass the silicon tracker undetected.

  The central part is made of silicon pixel detector modules for high resolution and is positioned at a distance of only a few centimetres from the interaction point. The original pixel detector, which was operated during 2016 and before, consisted of three layers and two additional disks at both ends, resulting in three hits per particle track within its coverage of $\eta \leq 2.5$ [44]. It was upgraded in March 2017 [47] to improve its performance and to cope with the increasing luminosity of the LHC, which started to exceed the design luminosity of $10^{34}\,\mathrm{cm^{-2}s^{-1}}$ at that time. Besides improved components and higher granularity, an additional layer and disk provide a fourth hit per track. The innermost layer is only $2.9\,\mathrm{cm}$ away from the beam.

  Around the pixel detector and at distances from the beam ranging from $20\,\mathrm{cm}$ to $116\,\mathrm{cm}$, several modules of silicon strip detectors form the remaining part of the silicon tracker. Altogether they provide at least nine hits per particle track passing the detector within $|\eta| \leq 2.4$. The total acceptance range of the silicon tracker is $|\eta| \leq 2.5$. The indicated numbers of hits are not taking into account efficiency limitations of the technical components. Moreover, technical defects e.g. due to radiation damage can cause permanently dead pixels/strips.

  One purpose of the tracker system is to determine the transverse momentum of charged particles. The accuracy of this is better than $2\,\%$ for $p_\mathrm{T} < 100\,\mathrm{GeV}$ in the barrel region and better than $10\,\%$ in the endcaps [44]. Another important task is the identification of vertices tracing back the reconstructed tracks. Pileup vertices and associated particles need to be distinguished from the primary vertex, which is chosen to be the one with the largest p_T sum of the associated physics objects. The spatial track resolution at the beamline reaches $10\,\mu\mathrm{m}$ in the transverse direction and $30\,\mu\mathrm{m}$ in the longitudinal direction for tracks at $p_\mathrm{T} = 100\,\mathrm{GeV}$ [44]. For tracks at smaller transverse momentum, this deteriorates to the order of $100\,\mu\mathrm{m}$ due to the larger curvature of the tracks in the magnetic field.

- **Electromagnetic Calorimeter:** It is located around the silicon tracker and is the innermost layer that is designed to absorb particles and measure their energy. It is made of lead tungstate ($\mathrm{PbWO_4}$) crystals, which serve as both the absorber

and scintillator material. The homogeneous material with a short radiation length of 0.89 cm and a Molière radius of only 2.2 cm makes it a very compact calorimeter with high resolution [44]. The crystals are 23 cm (barrel) or 22 cm (endcaps) long, corresponding to about 25 times the radiation length. As depicted in figure 3.3, the electromagnetic calorimeter absorbs electrons and photons while other particles usually pass through without major energy loss. Therefore, it is an essential ingredient for identifying and measuring these two particle types. The tracker, which detects electrons but no photons, helps to distinguish between them.

In the barrel, the crystals have a front face cross section of $22 \times 22 \, \mathrm{mm}^2$ corresponding to approximately 0.0174×0.0174 in η-φ [44]. In the endcaps, the crystals have a front face cross section of $28.62 \times 28.62 \, \mathrm{mm}^2$, which results in a better resolution in θ due to larger distance to the interaction point compared to the central barrel, but not in η or the longitudinal momentum of the measured particles. The energy resolution is of the order of $0.5 \, \%$ for particles below $500 \, \mathrm{GeV}$, where the energy loss due to the limited thickness of the calorimeter is negligible [44].

There is an additional component installed in front of the endcap calorimeter, called preshower, whose purpose is to distinguish photons from neutral pions. Neutral pions typically decay into a pair of photons before reaching the electromagnetic calorimeter. Depending on the energy, the two photons are very close to each other and in the forward region of the endcaps they are often so close that the granularity of the calorimeter is not large enough to resolve these as two separate photons. The preshower is a sampling calorimeter of lead absorber and silicon strip sensors used for tracking the electrons and positrons from the shower. It consists of two layers each, where the thickness of the front absorber layer is two and the one of the rear layer one radiation length, such that most photons initiate a shower. The granularity of the readout elements is about ten times larger than for the actual electromagnetic calorimeter, improving the resolution for pion identification [44].

- **Hadron Calorimeter:** Hadrons deposit only a small fraction of their energy in the electromagnetic calorimeter, which is rather thin for nuclear interactions. This is the case even if they are charged because the energy loss due to bremsstrahlung is smaller than for electrons or positrons because of the larger mass. Therefore, the hadron calorimeter surrounds the electromagnetic calorimeter and is about four times thicker. It is a sampling calorimeter and in the inner barrel and the endcaps, it is made of non-magnetic brass, acting as absorber, and a plastic scintillator [44]. In the barrel region, the main part is located inside the solenoid. But in order to achieve a better absorption it is complemented by the outer calorimeter. This consists of one scintillator layer outside the solenoid or, in case of the central detector wheel where particles pass the least material till they reach the outer calorimeter, of two scintillator layers with an additional iron absorber layer in between. The solenoid serves as an additional absorber material for the outer calorimeter and altogether the hadron calorimeter has an effective thickness of at least 11.8 interaction lengths [44].

The forward calorimeter extends the angular coverage of the detector up to $|\eta| = 5.2$

and therefore reduces the amount of undetected energy leaving the detector in the longitudinal direction. The used technology is different compared to the other parts of the hadronic calorimeter in order to resist the strong irradiation in the vicinity of the beam. It is made of steel absorber and quartz fibres where Cherenkov light is exploited as signal [44].

The energy resolution for hadrons is of the order of $10\,\%$ for momenta above $50\,\text{GeV}$ [48]. Towards smaller momenta of just a few GeV it increases up to $100\,\%$.

- **Muon Chambers:** While most detectable particles are absorbed in the calorimeters, muons are not because they do not interact strongly and the energy loss via bremsstrahlung is too small due to their large mass. The muon chambers are embedded in the iron return yoke outside the solenoid and form a second tracking system for detecting and identifying muons. The magnetic field of about $2\,\text{T}$ in this region bends the muon trajectories, which is used to complement the momentum measurement of the silicon tracker. With this setup, muons can be identified well and their momenta can be measured with a precision of better than $10\,\%$[1] for momenta up to $1\,\text{TeV}$, which makes the muon chambers an important ingredient for the presented analysis [44].

  With the larger distance to the interaction point there is no need for fine-granular and expensive silicon detectors to achieve the same angular resolution as the inner tracker. Drift tubes are used in the barrel region and more granular cathode strip chambers in the endcaps where the muon flux is larger and the magnetic field not homogeneous. Both are complemented by resistive plate chambers, which provide shorter response times and better time resolution, particularly improving the muon triggering system of the detector [44]. The muon chambers cover $|\eta|$ up to 2.4.

3.2 Event reconstruction

For the sake of a physics interpretation, the signals of all detector elements collected during a crossing of the proton bunches (event) are combined and associated with particles. Within each sub-detector, the single cells produce signals, which need to be combined: Tracks are reconstructed from adjacent tracker hits and also the clusters in the grid of calorimeter cells must be identified. Afterwards the preliminary results from the sub-detectors are combined, e.g. checking whether there is a track pointing to the position of a signal in the electromagnetic calorimeter, as expected for an electron, or not, as expected for a photon. This reconstruction of basic particles is done by the so-called Particle-Flow (PF) algorithm and the resulting collection of particles is referred to as PF candidates.

From this collection, several types of high-level objects that appear as a composite of multiple particles in the detector are reconstructed. So called jets are combined from closely adjacent particles. They are usually induced by gluons or quarks from the hard

---

[1]Even $5\,\%$ in the barrel region.

scattering and subsequent hadronisation. But also hadronically decaying τ leptons produce jets with a typical substructure, which is reconstructed as a separate high-level object. Furthermore, the negative sum of the transverse momenta is calculated of all PF candidates. Apart from certain technical shortcomings, only neutrinos would cause this sum to be non-zero and therefore it is a good estimate for the transverse momentum of all produced neutrinos, which cannot be measured otherwise with the detector.

3.2.1 Particle-Flow Algorithm

The Particle-Flow (PF) algorithm [49] is used to reconstruct the particles emerging from the proton-proton collisions and secondary processes. It can be understood as the translation from an image of detector hits into an event description in terms of particles and their momenta. It makes use of the high tracking resolution and granularity of the calorimeters and combines information from the tracking and the calorimeters in order to identify single particles. In contrast to simpler algorithms, which would not go into the substructure of jets and instead determine their total energy based on the corresponding calorimeter cluster, the PF approach makes it possible to distinguish different particle types within a jet, to apply a more dedicated energy calibration and therefore reach a better energy resolution for the total jet.

A first important step is to find the tracks in the image of tracker hits, involving both the silicon tracker and the muon chambers. In order to obtain a set of tracks that contains as few as possible falsely reconstructed tracks but also misses as few tracks as possible, an iterative approach is chosen. It starts with rather tight requirements to identify the best quality tracks first, making it unlikely that the related hits are associated with the wrong tracks which would in turn increase the likelihood of further reconstruction errors. Further iterations with increasingly looser quality requirements follow in order to reconstruct tracks from the remaining hits. In total, ten iterations are performed: The first iterations are seeded by hit clusters in the pixel detector, aiming to reconstruct tracks originating from the hard interactions. The seeds of further iterations include the silicon strip detector in order to identify displaced tracks (outside the pixel detector) which for example originate from interactions with the tracker material. The last iterations are seeded by remaining hits in the muon chambers. Seeds are the starting points for the algorithms to reconstruct complete tracks by iteratively adding further hits compatible with the extrapolation of the current track fragment. The extrapolation includes a hypothesis about the direction and momentum of the track. A Kalman filter is used to alternately select new hits and update this hypothesis. In case of electrons, the energy loss via bremsstrahlung typically causes a significant momentum change within the inner tracker, which is not taken into account by this method. If the track quality (χ^2 of hits with respect to the determined trajectory) exceeds a predefined threshold, it is refitted with a Gaussian-sum filter, which allows for momentum changes along the trajectory.

In the calorimeters, arriving particles cause showers spreading over several calorimeter cells such that their energy is not measured by single cells. Instead, clusters of energy depositions need to be identified. Calorimeter cells that provide a signal above a certain

threshold and stronger than in the neighbouring cells are taken as seeds. So-called topological clusters are grown from these seeds by adding together adjacent cells with a signal above twice the noise level. This way more than one seed can end up in the same topological cluster. A model of clusters with a Gaussian shaped energy distribution of fixed width but variable position and intensity is fitted to the calorimeter deposits within each topological cluster. The number of clusters in the model is chosen to be the respective number of seeds.

Finally, tracks and calorimeter clusters are linked together. Tracks are extrapolated to the calorimeter surface and linked to a cluster if they end up in the same area. If there is more than one overlapping cluster compatible with the track, the cluster with the smallest distance between the extrapolated track and the nominal cluster position is chosen. Similarly, this is done the other way around if there is more than one track. Furthermore, links between the preshower and the electromagnetic calorimeter in the endcaps or between the electromagnetic and hadron calorimeter are created if the nominal position of the cluster in the more granular calorimeter is located within the area of the cluster in the other calorimeter. Any two tracks are linked to each other if they share a secondary vertex. In order to take electron bremsstrahlung into account, tangents are extrapolated from tracks obtained with the Gaussian-sum filter to the electromagnetic calorimeter and matching clusters are linked. In addition, the likely conversion of such photons to an electron-positron pair is taken into account by looking for corresponding track pairs and checking their compatibility with the bremsstrahlung tangents.

From the resulting link collection, muon candidates are considered first because they provide the most distinct signature. They have to be either isolated or meet certain other preconditions in order to avoid the identification of hadron showers reaching the muon chambers (punch-through) as muons. The track and calorimeter entries of accepted muons are masked and electrons and isolated photons are identified from the remaining PF candidates. Due to bremsstrahlung and electron-positron pair production, the signatures of these two particle types have a strong resemblance and so their reconstruction is done in a single step. Finally, remaining clusters in the electromagnetic and the hadron calorimeter are identified as non-isolated photons or hadrons, charged hadrons in case of a linked track.

The resulting collection of particles is the basis on which more complex objects like jets and hadronically decaying τ leptons are reconstructed.

3.2.2 Jet clustering and tagging

The scattering of quarks or gluons and the hadronisation are typical sources of bunches of particles leaving the vertex in a similar direction. These so-called jets are treated as one object in the further analysis. They are not of primary interest but essential elements of specific process signatures. For example, the two quarks scattered during the VBF process (section 2.2.1) typically result in two detected jets.

The anti-k_t algorithm [50] is used to cluster jets from the particles reconstructed with the PF algorithm. The k_t refers to the transverse momentum of the particles, which plays a central role in this algorithm. For the following introduction however, the naming

convention to label it as p_T is kept. The anti-k_t algorithm is one of several algorithms that work with the same logic but differ in the distance measures used. The generalised expressions for these measures, a distance between two objects d_{ij} and a cut-off value d_i, are:

$$d_{ij} = \min(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p})\frac{\Delta R_{ij}^2}{R^2}$$
$$d_i = p_{\mathrm{T},i}^{2p}$$

(3.5)

It involves the angular distance ΔR_{ij} between particles i and j as defined in section 3.1 and a parameter R, which can be understood as cone radius of the jet and is chosen to be $R = 0.4$. With certain values of the parameter p, known clustering algorithms are obtained. It is $p = -1$ in the anti-k_t algorithm[2]. In an iterative procedure, the smallest of these distances is determined: If it is a d_{ij}, the two particles are merged and their momenta are added. If it is a d_i, the corresponding particle, which may be the result of previous iterations of merging, is considered a jet and masked for all further iterations. This step is repeated until no objects are left. The effect of the negative parameter p is that the distance measure does not depend on the momentum of the softer particle, which makes the jet clustering stable against the distribution of small momenta and results in a more uniform, cone-like jet boundary [50]. The given p_T-weight in the distance measure gives priority to high-p_T objects, which are already merged particles in particular. In the exemplary case of a high-p_T particle surrounded by much softer particles, the resulting jet will form a smooth cone around the high-p_T particle with a radius of $R = 0.4$.

If all PF candidates are taken into account for the clustering, the resulting jet collection is referred to as PF-jets. For this analysis however, a different collection is used in order to reduce the impact of pileup collisions. Charged hadrons can be associated with vertices via their tracks. Hence, all charged hadrons that unambiguously originate from pileup vertices are removed from the collection of PF candidates before running the jet clustering with the algorithm as described above [51].

There are many approaches to further classify such jets or reconstruct them in a more sophisticated way and they are chosen depending on the analysis needs. For the $H \to \tau\tau$ analysis in this thesis, the reconstruction of hadronically decaying τ leptons, which is discussed in the next section, is of major interest. A second method that is used is b-tagging. The algorithm used for this thesis only serves to identify jets that are initiated by bottom quarks (shortly b-jets). It does not involve a re-reconstruction or re-calibration of these jets. While b-jets are not part of the major signal processes targeted by this analysis, they appear as part of an important background process, the production of top quark pairs. The b-tagging method is used to reject such background events. The top quarks decay to leptons or hadrons, which resembles the decay of the τ pair. However, top quark decays also lead to b quarks in the final state, initiating b-jets that can be used to distinguish between the production of top quark pairs and the signal. The lifetime of b-flavoured hadrons is large enough that often secondary vertices or at least displaced tracks can be resolved.

---

[2]The choice $p = 1$ corresponds to the k_t algorithm, which motivated the name "anti-k_t".

Table 3.1: Branching fractions $\mathcal{B}$ of τ decays to electrons, muons and hadrons. Charged hadrons are denoted by $h^{\pm}$. Adapted from [53], data from [22].

decay mode	$\mathcal{B}\,(\%)$
$\tau^- \to \nu_\tau \bar{\nu}_e e^-$	17.8
$\tau^- \to \nu_\tau \bar{\nu}_\mu \mu^-$	17.4
$\tau^- \to \nu_\tau$ and hadrons	64.8
$\tau^- \to \nu_\tau h^-$	11.5
$\tau^- \to \nu_\tau h^- \pi^0$	26.0
$\tau^- \to \nu_\tau h^- \pi^0 \pi^0$	9.5
$\tau^- \to \nu_\tau h^- h^+ h^-$	9.8
$\tau^- \to \nu_\tau h^- h^+ h^- \pi^0$	4.8
remaining modes with hadrons	3.2

The b-tagging method used at CMS is called DeepJet [52] and makes use of a deep neural net. Aside from kinematic properties of the jet itself, it takes into account information from the individual charged and neutral PF objects in the jet, as well as secondary vertices. The DeepJet algorithm is trained to identify different jet flavours but only the b-tag information using the medium working point as defined in [52] is exploited in this analysis for the named purpose.

3.2.3 Reconstruction of hadronically decaying τ leptons

The mean lifetime of τ leptons is 290 fs [22] and at the LHC they typically travel a couple of millimetres before they decay [53]. Hence, only the decay products reach the detector. If the τ decays leptonically, only the electron or muon is visible in the detector and the track displacement with respect to the primary vertex is too small to clearly distinguish it from prompt electrons or muons. No dedicated reconstruction or discrimination against prompt leptons is run for leptonically decaying τ leptons. 65 % of the τ leptons decay hadronically [22] into several detectable particles, which can be reconstructed and identified with reasonable efficiency. The different τ decays and branching fractions are listed in table 3.1. A compressed scheme of decay modes, which reflects the types that are distinguished by the reconstruction algorithm, is given in the same table.

The reconstruction of hadronically decaying τ leptons is done with the so-called hadron-plus-strip (HPS) algorithm [53][54]. It operates on PF-jets, as defined in the previous section, identifies charged hadrons and reconstructs π^0 mesons from photons and electrons within the jet cone.

The π^0 decays promptly into two photons before it reaches the detector. Those photons are likely to convert into e^+e^- pairs within the tracker, which in turn can emit bremsstrahlung. In the magnetic field, the electrons and positrons are deflected in opposite directions of φ. Therefore, the electrons and photons with $p_T > 0.5$ GeV in the jet are clustered to strips, which have a rectangular shape in the η-φ-plane with a larger extent in the φ-direction. The size of the rectangles is chosen dynamically and depends on the

momenta of the involved particles in order to take into account the momentum-dependent collimation of the decay products. Within a minimum of 0.05×0.05 and a maximum of 0.15×0.30, the strip size is chosen to be

$$
\begin{aligned}
\Delta\eta &= 0.2 \cdot \left((p_\text{T}(e/\gamma)/\text{GeV})^{-0.66} + (p_\text{T}(\text{strip})/\text{GeV})^{-0.66} \right) \\
\Delta\varphi &= 0.35 \cdot \left((p_\text{T}(e/\gamma)/\text{GeV})^{-0.71} + (p_\text{T}(\text{strip})/\text{GeV})^{-0.71} \right)
\end{aligned}
\tag{3.6}
$$

for each merging step of an electron or photon (e/γ) [54]. The clustering starts with the electron or photon with the largest p_T and adds further electrons or photons within the given strip window, in decreasing order of p_T. After adding a particle, the strip location is always recalculated as the p_T-weighted mean of the merged objects. If no further particles to be added are found, the strip is closed and the procedure repeats with a new strip. All strips that reach $p_\text{T} > 2.5\,\text{GeV}$ are considered as π^0 candidates.

The π^0 candidates are combined with the charged hadrons $h^\pm$ in the jet, which must satisfy $p_\text{T} > 0.5\,\text{GeV}$ and fulfil some further requirement on the compatibility with a common secondary vertex or the primary vertex. The resulting collection of charged and neutral hadrons is compared to the possible decay modes listed before extended by objects with two charged hadrons. The following requirements are imposed for the respective hypotheses:

- $3\,h^\pm,\ 1\,\pi^0$: $0.9\,\text{GeV} - \Delta m_{\tau_\text{h}} < m(\tau_\text{h}) < 1.6\,\text{GeV} + \Delta m_{\tau_\text{h}}$, total charge ± 1

- $3\,h^\pm$: $0.8\,\text{GeV} < m(\tau_\text{h}) < 1.5\,\text{GeV}$, total charge ± 1

- $2\,h^\pm,\ 1\,\pi^0$: $0.4\,\text{GeV} - \Delta m_{\tau_\text{h}} < m(\tau_\text{h}) < 1.2\,\text{GeV} \cdot \sqrt{p_\text{T}(\tau_\text{h})/100\,\text{GeV}} + \Delta m_{\tau_\text{h}}$

- $2\,h^\pm$: $m(\tau_\text{h}) < 1.2\,\text{GeV}$

- $1\,h^\pm,\ 2\,\pi^0$: $0.4\,\text{GeV} - \Delta m_{\tau_\text{h}} < m(\tau_\text{h}) < 1.2\,\text{GeV} \cdot \sqrt{p_\text{T}(\tau_\text{h})/100\,\text{GeV}} + \Delta m_{\tau_\text{h}}$

- $1\,h^\pm,\ 1\,\pi^0$: $0.3\,\text{GeV} - \Delta m_{\tau_\text{h}} < m(\tau_\text{h}) < 1.3\,\text{GeV} \cdot \sqrt{p_\text{T}(\tau_\text{h})/100\,\text{GeV}} + \Delta m_{\tau_\text{h}}$

- $1\,h^\pm$:   -

The parameter Δm_{τ_h} serves to take into account the smearing of the τ_h mass induced by the strip reconstruction and depends on the strip kinematics as defined in [54]. Finally, all hadrons within the reconstructed τ_h are required to be located in a cone of

$$
\Delta R \leq
\begin{cases}
0.05 & p_\text{T}(\tau_\text{h}) > 60\,\text{GeV} \\
3.0\,\text{GeV}/p_\text{T}(\tau_\text{h}) & p_\text{T}(\tau_\text{h}) \in [30\,\text{GeV}, 60\,\text{GeV}] \\
0.1 & p_\text{T}(\tau_\text{h}) < 30\,\text{GeV}
\end{cases}
\tag{3.7}
$$

around the momentum axis of the τ_h. This takes into account the p_T-dependent collimation of the decay products. If there are several passing hypotheses, the one with the largest p_T of the resulting τ_h is chosen, which defines the reconstructed τ_h associated with the jet under consideration.

So far, this procedure defines τ_h candidates but it is not clear whether the underlying object is indeed a τ_h. They cannot be perfectly identified because their topology can well be mimicked by quark- or gluon-induced jets. It helps that a τ_h has only few particles inside and is typically isolated. But in turn, it can also happen that an electron or muon is misidentified as τ_h. Hence, it is very important to impose further identification criteria on the τ_h candidates.

The τ_h identification is performed with a convolutional deep neural network classifier called DeepTau [55]. It is trained to distinguish between four classes at the same time, in a way that is similar to the concept used for the event classification in this analysis and described in chapter 5. The classes are: genuine τ_h, quark- or gluon-induced jets, electrons, and muons. One set of inputs to the DeepTau classifier is low-level information like four-momenta from single objects in the τ_h cone, provided via $\eta \times \varphi$ grids. These are processed separately for electrons, muons and other PF candidates in two selections respectively. One selection takes into account particles within the inner cone as defined in equation 3.7. The other selection takes into account all particles within a cone of $\Delta R < 0.5$. The resulting six sets of low-level inputs are processed via convolutional layers respectively and subsequent dense layers. Another set of inputs contains high-level information, e.g. different isolation types taking into account only charged or neutral objects. It is processed through dense layers and finally merged with the low-level information via further dense layers. From the output values $\vec{\alpha}$ of the neural net, binary criteria to reject one of the named background types i are defined as

$$\frac{\alpha_{\tau_h}}{\alpha_{\tau_h} + \alpha_i} \geq W.P. \tag{3.8}$$

with *W.P.* being some value defining a working point of the classifier. Performance benchmarks for DeepTau and its different working points can be found in [55].

3.2.4 Missing transverse energy

Part of the energy from the τ decays leaves the detector as neutrinos and is not directly measured. Instead, the fact is exploited that the incoming protons enter with negligible transverse momentum and the transverse momenta of all produced particles except the neutrinos can be measured almost completely. Hence, the vectorial p_T sum of all produced particles, including neutrinos, is expected to be zero, which can be transformed to

$$\sum_{\nu} \vec{p}_T(\nu) = -\sum_{\text{PF cand.}} \vec{p}_T(\text{PF cand.}) \tag{3.9}$$

implying that the total p_T of all neutrinos can be determined as the negative vectorial p_T sum of all PF candidates. This is the basic concept of Missing Transverse Energy ($\not{E}_T$).

Like for jets, different versions exist in the context of pileup. In this analysis, the plain negative momentum sum is not used directly but a $\not{E}_T$ definition following the Pileup-Per-Particle-Identification (PUPPI) approach [56] as implemented by CMS [57].

In this approach, each single PF candidate is regarded and a probability w that it does not originate from a pileup interaction is derived from vertex information and surrounding

particles. Charged particles within the η range of the tracker can be associated with vertices in most cases and hence they are assigned $w = 1$ if they are associated to the primary vertex or $w = 0$ else. Charged particles that are not associated with any vertex are assigned $w = 1$ if their distance to the primary vertex along the beam axis is smaller than $0.3\,\text{cm}$ and $w = 0$ otherwise. Particles outside the tracker range, i.e. $|\eta| > 2.5$, or neutral particles cannot be assigned to vertices. Hence, the probability is estimated from surrounding particles with $\Delta R < 0.4$. Parton showering tends to result in clusters of particles [56] and therefore neutral particles are usually surrounded by other particles from the primary vertex. This does not hold for isolated photons, which is why they are assigned $w = 1$ provided that they are located in the range of the tracker and pass certain quality requirements. For the remaining particles i, a measure α is computed as

$$\alpha_i = \log \sum_j \left(\frac{p_{\text{T},j}}{\Delta R_{ij}} \right)^2. \tag{3.10}$$

The sum over j includes all particles within the surrounding cone defined above. Within the η range of the tracker, only charged particles from the primary vertex are taken into account. Taking the median and standard deviation of α for pileup particles, a p-value is calculated for a given α_i via a χ^2 approximation. The parameters used in this calculation are optimised for different η intervals and are documented in [57]. The p-value is small if α_i comes out too large for typical pileup and hence the probability w is then taken as the difference of the p-value to 1. The resulting probabilities are set to zero if they are below 0.1 or if $w_i \cdot p_{\text{T},i} < A + B \cdot N_{\text{vtx}}$. This is to prune the particle collection and take into account the pileup activity as determined from the reconstructed number of vertices N_{vtx}. A and B are phenomenological constants [57].

The derived probabilities are applied as weights to the momenta of the particles, i.e. suppressing their contribution if the probability to be due to pileup is large. The following $\not{E}_{\text{T}}$ definition based on this approach is used in this analysis:

$$\vec{\not{E}}_{\text{T}} = - \sum_{\text{PF cand.}} w_{\text{PF cand.}} \cdot \vec{p}_{\text{T}}(\text{PF cand.}) \tag{3.11}$$

3.2.5 A note on misidentifiation

Only part of the background processes in this analysis yield genuine τ leptons. These are mostly the production of a Z boson decaying into a τ pair and minor contributions from top quark decays or the production of two vector bosons.

Besides this, a large part of the background is due to the misidentification of objects in the detector. As outlined before, τ leptons decaying into electrons or muons cannot be distinguished from prompt electrons or muons. Hence, even though the correct final-state particle may be picked in the $e\mu$, $e\tau_{\text{h}}$ and $\mu\tau_{\text{h}}$ channels, it may be a prompt lepton without a τ being involved.

In addition, there are cases where the reconstructed PF-type is incorrect, which happens frequently with hadronically decaying τ leptons. Quark- or gluon-induced jets may pass the τ_{h} identification discriminant if they contain a relatively small number of tracks

in the detector. Moreover, electrons and bremsstrahlung can be misidentified as τ_h, in particular if some energy deposits reach the hadron calorimeter. Other particles or pileup going in the same direction may contribute to this. The same can happen with muons if the reconstruction in the muon chambers is incomplete or the muon decays early. In the transition regions from barrel to endcap or between the muon wheels, the misidentification rate of leptons is higher.

Jets induced by quarks or gluons can also be misidentified as leptons, usually as electrons if the shower develops early and the energy deposit in the hadron calorimeter is small. In rare cases, a jet may be identified as muon in case of "punch-through" where part of the shower reaches the muon chambers, but this is suppressed by the usually large energy deposit in the calorimeters and the isolation requirements imposed on the muon.

Several processes that constitute background due to these forms of misidentification have to be taken into account. Most prominent is the production of jets via strong interactions, which are misidentified as τ_h, contain leptons or involve signatures misidentified as leptons. In the (semi-)leptonic decay channels, the production of W bosons decaying to light leptons is an additional important background. The lepton can be misidentified as the leptonic τ candidate and a jet from the W recoil as the τ_h, like in the previous case. The production of a Z boson decaying into electrons or muons constitutes another prominent background. Given the misidentification of a lepton or associated jets as a τ_h, it is mainly present in the semi-leptonic decay channels. Apart from this, processes where two genuine τ leptons are produced also typically yield a small fraction where one of the τ leptons is not selected as τ candidate but a misidentified jet.

3.3 Data acquisition

The presented reconstruction techniques are too complex to run in real time during data-taking and hence not suitable for selecting the events to be stored on tape. The LHC delivers bunch crossings at a rate of 40 MHz [59], which means that every 25 ns new data must be evaluated and the decision made whether the event is stored or discarded. Therefore, a system of alternating reconstruction and selection steps, in which the data rate is reduced and the reconstruction effort increased, is applied. This is set up as a two-tier trigger system [59], i.e. there are two major stages where data is selected.

The first level, referred to as L1 trigger, performs a very basic reconstruction. In a first selection it reduces the event rate from 40 MHz to 100 kHz [59]. It is realised with dedicated hardware components. The reconstruction only involves the calorimeters and muon chambers because track reconstruction in the silicon tracker is not feasible at the required rate. At this stage, merely the presence of energetic particles is checked. Unless they are muons or not detectable they end up in the calorimeters.

The workflow of the L1 trigger is depicted in figure 3.5. The calorimeter part takes into account both the electromagnetic and the hadron calorimeter. In a first step, particle candidates and energy sums are reconstructed in regional units in parallel. These are subsequently combined by a global instance, jets are clustered and a basic version of $\slashed{E}_\mathrm{T}$ is computed. In parallel, tracks are reconstructed in the muon chambers, which
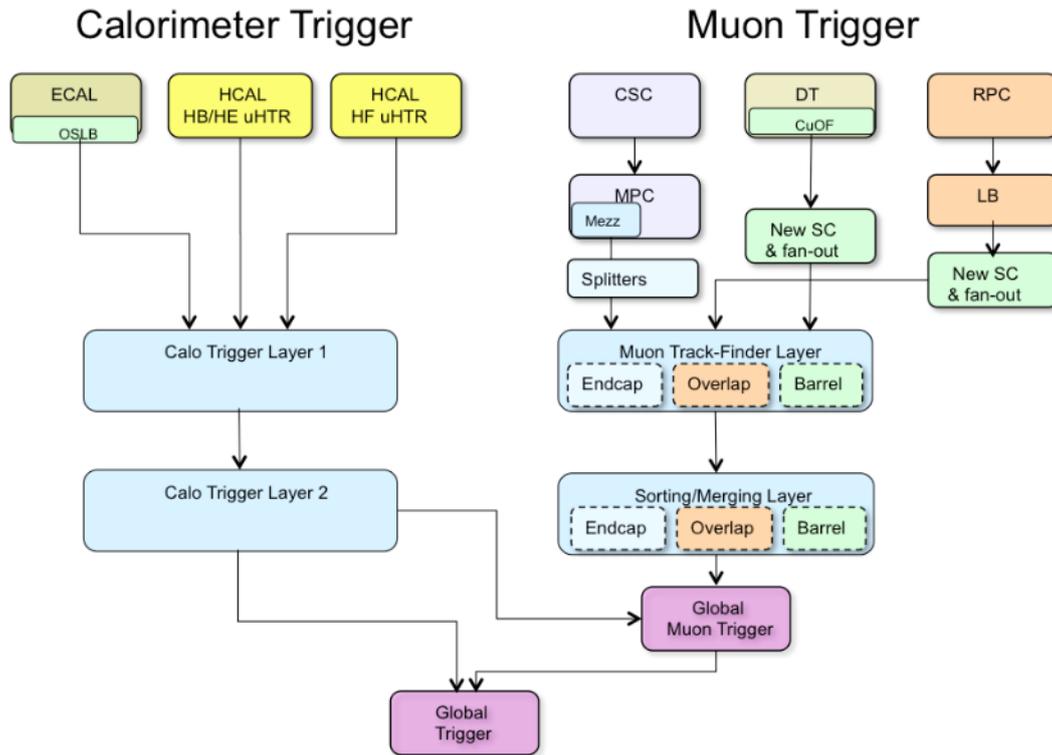
Figure 3.5: Workflow of the L1 trigger system (from [58])

happens independently for three regional units where either only the drift tubes or the cathode strip chambers are located or both systems overlap. The resulting candidates are merged at a global stage and quality requirements are applied including isolation retrieved from the calorimeter results. Various trigger paths are defined to look for certain patterns, e.g. just muons above a certain energy. If at least one of these claims interest, corresponding candidate collections are created and passed on as seeds to the next trigger stage. Otherwise the event is discarded.

The High-Level Trigger (HLT) operates on top of the L1 trigger. At the reduced data rate of about 100 kHz, a more sophisticated reconstruction involving all detector components can be performed, albeit with a simplified PF reconstruction. It is implemented as a software trigger on conventional hardware. Various HLT paths are defined that look for specific signatures. Corresponding L1 trigger results are used as seeds and for each HLT path additional reconstruction steps in combination with selection criteria are applied. In principle, they have access to the information from all detector components in full detail. But the first part only relies on the information from the L1 seeds and a more detailed reconstruction in the calorimeters and muon chambers. A preselection is made based on this such that the computationally more expensive track reconstruction for the inner tracker is only run for few events. Most events are processed within 50 ms but with the simplified PF reconstruction it takes up to 1 s for some events [60]. Inner tracks are

matched to calorimeter clusters and outer muon tracks. Subsequently electron, muon, photon, and jet candidates are reconstructed. Also $\not{E}_T$ is computed and a reconstruction of τ_h candidates is performed. Based on this, the HLT finally decides whether an event is stored or not. If yes, information about which HLT paths were triggered and the corresponding objects reconstructed at this level are stored in addition. The selected events undergo the final event reconstruction as outlined in the previous sections without technical time constraints.

The HLT reduces the data rate to the order of $1\,\mathrm{kHz}$ [60]. Depending on the instantaneous luminosity, the selection requirements have to be adapted. HLT paths exist with different p_T thresholds and within a data-taking period the rates of the ones with lower threshold are reduced whenever necessary, dropping a certain fraction of events. From 2016 to 2018, the instantaneous luminosity increased on average and hence the p_T thresholds were raised. HLT paths that require two objects, e.g. an electron and a τ_h, naturally fire at a lower rate than paths with only one required object. For these, lower p_T thresholds are imposed. All this influences the data selection applied by the analyses. For example, the combination of data triggered by either single-object paths or double-object paths can be beneficial for the size of the selected dataset. Another detail that is relevant for the analysis is the data modelling close to the p_T threshold. The clear cut on the p_T of the HLT objects is smeared out in the final reconstruction and is difficult to model simulation. Therefore, a p_T threshold increased by $1\,\mathrm{GeV}$ cutting away the most critical region is part of the offline selection.

Data selection and modelling

From the reconstructed events, a subset containing potential $H \to \tau\tau$ decays is selected. Only a small part of these events yields genuine $H \to \tau\tau$ decays and most events originate from various background processes. Therefore, the different kinematic distributions of the processes are taken into account in order to interpret the measured data. This interpretation is done via a fit of a model to the data, where the model is based on predictions by the SM and the knowledge of the detector performance. The parameters of interest, which scale the signal, are part of the model such that the fit yields the measurement of these parameters as a result. The kinematic distributions that appear in this analysis after the event selection are non-trivial. Because of that, both data and the model are used in the form of histograms, which are either filled with single events, observed or simulated, or derived by certain extrapolation techniques.

The topics of this chapter are the selection of events that are of interest for this analysis, the model used to describe this data and additional corrections that are applied in order to improve the model.

4.1 Offline selection criteria and choice of the τ pair

This analysis takes four of the six di-τ decay channels into account, i.e. $e\mu$, $e\tau_{\text{h}}$, $\mu\tau_{\text{h}}$ and $\tau_{\text{h}}\tau_{\text{h}}$. The ee and $\mu\mu$ channels are not beneficial to the sensitivity of the analysis due to the large background from $Z \to ll$ decays and their small share of only $6.2\,\%$ of all di-τ decays (table 2.1).

For each of the investigated channels, events with a suitable τ pair candidate are selected. This is based on dedicated trigger selections as listed in table 4.1. In the $e\mu$ channel, cross-triggers looking for pairs of an electron and a muon are used. In the semi-leptonic channels, combinations of cross-triggers looking for lepton-τ_{h} pairs and single-lepton-triggers are used. The cross-triggers impose lower p_{T} thresholds on the lepton and the single-lepton-triggers do not impose any requirements on the τ_{h} candidate, which makes it possible to apply better object identification based on the offline reconstruction. The selection for the $\tau_{\text{h}}\tau_{\text{h}}$ channel is done using di-τ_{h} cross-triggers. Combinations of the same trigger type with different p_{T} thresholds are used in some cases, because of the artificially reduced trigger rate of the HLT paths with the lower threshold, which is applied by the

Table 4.1: Trigger selection per final state and year. Events are selected if they fire at least one of the indicated trigger paths. Each path requires either one or two HLT objects at a minimum $p_\mathrm{T}/\mathrm{GeV}$ as indicated in brackets.

channel	year	trigger selection	
$e\mu$	all	$e(12)\mu(23)$	$\vee\ e(23)\mu(8)$
$e\tau_\mathrm{h}$	2016	$e(24)\tau_\mathrm{h}(20)$	$\vee\ e(25)$
	2017	$e(24)\tau_\mathrm{h}(30)$	$\vee\ e(27) \vee e(32) \vee e(35)$
	2018	$e(24)\tau_\mathrm{h}(30)$	$\vee\ e(32) \vee e(35)$
$\mu\tau_\mathrm{h}$	2016	$\mu(19)\tau_\mathrm{h}(20)$	$\vee\ \mu(22)$
	2017 & 2018	$\mu(20)\tau_\mathrm{h}(27)$	$\vee\ \mu(24) \vee \mu(27)$
$\tau_\mathrm{h}\tau_\mathrm{h}$	2016	$\tau_\mathrm{h}(35)\tau_\mathrm{h}(35)$	
	2017 & 2018	$\tau_\mathrm{h}(35)\tau_\mathrm{h}(35) \vee \tau_\mathrm{h}(40)\tau_\mathrm{h}(40)$	

experiment to stay below the maximum data rate. The HLT objects are required to match the considered τ-candidates from the offline reconstruction within $\Delta R < 0.3$ in the $e\mu$ channel and $\Delta R < 0.5$ in the remaining channels. Further quality requirements are imposed on each electron, muon or τ_h serving as τ-candidate. Those are listed in table 4.2 and further explained in the following.

Requirements on the impact parameters are imposed to reject objects that originate from pileup vertices.

Each object must pass the dedicated identification (ID) requirements for the given object type. For τ_h, this is the DeepTau ID referenced in section 3.2.3, which discriminates against electrons, muons, and jets induced by quarks or gluons. The τ_h candidates that are reconstructed with two charged pions, which are not genuine unless they are not correctly reconstructed, are not taken into account. For electrons, a multivariate identification based on a boosted decision tree is applied. Besides kinematic quantities, it consumes information about track quality and the shower shape, in particular the ratio of energy deposits in the electromagnetic and hadron calorimeter. The identification requirement of muons is simpler as their signature is naturally more unique. Muons must be reconstructed as either tracker or global muons, where the former are only reconstructed as a track in the inner tracker and a hit in the muon chambers. Global muons comprise reconstructed tracks in both parts. The tracks must satisfy requirements on the track fit quality and the compatibility between the detector segments.

Electrons and muons from τ decays are generally isolated, which means that there is not much momentum flux related to other particles in the vicinity of the lepton. This is different if they originate from heavy-flavoured quark decays. Therefore, a corresponding requirement is imposed, defining the relative isolation as

$$I_\mathrm{rel}(l) = \frac{\sum_{h^\pm} p_\mathrm{T}(h^\pm) + \max\left(0, \sum_{h^0} p_\mathrm{T}(h^0) + \sum_\gamma p_\mathrm{T}(\gamma) - \Delta_\mathrm{PU}\right)}{p_\mathrm{T}(l)}. \tag{4.1}$$

This corresponds to the sum of transverse momenta of charged hadrons $h^\pm$ from the primary vertex, neutral hadrons h^0 and photons γ relative to the transverse momentum

Table 4.2: Selection criteria imposed on τ candidates per final state

criterion type	specification for		
	e	μ	τ_h
impact parameter in cm	$d_{xy} < 0.045$ $d_z < 0.2$	$d_{xy} < 0.045$ $d_z < 0.2$	$d_z < 0.2$
ID	MVA $90\,\%$	medium	DeepTau: tight vs. jets & tight vs. e ♠ VVloose vs. e ♡♣ tight vs. μ ♡ Vloose vs. μ ♠♣
isolation	$I_\mathrm{rel} < 0.1$	$I_\mathrm{rel} < 0.2$ ♢ $I_\mathrm{rel} < 0.15$ ♠♡♣	-
minimum p_T in GeV †	$p_\mathrm{T} > 15$ ♢ $p_\mathrm{T} > 25$ ♠	$p_\mathrm{T} > 15$ ♢ $p_\mathrm{T} > 20$ ♡	$p_\mathrm{T} > 30$ ♠♡ $p_\mathrm{T} > 40$ ♣
η range	$|\eta| < 2.4$ ♢ $|\eta| < 2.1$ ♠	$|\eta| < 2.4$ ♢ $|\eta| < 2.1$ ♡	$|\eta| < 2.3$

† higher thresholds may apply to individual objects depending on the matched trigger path
♢ $e\mu$ channel, ♠ $e\tau_\mathrm{h}$ channel, ♡ $\mu\tau_\mathrm{h}$ channel, ♣ $\tau_\mathrm{h}\tau_\mathrm{h}$ channel

of the tested particle. In case of electrons, the sum runs over all particles within a cone of $\Delta R < 0.3$, in case of muons it is $\Delta R < 0.4$. The neutral part is corrected for pileup heuristically, denoted as Δ_{PU}.

In general, the p_{T} of the candidates is required to be at least 1 GeV above the threshold of one of the firing triggers, 5 GeV in case of τ_{h}, in order to exclude the turn-on regions of the triggers, which are difficult to model. Further minimal p_{T} requirements according to table 4.2 are applied in order to ensure proper modelling of the data.

The η ranges are chosen such that the candidates are within the coverage of the tracker, including the cone size in case of τ_{h}. In the channels with one lepton, the η range is chosen more narrow in order to have the isolation cone covered as well and therefore have a more reliable isolation criterion.

Possible τ pairs are constructed by combining two of the τ candidates identified so far. The candidates forming a τ pair are required to be separated by $\Delta R > 0.5$, in the $e\mu$ channel $\Delta R > 0.3$. If there is more than one valid τ pair in an event, which can happen due to the occurrence of multiple τ_{h} candidates, the one containing the τ_{h} with the larger DeepTau vs. jets raw value is chosen. In the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ channel, the τ_{h} candidates with larger p_{T} in each pair are compared first, then the other one if necessary.

Multiple τ pairs due to electrons or muons do not appear because additional leptons are vetoed. This is in order to prevent overlap with other decay channels. Events are discarded if they contain additional electrons or muons, which satisfy the selection requirements of the $e\mu$ channel as listed in table 4.2 with the following adaptations: $I_{\mathrm{rel}} < 0.3, p_{\mathrm{T}} > 10$ GeV and additionally for electrons $|\eta| < 2.5$.

Events where a suitable τ pair is identified must pass further requirements to end up in the final analysis selection as detailed in the following. Some of these criteria rely on the transverse mass, which is defined as:

$$m_{\mathrm{T}}(x, y) = \sqrt{2p_{\mathrm{T}}(x)p_{\mathrm{T}}(y)\left(1 - \cos\left(\Delta\phi(x, y)\right)\right)} \tag{4.2}$$

- The τ candidates must carry opposite charge.

- Events with a pair of oppositely charged electrons or muons are discarded in order to suppress background from $Z \to ll$ decays. This goes beyond the veto of additional leptons mentioned above because the selection requirements for the leptons of these pairs are loosened. Muons are selected according to the requirements of the $e\mu$ channel but without the identification. It is sufficient if they are reconstructed as tracker or global muons. Electrons for this veto are chosen according to the $e\mu$ channel selection but with $|\eta| < 2.5$ and looser identification.

- Events with pathological $\not{E}_{\mathrm{T}}$ [61] are discarded.

- In the $e\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels, $m_{\mathrm{T}}(e/\mu, \not{E}_{\mathrm{T}}) < 70$ GeV is required. The rejected region serves as a sideband for the data driven modelling technique described in section 4.3.2.

- In the $e\mu$ channel, events containing at least one b-tagged jet are vetoed in order to remove a large amount of background while preserving sensitivity of the analysis.

- In the $e\mu$ channel, $m_{\mathrm{T}}(e+\mu, \not{E}_{\mathrm{T}}) < 60\,\mathrm{GeV}$ is required in order to exclude events that enter the analysis of $H \to WW$ events, which is necessary for future combinations of the analyses.

Jets do not have an influence on the event selection but are relevant for the event classification discussed in the next chapter. Jets for the purpose of this analysis are jets as described in section 3.2.2 with charged hadrons from pileup subtracted and the following further properties. A jet has to pass a jet ID setting constraints on the energy distribution over electron, photon, neutral and charged hadron constituents. The exact conditions can be found in [62]. Furthermore the jets must satisfy $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 4.7$. If there are at least two jets in an event, the di-jet system is defined as the sum of the momenta of the two jets with the largest p_{T}.

Jets that are b-tagged as referenced in section 3.2.2 are referred to as b-jets. For these, the p_{T} threshold is lowered to $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and the η range tightened to $|\eta| < 2.4$.

4.2 Event simulation

An essential ingredient to the model of this analysis is the simulation of events. Simulating the hard proton-proton scattering for certain physical processes, subsequently the energy deposits in the detector, trigger behaviour and selection allows to model the typically non-trivial kinematic distributions for this physical process after the selection. Such simulations have their limitations, which is why a major part of the model is derived from actually observed data or from hybrid approaches in this analysis. Those methods are discussed in the next section. However, simulated events represent part of the model, in particular the signal model, and also serve as an ingredient to some of the data driven methods.

The hard scattering process is simulated with Monte Carlo event generators like PYTHIA 8.2 [63], MadGraph 5 (MG5) aMC@NLO [64] and POWHEG [65]. A list of the simulated processes and the used generators is given in table 4.3. The event simulations are complemented with a simulation of hadronisation, parton showering and underlying event using PYTHIA 8.2. This includes the simulation of pileup events, where the number of pileup vertices is drawn from a Poisson distribution with an expectation value matching the observed data. The events are later on re-weighted according to the observed pileup profile [62].

The interactions of the produced particles with the detector and resulting detector signals are simulated with GEANT4 [66]. These are passed on to the same trigger and reconstruction algorithms as used for observed data in order to get a realistic model that reflects not only the physics but also the detector acceptance and resolution. The simulation is further tuned with corrections, which is going to be described in section 4.4.

Since this analysis targets differential measurements, a further measure is taken to improve the modelling of the ggH process at the corresponding level of granularity. Based on POWHEG NNLOPS [67], the cross sections of the single STXS bins are rescaled to NNLO precision in QCD.

Table 4.3: List of simulated processes and used event generators. The first column indicates the process as treated as entity in the analysis. Different datasets as specified in the second column may belong to the same entity.

Process	Subprocess	Generator	Precision (different precision of inclusive cross section if applicable)
ggH	ggH	POWHEG	NLO (N³LO) QCD
qqH	VBF	POWHEG	NLO (NNLO) QCD
	VH ($V \to$ hadrons)	POWHEG	NLO (NNLO) QCD
VH$_{lep}$	VH$_{lep}$	POWHEG	NLO (NNLO) QCD
ttH	ttH	POWHEG	NLO QCD
		MG5 aMC@NLO	NLO QCD
$Z \to ll$	Drell-Yan $Z \to ll$	MG5	LO (NNLO) QCD, NLO EW
	Electroweak $Z \to ll$	MG5	LO QCD, NLO EW
$t\bar{t}$	$t\bar{t}$	POWHEG	NLO QCD
di-boson	WW, WZ, ZZ	MG5 aMC@NLO	NLO QCD
	Single t	POWHEG	NLO QCD
W+jets	$q\bar{q}' \to W$	MG5	LO (NNLO) QCD, NLO EW
	Electroweak W	MG5	LO QCD, NLO EW

The unavailability of bbH datasets is compensated by increasing the cross section of the ggH sample by 1 %. Differences in the kinematic distributions are negligible at the given rate of this process. Also the ttH has shown to impact the results of this analysis by less than 1 % but is taken into account for completeness.

4.3 Data driven modelling

In this analysis, between 75 % and 95 % of the background is modelled with data driven techniques, depending on the di-τ decay channel. This means that observed events that do not enter the analysis selection are used to deduce the model within the selection. Hence, conditions from the beam, pileup, detector performance directly originate from the observed data and are thus perfectly modelled. The three data driven methods that are used in this analysis are explained in the following subsections. They are used to model events with genuine τ candidates or with certain types of misidentification. Such events are excluded from the simulated datasets based on the truth information about the underlying objects, which is available for simulated events.

4.3.1 Tau embedding

This technique is used to model background events, where two genuine τ leptons are correctly identified as the di-τ pair. Such background mostly consists of $Z \to \tau\tau$ events, but there are further minor contributions from $t\bar{t}$ or di-boson events with the top quarks or bosons decaying into τ leptons.
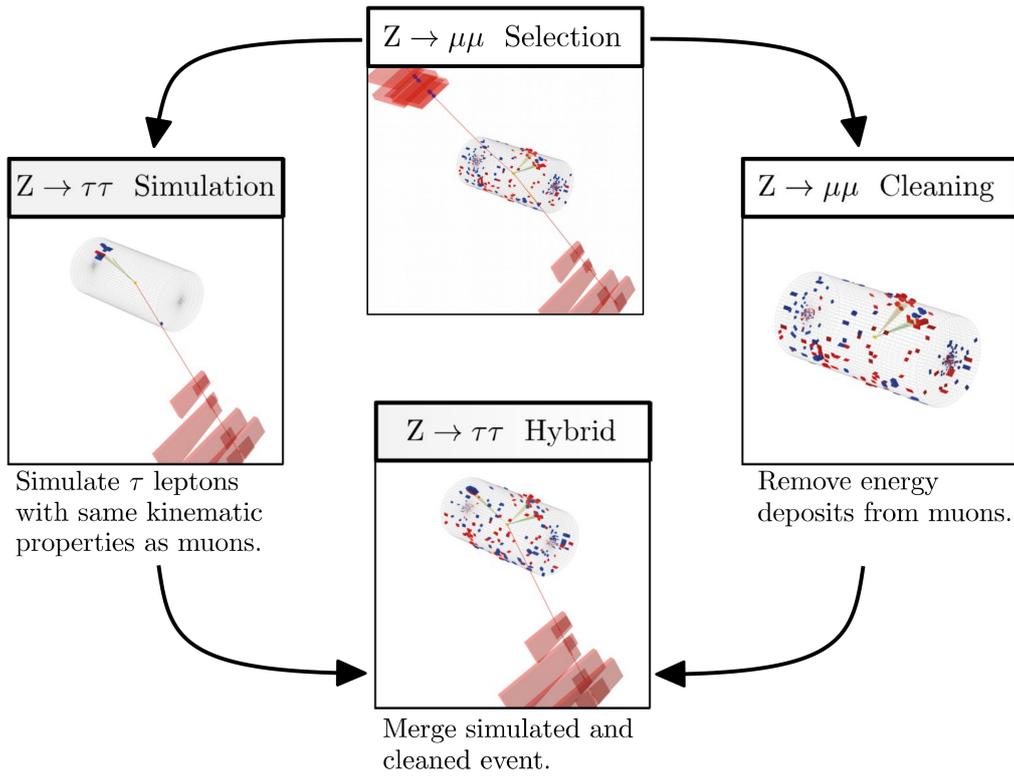
Figure 4.1: Illustration of the τ-embedding technique (from [68])

The τ-embedding technique is a hybrid approach using observed events and replacing part of each event by simulation. The procedure is illustrated in figure 4.1. A detailed documentation of the technique is given in [68]. Based on lepton universality, observed events with a pair of muons, which can be well identified and measured, are taken and the muons are replaced by simulated τ leptons.

The hits associated with the muon tracks as well as the energy deposits in the calorimeter cells are removed. The reconstruction is rerun on the remaining event record to obtain the modified calorimeter clusters and tracks. In parallel, decays of τ leptons with the same kinematic properties as the previously selected muons are simulated. The reconstruction of tracks and calorimeter clusters happens in an otherwise empty detector. Afterwards, the tracks and calorimeter clusters from the cleaned event and the simulated τ pair are merged into a hybrid event. The reconstruction of the physics objects is finally run on the hybrid event.

The resulting events do only rely on simulation in terms of the τ decay and the interaction of its decay products with the detector. All other objects in the events like jets originate from real data. Even the hard scattering, in which the muon pair, or τ pair after the replacement, is produced, occurred in reality. This includes recoiling objects. No theoretical uncertainties on the cross section or p_T spectrum need to be taken into account, instead only experimental uncertainties, e.g. from the measured trigger efficiencies. The selected di-muon events originate from the same data-taking and therefore have the same luminosity, pileup and detector conditions as the data to be modelled. Also the modelling of $E\!\!\!/_T$ is improved.

4.3.2 Fake rate method

The fake rate method (F_F method) is an extrapolation technique that is used to estimate the background due to jets that are induced by quarks or gluons and misidentified as τ_h. The method is applied in the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channels. It was established in [69] and used in other analyses of $H \to \tau\tau$ events like [32][70].

The basic idea is to model this background from a sideband region where it is very pure and where it has the same kinematic distributions apart from some corrections to be taken into account. This sideband is referred to as application region. The event selection of the application region is aligned with the signal region, i.e. analysis selection, but the requirement on the τ_h ID vs. jets is inverted. It comprises events that pass a looser working point of the DeepTau ID, but not the tight working point used for the signal region. The τ_h candidates in the application region are therefore less isolated, i.e. more jet like, and the purity of background from jets induced by quarks or gluons is strongly enhanced.

In order to generate the histograms needed as model in the signal region, they are filled with observed data in the application region. The minor contribution from events without misidentified jets is modelled with simulated data and τ-embedding and then subtracted from the observed data. Leading uncertainties related to the subtracted processes and the statistical uncertainty from the limited number of events are kept track of and propagated to the statistical inference. Particularly the latter is relevant. The subtraction leads to
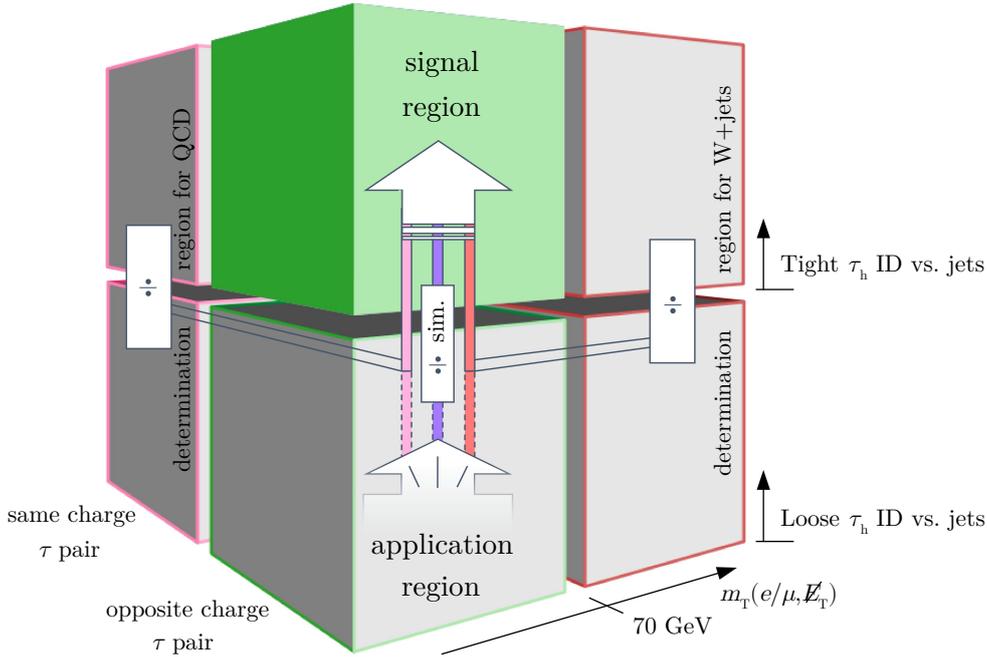
Figure 4.2: Illustration of the F_F extrapolation method: Events from the application region are used to model the jet-induced background in the signal region. A combined extrapolation factor is composed from three factors dedicated to the QCD multijet, W+jets and $t\bar{t}$ background according to the share of these processes in the application region. The QCD multijet factor is derived in a sideband region with τ pairs of same charge. The W+jets factor is derived in a sideband region with large transverse mass. The $t\bar{t}$ factor is derived from simulated data.

larger statistical fluctuations such that a simple Poisson uncertainty according to the resulting yield is insufficient. This is confirmed by the model validation based on goodness of fit (GoF) tests.

The events are weighted with extrapolation factors in order to account for differences between the application and the signal region. This is illustrated in figure 4.2. First of all, the total yield is expected to differ because it depends on the definitions of the DeepTau ID working points. Furthermore, different physical processes contribute, which has to be taken into account. Since different mechanisms of jet production are involved at different rates in these processes, the ratio of isolated to non-isolated jets is different. The following physical processes are distinguished:

- QCD multijet production

- $t\bar{t}$ production and associated jets - only relevant in $e\tau_{\mathrm{h}}$, $\mu\tau_{\mathrm{h}}$; not taken into account in the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ channel

- W+jets - only relevant in $e\tau_h$, $\mu\tau_h$; not taken into account in the $\tau_h\tau_h$ channel

For each background, dedicated extrapolations are derived in sidebands, the so-called determination regions, which mirror the signal and application regions, but contain a pure sample of the considered background. For the QCD multijet background, the determination region is defined by requiring same charge for the two τ candidates instead of opposite charge. For the W+jets background, the requirement on the transverse mass of the lepton and $\not{E}_T$ is inverted. This quantity is a good estimate for the mass of the W boson as it decays to the lepton and a neutrino. Related to that, the process dominates the region at $m_T(e/\mu, \not{E}_T) > 70\,\text{GeV}$. For $t\bar{t}$, no suitable determination region is available and the extrapolation factor is instead derived from simulation. Finally, a combined extrapolation factor is applied to the events from the application region as a weighted mean of the process specific extrapolation factors according to the fractions of the processes in the application region. These fractions are determined from simulation, which is done differentially in the number of jets and either $m_T(e/\mu, \not{E}_T)$ in the $e\tau_h$ and $\mu\tau_h$ channels or the invariant mass of the τ pair in the $\tau_h\tau_h$ channel. Due to the lack of proper simulation for QCD multijet background, its fraction is determined from observed data and by subtracting all other contributions from both simulation and τ-embedding. Therefore this part does not exclusively comprise QCD multijet background, but also further minor processes that are not among the three listed above.

Going one level deeper into detail, also the extrapolation factors are determined differentially in certain quantities. For three bins of the number of jets $[0, 1, \geq 2]^1$, which aligns with the STXS scheme, the extrapolation factor is measured as a function of $p_T(\tau_h)$. In the $\tau_h\tau_h$ channel, this is done for each τ_h and the mean extrapolation factor is applied in the end. The ID requirement of the τ_h, which represents its isolation to a large extend, is not independent from $p_T(\tau_h)$. Therefore, this differentiation is necessary. Non-closure correction factors C, differential in p_T of the other τ candidate, the invariant mass of the τ pair m_{vis}, $m_T(e/\mu, \not{E}_T)$ or I_{rel} of the electron or muon are applied depending on the considered physical process and di-τ decay channel:

- $t\bar{t}$: $\qquad\qquad F_F^{t\bar{t}} = F_{\text{raw}}^{t\bar{t}}\left(\#_{\text{jets}}, p_T(\tau_h)\right) \cdot C\left(m_{\text{vis}}\right)$

- W+jets: $\quad F_F^{W+\text{jets}} = F_{\text{raw}}^{W+\text{jets}}\left(\#_{\text{jets}}, p_T(\tau_h)\right) \cdot C\left(p_T(e/\mu)\right) \cdot C\left(m_T(e/\mu, \not{E}_T)\right)$

- QCD multijet, $e\tau_h$ & $\mu\tau_h$: $F_F^{\text{QCD},e\tau_h\&\mu\tau_h} = F_{\text{raw}}^{\text{QCD}}\left(\#_{\text{jets}}, p_T(\tau_h)\right)$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdot C\left(m_{\text{vis}}\right) \cdot C\left(I_{\text{rel}}(e/\mu)\right)$

- QCD multijet, $\tau_h\tau_h$: $\qquad F_F^{\text{QCD},\tau_h\tau_h} = F_{\text{raw}}^{\text{QCD}}\left(\#_{\text{jets}}, p_T(\tau_{h,1})\right) \cdot C\left(m_{\text{vis}}\right) \cdot C\left(p_T(\tau_{h,2})\right)$

Within the determination regions, the yield of the considered process is derived by subtracting simulation and τ-embedding of other processes from the observed data. Since there is no adequate simulation of QCD multijet background, its contribution to the W+jets determination region is estimated via another extrapolation from the aligned sideband with τ pairs of same charge.

---

[1] It is only two bins for the $t\bar{t}$ extrapolation factor, i.e. $[\leq 1, \geq 2]$ because there are hardly any $t\bar{t}$ events without jets.

4.3.3 QCD multijet extrapolation

The QCD multijet production is not well modelled by simulation. In the channels with hadronically decaying τ leptons, it is modelled by the F_F method. But it constitutes part of the background in the $e\mu$ channel as well. Leptons from decays of heavy flavour jets can be misidentified as originating from a τ pair. Regarding the electron candidate, a light flavour jet may cause an electron-like shower in the detector as well. The QCD multijet background is modelled in the $e\mu$ channel with an extrapolation method similar to the F_F method but simpler as it does not involve process specific extrapolation factors. It has been used in multiple analyses already, e.g. [32].

If the two τ candidates are required to carry same charge, this results in a phase space region dominated by the QCD multijet background. This region is used as determination region in the F_F method. Here, it serves as the application region. The extrapolation factor is derived from a sideband where the isolation of the muon is relaxed, i.e. $0.2 < I_{\mathrm{rel}}(\mu) < 0.5$. It is taken differentially in three bins of the number of jets $[0, 1, \geq 2]$ and depending on $\Delta R(e, \mu)$ since the ratio of light and heavy flavour jets varies depending on $\Delta R(e, \mu)$. Two non-closure corrections are applied and they depend on the p_{T} of both τ candidates. One of them is derived from the residual differences in the sideband that become visible with resolved p_{T}. The other one accounts for the dependency of the extrapolation factor on the muon isolation and is derived as a ratio of extrapolation factors measured in parallel in the sidebands with $0.2 < I_{\mathrm{rel}}(e) < 0.5$ and either $I_{\mathrm{rel}}(\mu) < 0.2$ or $0.2 < I_{\mathrm{rel}}(\mu) < 0.5$. The total extrapolation factor computes as

$$F^{\mathrm{QCD}} = F_{\mathrm{raw}}^{\mathrm{QCD}} \left(\#_{\mathrm{jets}}, \Delta R(e, \mu)\right) \cdot C_{\mathrm{non-closure}} \left(p_{\mathrm{T}}(e), p_{\mathrm{T}}(\mu)\right)$$
$$\cdot C_{\mathrm{non-iso \to iso}} \left(p_{\mathrm{T}}(e), p_{\mathrm{T}}(\mu)\right). \quad (4.3)$$

Like in the F_F method, histograms as needed for the analysis are generated with observed data in the application region. Corresponding histograms representing the minor contributions from other processes are generated with simulated and τ-embedded data and subtracted from the observed data. All events entering this calculation are weighted by the differential extrapolation factor.

4.4 Corrections

Several corrections are applied to the model on top of the event reconstruction in order to improve the description of the observed data. These primarily concern the fully simulated events and the simulated parts of τ-embedded events, where simulation cannot include all details of the detector performance for example. Indirectly, these corrections have an influence on the extrapolation methods via the subtraction of contributions modelled by simulation or τ-embedding.

The applied corrections are summarised in the following. They comprise corrections of selection efficiencies, mostly by the application of weights to the events, and corrections of energy scales, which change momenta of the concerned objects and depending quantities, in particular $\not{E}_\mathrm{T}$. The corrections are taken into account in the event selection described in section 4.2.

- Efficiencies of the identification, isolation and trigger requirements on electrons and muons are corrected in both simulated and τ-embedded samples. The efficiencies are measured for simulated, τ-embedded and observed data respectively with the tag-and-probe method. This is shown in [12][71]. The respective ratio of observed data to simulation or τ-embedding is applied as a weight to the events. These corrections are differential in p_T and $|\eta|$ of the concerned electrons or muons.

- Efficiencies of the identification and trigger requirements on τ_h candidates are corrected in both simulated and τ-embedded samples. The correction factors are differential in $p_\mathrm{T}(\tau_\mathrm{h})$. The τ_h decay mode is taken into account where necessary. For electrons and muons misidentified as τ_h, independent corrections are applied respectively. These are differential in $|\eta|$ such that the bins correspond to the different segments of the electromagnetic calorimeter or muon chambers depending on the type of the misidentified object.

- Tracking efficiencies are corrected in τ-embedded samples in order to take into account differences due to the track reconstruction in the empty detector and migration between decay modes due to remnants of the original muon signature.

- The energy scale of electrons is corrected in simulated and τ-embedded samples. In simulated samples, this comprises a correction of the resolution as well, which is of minor impact for this analysis.

- The energy scale of τ_h is corrected in simulated and τ-embedded samples. This is done separately for each decay mode in order to take reconstruction effects of the different involved particles into account.
  Dedicated energy scale corrections are applied to electrons or muons misidentified as τ_h. A distinction is made between the decay modes with a single charged hadron and with or without a π^0. Decay modes with more particles are hardly reconstructed from misidentified electrons or muons.

- The energy scale and resolution of jets is corrected in simulated samples. In τ-embedded samples, this is not necessary since the jets originate from real data.

- The efficiency of the *b*-tagging of jets is corrected in simulated samples. This is not done with weights since process related correlations between the presence of *b*-jets and the kinematics of leptons selected as the τ pair exist. Instead, the result of the *b*-tagging is altered for some randomly chosen jets in order to reach the correct rate. This is done independently for the identification rate of *b*-jets and the misidentification rate of all other jets based on truth information from the simulation.

- The recoil in events with heavy resonances and its impact on $\not{E}_\mathrm{T}$ is measured with a selection of high purity of $Z \to \mu\mu$ events [62] and used to correct $\not{E}_\mathrm{T}$ of simulated events with a massive boson involved. The correction is differential in the number of jets in the event.

- The $\not{E}_\mathrm{T}$ scale of τ-embedded events without jets is corrected for the influence of the simulated part on the PUPPI approach. In τ-embedded events with jets, it is already well described.

- Simulated $Z \to ll$ events are re-weighted depending on the truth information about the p_T and the invariant mass of the simulated di-τ system. The weights are derived from a comparison to data in a selection of high purity of $Z \to \mu\mu$ events.

- Simulated events of $t\bar{t}$ production are re-weighted in order to improve the modelling of the p_T spectrum of the top quarks according to an improved theoretical prediction.

- In 2016 and 2017, the CMS detector had a trigger inefficiency that was due to a deterioration of calorimeter crystals in the forward region leading to a timing shift. Events with jets or photons in that region could cause the triggering of the previous event, which vetoes the triggering of the present event in the current system of data acquisition. A weight depending on the event topology is applied to simulated events in order to account for the reduced probability of being triggered due to this effect. The VBF process is impacted relatively strongly by this effect and its selection efficiency is reduced by about $4\,\%$ on average.

Machine learning based event classification

The phase space of the final state is going to be treated differentially by subdividing it into event categories and histogram bins[1], for two reasons: Firstly, the signal to background ratio varies with certain kinematic quantities and splitting by kinematic regions yields the possibility of profiting from the different ratios and related sensitivity levels. Secondly, the contributions from different processes and the related impact of systematic variations vary across the phase space as well and the splitting supports the statistical analysis in keeping such variations under control. The first one is the primary purpose and, as long as the resulting measurement is not dominated by systematic uncertainties, it is basically enough and optimal to distinguish between different phase space regions only according to the probability that a given event is signal or background. This is also the typical use case for machine learning based classification, in particular because the signal-background composition depends on several kinematic quantities in a non-trivial way due to the different signal and background types and reconstruction effects. With machine-learning based multi-classification, it is even possible to distinguish between more than two classes. This is very useful for this analysis because it is targeting the STXS measurement with more than just one signal class. Having introduced one class for each signal, this can also be done for the different background processes. This way, the modelling of the backgrounds can be better controlled, particularly addressing the second point mentioned above.

The first application of multi-classification based on neural networks in an analysis of H $\rightarrow \tau\tau$ events was developed in [72] for a measurement of inclusive signal as well as ggH and VBF. The classification presented in this chapter is based on this approach. It is extended to fit the target of STXS stage-1 measurements. An intermediate step with a sub-categorisation via simple cuts was published in [32], which is a preliminary version of this analysis. The approach for the more inclusive measurements is improved with respect to [32] as well. The following section focusses on basic aspects of multi-class neural networks and the technical setup used for this thesis. Afterwards the concrete event categorisation that is implemented with the neural networks is discussed in detail and in relation to the envisaged measurement.

---

[1]Technically there is no difference between these two types of phase space splitting. However, categories additionally represent logical units associated with a specific purpose.
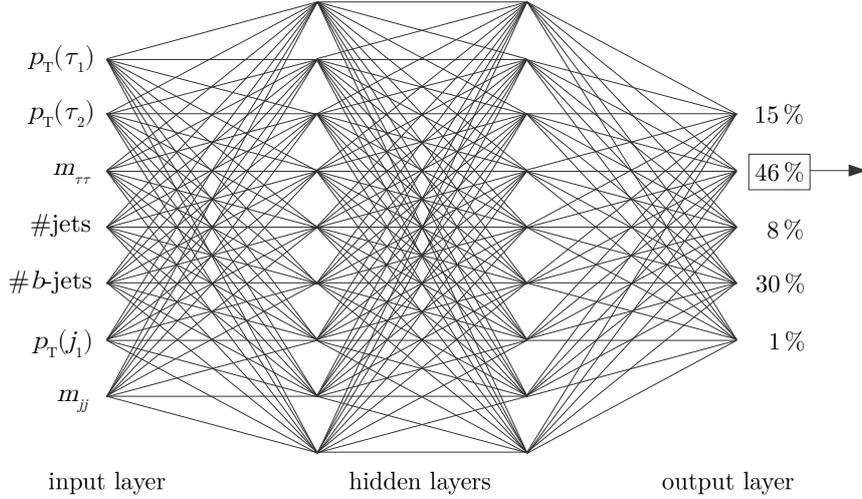
Figure 5.1: Illustration of a multi-class neural network with seven input quantities, two hidden layers with nine nodes each, and five output nodes. The softmax function applied to the output layer makes the outputs sum up to one and allows for a Bayesian probability interpretation. An exemplary configuration of input quantities and output values is given. For example the largest output value of 46 % could be used to determine the category for this sample.

5.1 Multi-class neural networks and technical setup

The expression "multi-class" refers to the existence of multiple output nodes of the neural network, where each node generates a neural network output value (NN score). They can be trained to address multi-dimensional problems like the separation of not only two but multiple event classes. An exemplary architecture of a multi-class neural network is illustrated in figure 5.1. Using a softmax activation for the output layer, which ensures that the output nodes sum up to one, and cross entropy as loss function, the neural network is trained to indicate the process composition in the phase space region defined by the input quantities. In the given example, the sample most likely belongs to the class that is associated with the second node with a value of 46 % and would be assigned the corresponding event category.[2] Within each category, the largest NN score is used for further discrimination based on the purity of the entering events. A histogram is filled resulting in a distribution with increasing purity of the process associated with the respective category.

The utilised neural networks are fully connected feed-forward networks with two hidden layers and 200 nodes each. The technical parameters of the trainings are listed in table 5.1.

---

[2]There are different ways how the output values can be used for the category definition. This will become an issue in the following sections. Sending an event to the category with the highest score is a straight forward method making sure that events are not double-counted in the final statistical inference.

Table 5.1: Configuration of the neural networks and training

Type	feed-forward, fully connected layers
Number of output nodes	5 to 20 (see section 5.3)
Number of hidden layers	2
Number of nodes per hidden layer	200
Activation function	hyperbolic tanget (+softmax for output layer)
Input preprocessing	zero mean, unit standard deviation
Weight initialisation	uniform Glorot
Batch size	30 per class and year of data-taking
Validation split	25 %
Optimiser	Adam (learning rate: 10^{-4})
Descent method	mini-batch
Loss function	cross entropy
Regularisation	dropout (30 % per layer), L2 (10^{-5})
Early stopping	after 50 epochs without improvement

A more detailed explanation can be found in e.g. [73]. The training is performed on the same data samples that are also used for the model in the statistical inference. For the processes that are modelled with the QCD-extrapolation or F_F method, observed data events from the respective sideband with extrapolation weights applied are used in the training. While the exact model of these backgrounds foresees the subtraction of minor contributions from other backgrounds based on histograms, this is neglected for the training since this requires single events as inputs. This procedure is sufficient to achieve a high purity of these backgrounds in the respective events categories.

A two-fold training is used in order to guarantee the applicability of the training, to the observed data in particular. This implies that the whole dataset is split into two folds. On each half, a neural network is trained and used to classify the other half. This way, there is no event in the final model that obtains its NN score from a neural network that was trained on the same event. This detail is however not relevant for the other aspects of the classification and such a pair of neural networks is therefore referred to as a single classifier in the following.

The batches of events that the neural network is trained on are created such that a fixed number of 30 events is taken from each class in each year of data-taking. Each epoch processes a fixed number of 1000 batches. On one hand, this is a technical necessity in order to cope with the very different numbers of available events per class. Random selection of events from the collection of all classes together would result in the dominance of τ-embedded events, which outnumber the simulated events. This was investigated in [73]. On the other hand, this means that apart from event weights, which for example reflect trigger efficiencies and are at the order of 1, all classes enter the training at equal prevalence. For the neural net, the total event numbers of all classes appear to be the same. This is intended for the following reason:

For each class, there is a phase space region where the class seems to be dominant to the neural net. This in turn leads to a population of the corresponding category.
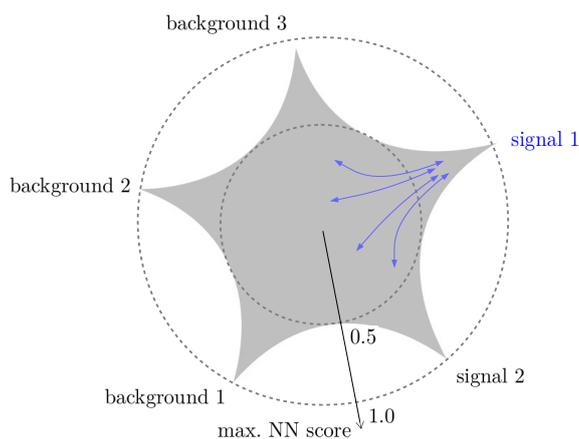
Figure 5.2: Qualitative sketch of the NN score space and the position of events depending on the process purity and normalisation: The shown star-like shape depicts the actually (5-1)-dimensional output space of a multi-class neural network with five classes. If a NN score is larger than 0.5 it is necessarily the largest and events are assigned to the corresponding category for sure. Below a value of 0.5, the assignment depends on how the other NN scores are distributed. As highlighted for signal 1, events of this class with low purity are spread over all categories. Only if a class is represented in the training at a sufficiently large rate to be perceived as pure in certain regions of the phase space, the corresponding events will aggregate in the category at a high NN score.

The efficiency of assigning the correct category will be of the same order for all classes. For a one-dimensional discriminator, this is not very important because the events are ordered by purity along the one-dimensional output scale and a different prevalence of the classes during the training will lead to a consistent shift of events along this scale without changing the order. For the multi-class output however, it makes a difference, which is illustrated in figure 5.2. Within a category, the same argument about the order holds partially and therefore it is also acceptable to have higher prevalence of the signal in the training than in reality. But if events were shifted too much to a smaller NN score such that other NN scores would become larger, they would start to spread over the other categories. This is not desirable for the signal events in particular, which in reality do not dominate the background even in the purest phase space regions. The chosen method for creating the training batches ensures the population of the signal categories in this analysis.

5.2 Input quantities

The input quantities of the classifiers must meet two requirements, i.e. improving the separation of the processes and being well modelled. A quantity that yields no separation power does not directly harm the training or the classification result but it increases complexity and required computation power. On the other hand, a quantity that yields

separation power is only useful if its modelling is under control. Otherwise, it potentially deteriorates the modelling of the NN score and biases the measurement. Detailed studies were carried out to identify a set of observables from a larger superset that is as small as possible but does not result in a reduced sensitivity. This is documented in [74] and briefly summarised in the following.

At first, the candidate observables are sorted by their importance for the classification task. This is done with a Taylor expansion technique of the neural networks [75] determining the impact of input quantities on a given output node. The sorting defines in which order input quantities are subsequently removed from the full set for further investigation. This significantly reduces the number of investigated subsets, compared to all possible subsets, and keeps the effort at an affordable level. The performance measured via the F1 score [74] of the resulting classification is monitored for all classes. Variables that do not clearly lead to an improvement are dropped.

As outlined in the previous chapter, the analysis is conducted in four decay channels of the di-tau pair. Due to the differing types of tau candidates, related selection requirements and reducible backgrounds, dedicated classifiers are trained for each channel. Accordingly, the pruning procedure of input quantities is done for each channel but with the objective to identify a commonly usable set. The final set consists of 15 quantities, which are mostly the same for all channels with one exception. In first place, these are kinematic quantities related to the two τ-candidates, i.e. $p_{\mathrm{T}}(\tau_1)$, $p_{\mathrm{T}}(\tau_2)$, $p_{\mathrm{T}}(\tau_1, \tau_2)$, their invariant mass (visible di-τ mass), angular separation ΔR and the likelihood based estimate of the invariant di-τ mass $(m_{\tau\tau})$ [76] taking into account $\not{E}_{\mathrm{T}}$. These variables are supplemented by jet quantities, which are relevant for the identification of qqH and some of the background processes. These are the number of jets, number of b-tagged jets and kinematic properties of the two jets with the largest transverse momentum $p_{\mathrm{T}}(j_1)$, $p_{\mathrm{T}}(j_2)$, $p_{\mathrm{T}}(j_1, j_2)$, m_{jj}, $\Delta\eta(j_1, j_2)$. Furthermore, two quantities are included which are derived from a Matrix Element Likelihood Analysis (MELA) algorithm [77] exploiting information from the τ-candidates and the two leading jets. They characterise the energy transfer via the vector bosons under the VBF hypothesis. In the $e\mu$ channel, the number of b-tagged jets is obsolete since a b-veto is imposed on the initial event selection. But in addition, the transverse mass of the di-tau system and $\not{E}_{\mathrm{T}}$, i.e. $m_{\mathrm{T}}(e\mu, \not{E}_{\mathrm{T}})$ is used.

Control plots of the input quantities are shown in appendix A. The visual inspection of these distributions is complemented by the objective quantitative measure of GoF tests for validating the modelling. Likelihood based saturated GoF tests [78], which take into account systematic uncertainties in the model, are used for this and applied at two stages. Firstly, the single distributions are tested at a granularity of ten equipopulated bins capturing all essential features at the given level of statistical precision of the measurement. Secondly, two-dimensional distributions with 5×5 bins are tested, going through all pairwise combinations of the quantities. This is meant to not just validate the marginal distributions but also their pairwise correlations since the neural network explicitly exploits these. Altogether, the number of tests is 120 per channel and year for the given number of input quantities. The number grows quadratically with the number of input quantities, which illustrates the motivation for the input pruning mentioned above. At this amount, a few tests are expected to result in low p-values even with a
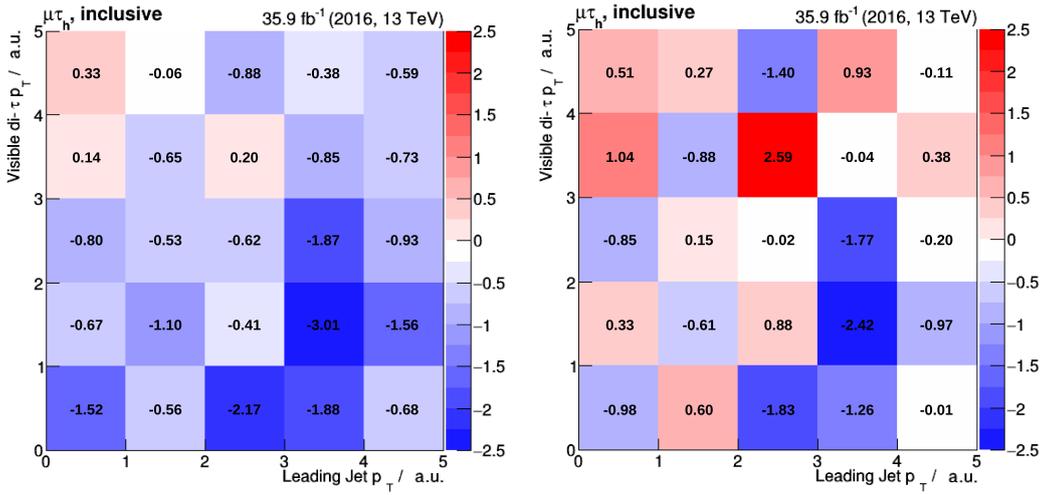
Figure 5.3: Exemplary inspection of the 2D residuals of two input quantities: The differences between the model and observed data normalised to the uncertainty in each bin are shown depending on the p_T of the visible di-τ-system and the p_T of the jet with largest p_T. Left: Residuals before running a fit. Right: Residuals after fitting the model to the data based on the uncertainty model detailed in chapter 6. Note that the fit imposes additional constraints on the uncertainties resulting in a smaller denominator of the residuals.

perfect model. Hence, low p-values do not necessarily imply a problem of the model. Among the 1D tests, 1.1 % result in a p-value below 5 %. Among the 2D tests, it is 9.8 %. Keeping in mind that the involved quantities are not fully independent, these fractions fit the assumption of a proper model. 2D distributions with low p-values and identified to be of major importance to the neural network are further inspected manually and checked for homogeneous deviations that would result in different correlations in data and corresponding the model. An example of such an inspection is shown in figure 5.3. For the given combination of quantities, the GoF test results in a p-value of 0.2 %. However, there is no clear global shift in the residuals visible. The tension in the fit is mostly due to two disjoint bins with large residuals visible in the distribution after fitting, i.e. 2.59 and -2.42, which hints at a statistical fluctuation. The correlation of this 2D distribution is 0.391 in observed data and 0.392 in the model both before and after the fit, which does not hint at a problem in the model, either. Given such cross-checks, variables can be included in the training even if some related p-values are small. With this procedure, the proposed set of 15 input quantities is verified as good to go for the training.

In order to harmonise the approach and increase the training statistics, common classifiers are trained on the full dataset of all three years of Run-2. It should still be taken into account that the beam and detector conditions are not the same for all years and particularly the trigger thresholds differ. Therefore, the year is provided as additional input such that the classifiers get the freedom to apply adaptations depending on the year [74]. This is done via a one-hot encoding technique introducing each year as a binary input node and treating the years equally.

Table 5.2: Background categories per channel. The listed classes comprise the following processes: genuine τ = {τ-embedded events}, zll = {$Z \rightarrow ll$ with an e or μ misidentified as τ_{h}}, tt = {fraction of top pair production that is modelled via simulation}, qcd = {jet induced background modelled by the QCD extrapolation method}, jet$\rightarrow \tau_{\mathrm{h}}$ = {jet induced background modelled by the F_F method}, diboson = {fraction of di-boson production that is modelled via simulation}, misc = {remaining backgrounds}

$e\mu$	$e\tau_{\mathrm{h}}$	$\mu\tau_{\mathrm{h}}$	$\tau_{\mathrm{h}}\tau_{\mathrm{h}}$
genuine τ	genuine τ	genuine τ	genuine τ
	zll	zll	
tt	tt	tt	
qcd	jet$\rightarrow \tau_{\mathrm{h}}$	jet$\rightarrow \tau_{\mathrm{h}}$	jet$\rightarrow \tau_{\mathrm{h}}$
diboson			
misc	misc	misc	misc

5.3 Categorisation

The methods previously introduced in this chapter are used to define a process oriented event categorisation taking into account both different signal and background processes.

The composition of the background processes depends on the di-τ decay channel. For this reason the sets of background categories are chosen differently for each channel. For example, leptons from $Z \rightarrow ll$ events misidentified as τ_{h} are a relevant background in the $e\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels. In contrast, they hardly contribute in the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ channel due to the small probability to have two cases of misidentification at the same time. Only the dominant background processes that appear with sufficient rate to populate a distinct event category are taken into account as a separate event class. The remaining background processes are collected in an additional class labelled "misc" in order to still have them taken into account in the training avoiding migration into the signal categories without penalty on the training loss. The chosen sets of background categories per channel are listed in table 5.2.

The treatment of the signal categories is the same for all decay channels but separate approaches are followed depending on the envisaged measurement. Targeting the differential measurement defined by the STXS stage-1 scheme, event classes are introduced for each STXS bin so that the neural networks do not only accomplish the separation of signal and background but also the best possible decorrelation of the single signal fractions. This increases the number of categories significantly. However, this categorisation does not lead to optimal results for the STXS stage-0 and inclusive measurements. The reason for this becomes clear from the distributions shown in figure 5.4. Once for ggH and once for qqH, a scatter plot shows the dependency between the maximum NN score and one of the quantities that define the STXS bins. The colors indicate the true STXS bin to which the events belong. Events that are close to a bin boundary are in general assigned a smaller NN score. This is because neither the reconstruction nor the neural networks can perfectly resolve the truth information that is used to define the bins. Therefore, it is not clear for the neural network to which of the two neighbouring classes the event
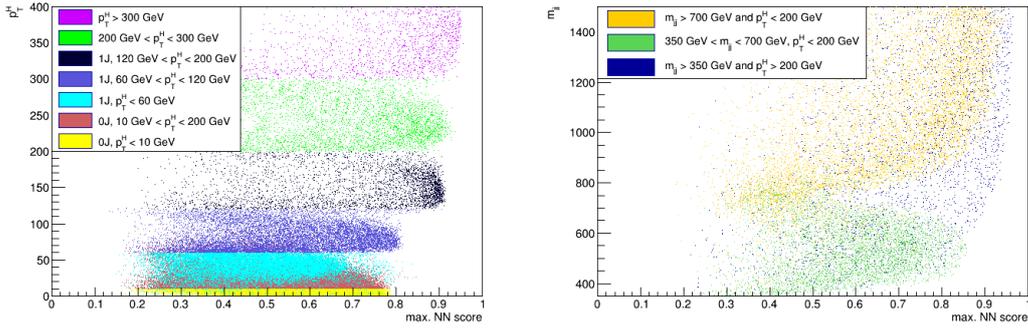
Figure 5.4: Rejection of events close to STXS bin boundaries by the classifier that is trained on stage-1 signal classes. Left: Simulated ggH events that belong to the indicated stage-1 classes and are assigned to any of the ggH categories. The x-axis indicates the maximum NN score, which is usually the one of the respective class. The y-axis indicates the true p_T of the Higgs boson, which defines several STXS bins. The reconstruction and the classifier are not able to fully resolve it. Therefore, events that are located close to a bin boundary are not well classifiable and assigned smaller NN scores. This becomes visible as clear dips in the shown distributions. Right: Corresponding plot for qqH events and classes. On the y-axis, the mass of the di-jet system being important for the STXS bin definitions of qqH is plotted. This time the reconstructed quantity is shown because it is already close to the di-jet mass the classifier can resolve. The green and the orange class overlap and events in the transition region obtain smaller NN scores. The high-p_T class shown in blue does not share this STXS bin boundary and is therefore not affected.

belongs if it is located close to the boundary. The event will be assigned similar NN scores for both classes. Such events do indeed not contribute to the sensitivity of the stage-1 measurement due to this ambiguity. Therefore it is good to separate them from signal events without this ambiguity and put them to smaller NN scores together with events with a larger background contamination. However, for the more inclusive measurements where no difference is made between the stage-1 signal fractions, such events would be valuable and should not be rejected. For this reason, two different event categorisations are used depending on the STXS stage of the measurement as detailed in the following subsections. Since each stage is performed as a separate measurement, there is no need to stick to a common set of event categories.

5.3.1 STXS stage-0

For the measurement of ggH and qqH, trainings are performed with two corresponding signal classes in addition to the background classes listed in table 5.2. Instead of introducing two signal categories with that, a different approach is chosen in order to efficiently run a fit of inclusive signal with the same setup as well. The issue discussed before and motivating the different approaches for stage-1 and more inclusive measurements does in fact impact the measurement of inclusive signal based on a training and categorisation with two signal classes in a similar way. But the smaller dimensionality of the problem

allows for a different approach that is technically feasible and clearly representable and enables an efficient measurement at both levels.

The two signal categories that would have been defined with the given training are merged into a single one, aggregating all signal-like events. Within this category, the information of both NN scores related to the signal is used to disentangle the two signals and the background at the same time. Hypothetically, one could use an n-dimensional histogram according to the n output nodes of the neural network instead of introducing n categories in order to optimally exploit the available information. But this would immediately run into technical problems concerning population with events and traceability. At least a reduced version of this is realised here, creating a 2D distribution of the ggH- and qqH-NN score within a single signal category. For technical and visualisation reasons this distribution is still mapped into a one-dimensional histogram. The mapping is illustrated in figure 5.5. The granularity of the bins is adapted to the expected event distribution and the variation of the signal to background ratio. Some qqH events with their typical di-jet topology appear in a very pure phase space region such that a fine binning is beneficial at high qqH NN score. Events of the ggH process are mostly not as pure and concentrate more at a medium NN score. Still, it is sensitive due to the large amount of events. Events that are clearly identified as ggH or qqH end up in the upper or right corner of the displayed triangle at high values of the respective NN score. Events that are more background-like are located more towards the lower left. Separated from these, events that are not clearly identified as ggH or qqH but still signal-like are located close to the center of the hypotenuse of the triangle.

The resulting distribution for the 2018 $\mu\tau_h$ channel is shown in figure 5.6. Plots for the other categories, channels, and years are provided in appendix B. Note that the presented binning is a template on the basis of which bins are merged independently for each channel and year in order to make sure that every bin is populated with a minimum background expectation of ten events to avoid shortcomings in the statistical model. The dashed lines indicate the vertical slices through the triangle of figure 5.6 such that the bins at the hypotenuse are the most right ones in each slice. These bins happen to have a small event yield but a comparably high signal to background ratio. With two one-dimensional discriminators, these bins would disappear in the bulk of bins 1 to 4. The yielded signal is mainly ggH. This is because ggH is not as much amplified in the training as qqH given its larger natural production cross section. Therefore, this approach is also beneficial for the sensitivity to ggH, improving it by about 15 % and the inclusive sensitivity by about 10 %. The sensitivity to qqH remains basically unchanged with an improvement at percent level.

Figure 5.7 shows the confusion matrix of the stage-0 classifier for the $\mu\tau_h$ channel. Corresponding plots for the other channels are provided in appendix C. For all classes the largest fraction of events ends up with the correct prediction. As lined out before, qqH events can be well identified due to the additional di-jet signature. Events of the ggH process are quite similar to $Z \to \tau\tau$ events, which dominate the genuine τ class. Therefore, the confusion between these two classes is relatively large, which is unavoidable. Also 22 % of ggH are identified as qqH matching the observation made earlier in this section.
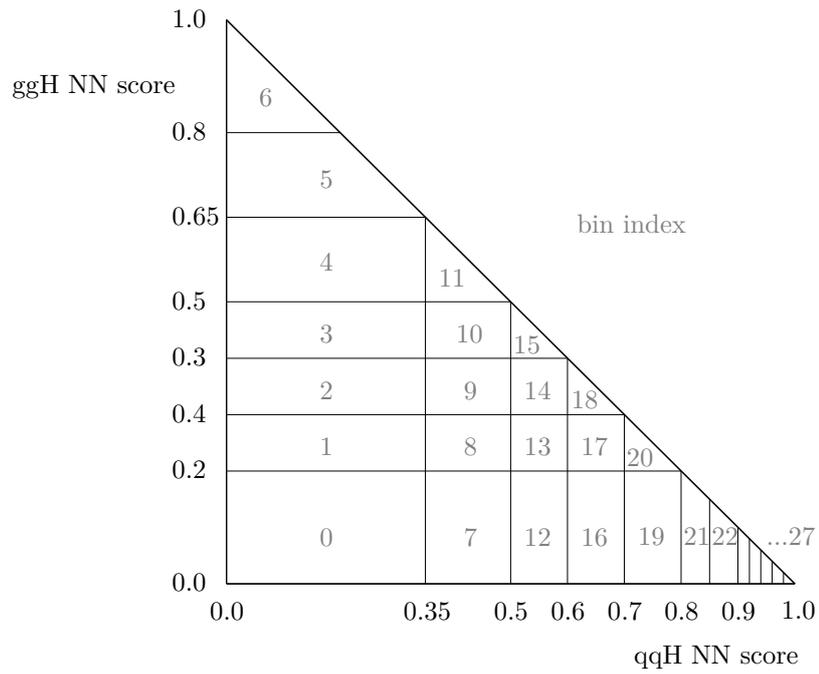
Figure 5.5: Bin mapping of the 2D signal discriminator. The triangle represents the space spanned by the two signal NN scores. It is mapped onto a one-dimensional histogram according to the bin indices indicated in grey.
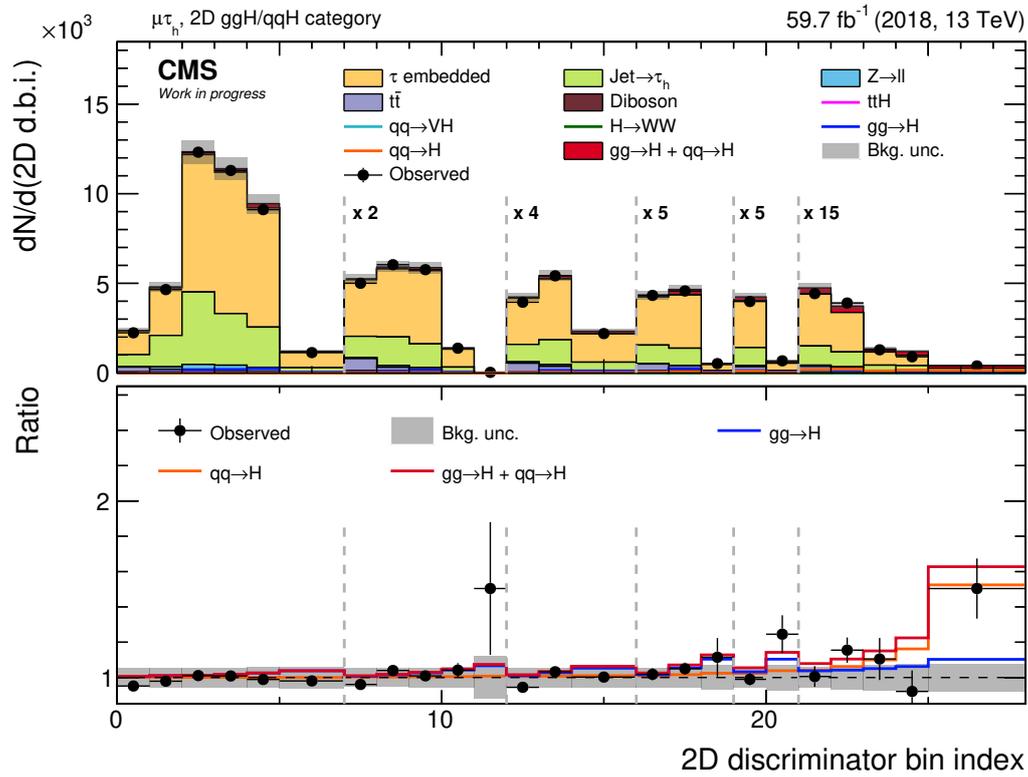
Figure 5.6: 2D signal category in the 2018 $\mu\tau_h$ channel. In the upper panel, slices are scaled up by the indicated factor for visibility reasons. In the lower panel, the observed data and the expected signal excess on top of the background are shown relative to the background prediction. Some bins extend over more than one index of the template given in figure 5.5 as a result of an individual merging of bins within each category ensuring sufficient population.

Figure 5.7: Confusion matrix of the stage-0 $\mu\tau_h$ classifier. The columns are normalised such that given numbers indicate the fraction of a true class directed to a predicted class.

5.3.2 STXS stage-1

For the STXS stage-1 measurement, more signal categories are introduced. Due to sensitivity limitations however, not the full granularity of the stage-1 scheme as described in section 2.3 is envisaged. The reduced STXS schemes used in this analysis are depicted in figures 5.8 and 5.9.

In the ggH scheme, only two classes at $p_T^H > 200$ GeV are introduced. Also the granularity of the classes with at least two jets is reduced because of the concurrence with qqH events. While in the measurement only a single 2-jet class is going to be measured, the categorisation keeps a granularity of four classes for the future combination with other analyses, where these bins will be resolved. In the qqH scheme, the granularity is reduced to three classes with what is referred to as VBF topology, i.e. $m_{jj} > 350$ GeV, and one class containing the remaining events with at least two jets, in particular accumulating the fraction of events originating from VH. The small part of qqH events with less than two jets is not included in the training. In the measurement, its yield is going to be correlated with the latter class due to the lack of sensitivity. Altogether, 15 signal classes and event categories are introduced.

Four examples of resulting event distributions as they are used for the statistical inference are shown in figure 5.10. The remaining distributions in the stage-1 categories can be found in appendix B. The distribution of the signal class associated with the respective categories is shown as a red line. The purity of this signal increases with increasing NN score within each category. This holds with respect to both the background processes and the remaining parts of the ggH and qqH processes.

Figure 5.11 shows the confusion matrix of the classifier in the $\mu\tau_h$ channel as an example. Corresponding plots for the other channels are provided in appendix C. Even with this large amount of classes, a large fraction of events is identified correctly for most of the classes. Only among the ggH and qqH classes with two jets the confusion is large and leads to less than 30 % correctly identified events for two of the classes. This is hardly avoidable because of the large overlap in the observable phase space. Note that compared to stage-0 the signal classes dominate the training more due to the larger number of signal classes. The ggH process is identified more efficiently but in turn more $Z \to \tau\tau$ events from the genuine τ class are assigned to ggH categories.

In this chapter, it was discussed how events are distributed to categories with the help of neural network based multi-classification optimising the separation between all considered physics processes. This is done for the inclusive and stage-0 measurement and for the stage-1 measurement individually, creating a set of categories representing the respective granularity. Within the categories, the associated NN scores are used to further resolve the pure phase space regions of each respective process. In the exceptional case of the stage-0 signals, the two potential signal categories are merged and both associated NN scores are used for further differentiation within the category. The resulting event distributions are the basis on which the statistical inference for measuring the signal rates is applied.
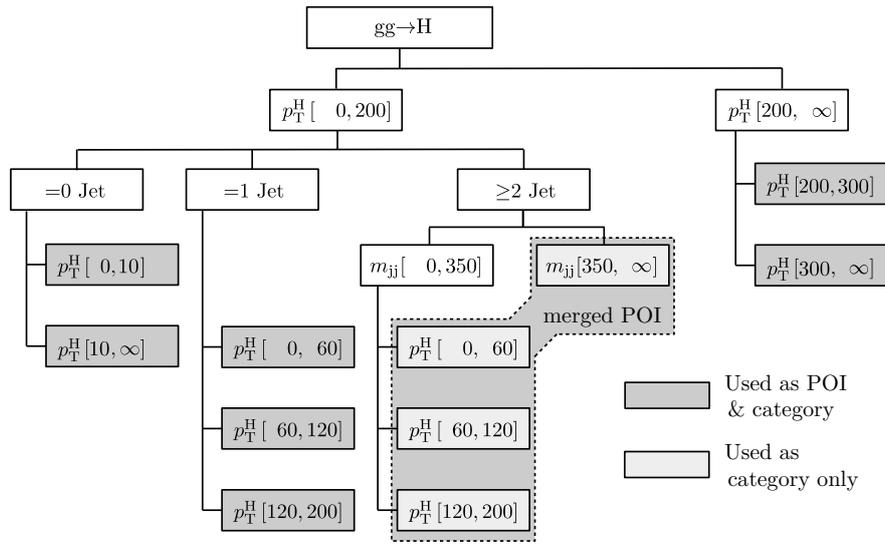
Figure 5.8: Reduced ggH STXS scheme. The granularity of the original scheme shown in figure 2.4 is reduced to fit the sensitivity of this analysis. Processes grouped as "merged POI" are going to be measured as a single entity in the statistical inference scaled by a shared parameter of interest (POI).
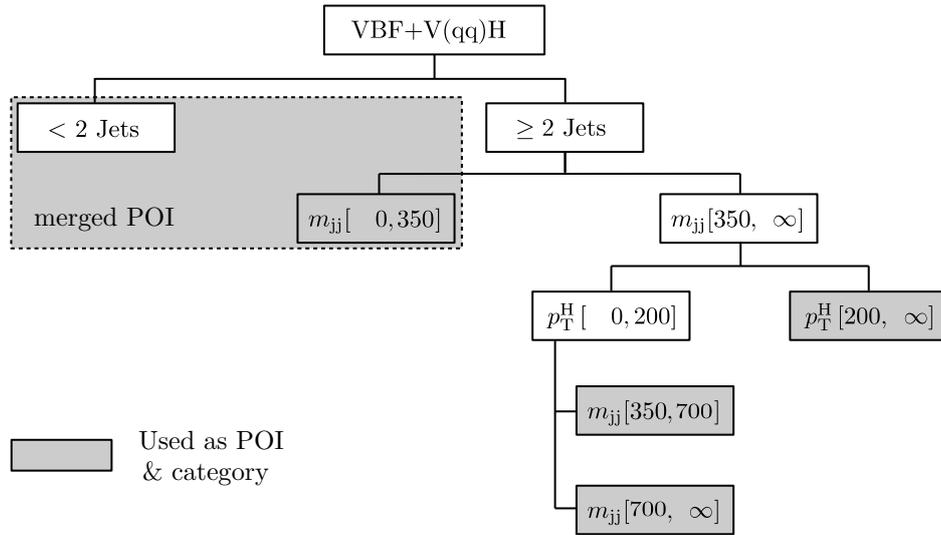
Figure 5.9: Reduced qqH STXS scheme. The granularity of the original scheme shown in figure 2.5 is reduced to fit the sensitivity of this analysis. Processes grouped as "merged POI" are going to be measured as a single entity in the statistical inference scaled by a shared POI.
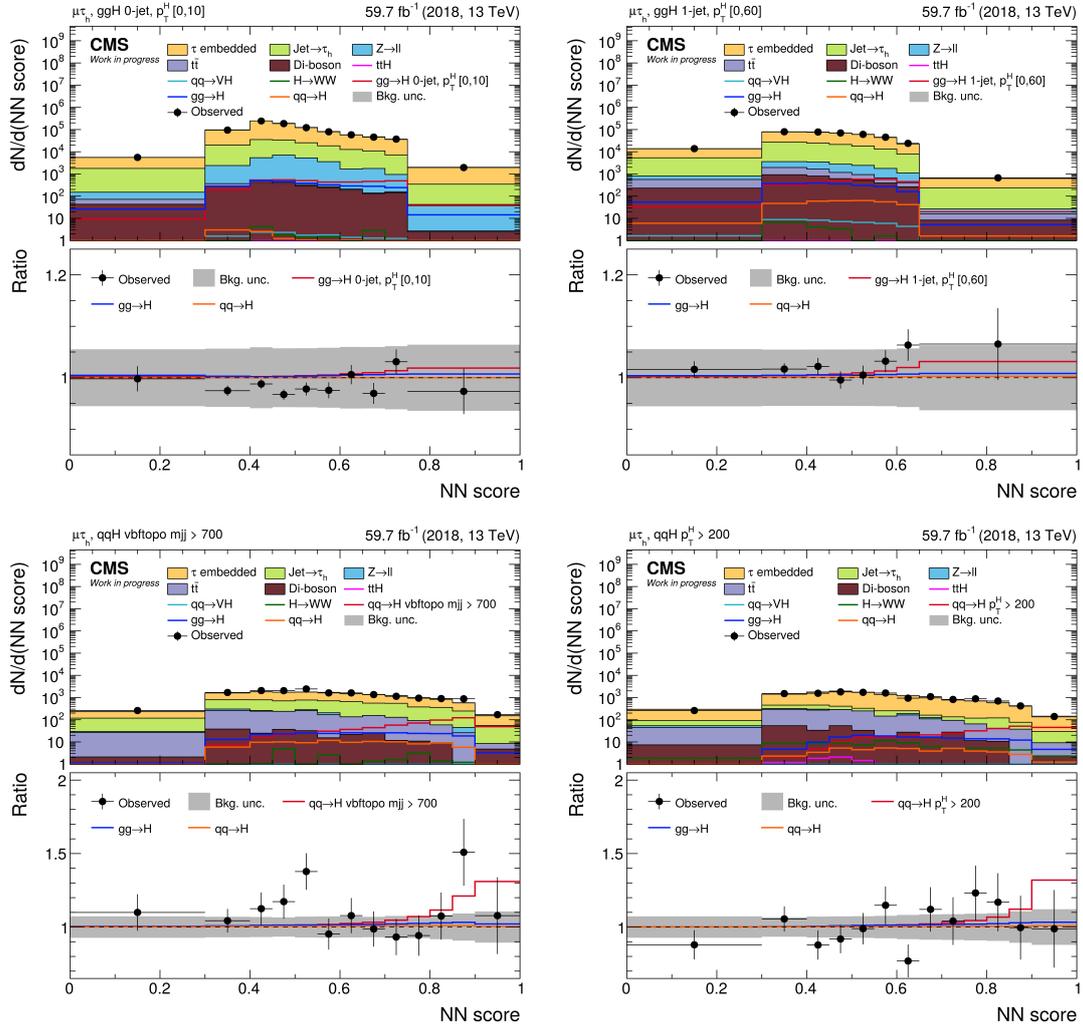
Figure 5.10: STXS stage-1 signal categories in the 2018 $\mu\tau_\mathrm{h}$ channel. Four of the 15 signal categories with the respective prefit distributions of the NN score are shown as examples. The category names are indicated on the upper left of each plot. The signal component that is associated with the respective category is shown in red. Remaining ggH and qqH contributions are shown as blue and orange lines. In the ratio panels, the observed data and the expected signal excess on top of the background are shown relative to the background prediction.
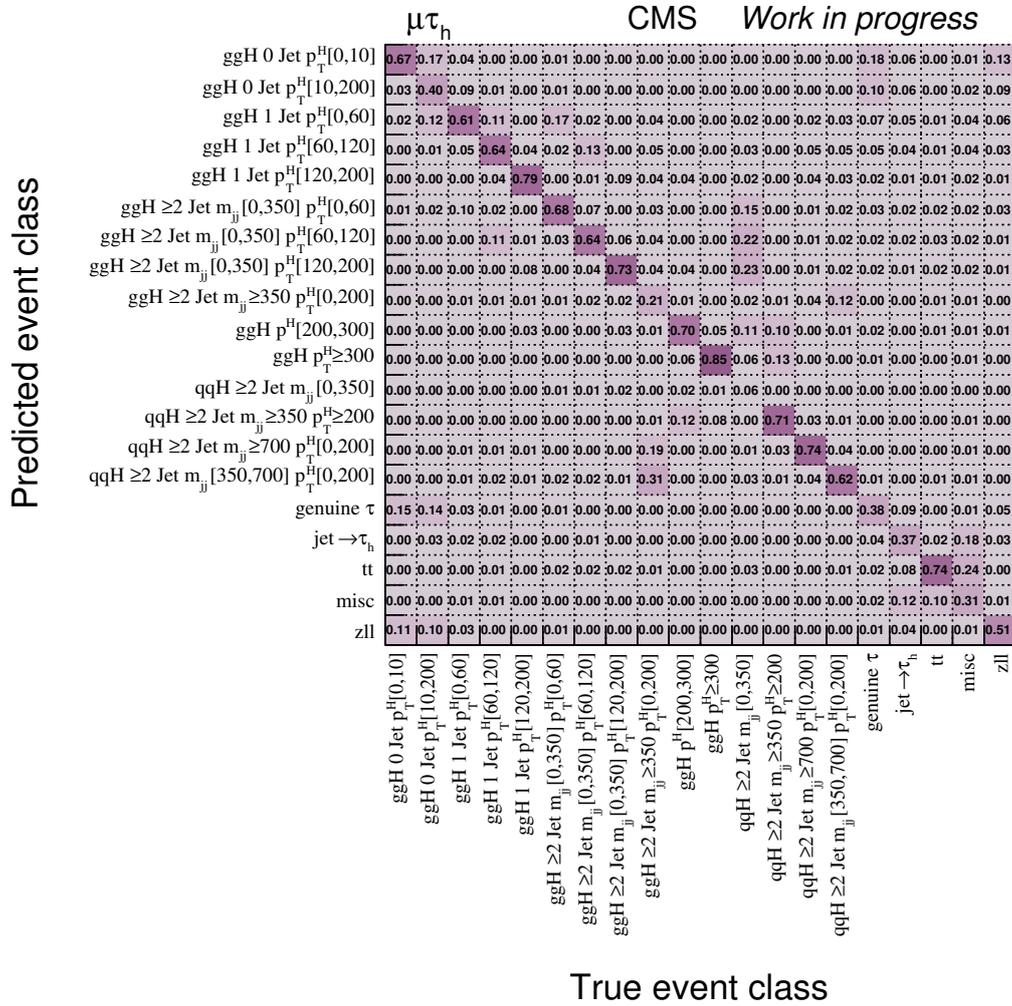
Figure 5.11: Confusion matrix of the stage-1 $\mu\tau_h$ classifier. Given numbers indicate the fraction of a true class directed to a predicted class. A random distribution would result in only 5 % of the events assigned correctly.

Statistical inference and results

The figures of merit in this analysis are best-fit values of signal strengths and fiducial cross sections as well as their uncertainties. These are obtained via maximum likelihood fits, which are the basis of most analyses in particle physics. They allow for sophisticated models taking into account not only statistical uncertainties but also a large set of systematic uncertainties, which is needed for such measurements.

6.1 The likelihood function

The likelihood function L is supposed to represent the probability density of certain observations for a given model. Of course, the actual observation is of particular interest and the fit algorithm is used to find the model parameters $\vec{x}$ for which the maximum value of the likelihood function is obtained. These parameter values represent the result of the measurement.

This analysis makes use of a binned likelihood meaning that events are filled into histograms where basically each bin corresponds to a counting experiment such that the likelihood factorises into

$$L(\vec{x}) = \prod_i L_i(\vec{x}), \tag{6.1}$$

where the index i runs over all bins in the histograms. Due to the setup of the collider experiment and the probabilistic character of quantum mechanics describing the proton-proton collisions, the likelihood of a single counting experiment can be modelled with a Poisson distribution $P(k, \lambda)$ where k denotes the integer number of observed events and λ the expected number of events. Only a small fraction of the measured collisions is selected for the analysis and finally for the considered bin.[1] The number of expected events is given by the model based on the event simulation and the data driven methods described in chapter 4. It is modified by the model parameters $\vec{x}$ where a distinction

---

[1]Note that this corresponds to binomial distributions in the limit of many trials at small probability. In fact, we would never observe more events in a bin than there were bunch crossings in the experiment, which in contrast would be possible according to the Poisson distribution. This difference is however negligible because the number of bunch crossings is more than ten orders of magnitude larger than the number of events that is expected in the bins of this analysis.

is made between the modifiers of the signal strength $\vec{\mu}$, also known as POIs, and the parameters controlling the effects of systematic variations $\vec{\theta}$, called nuisance parameters. In first place, this distinction is without technical motivation and only expresses our interest in the respective parameters. A technical difference comes up when probabilistic priors are added to the nuisance parameters. This is detailed in the following section. So far, the likelihood function reads:

$$L(\vec{\mu}, \vec{\theta}) = \prod_i P(k_i, \lambda_i(\vec{\mu}, \vec{\theta})) \tag{6.2}$$

6.2 Modelling of uncertainties

The Poisson distribution already models the statistical variation of event numbers that would be observed in repeated measurements. There are many systematic uncertainties that need to be taken into account as well, e.g. related to the energy scale of the considered particles. They are the reason why the expected number of events per bin cannot be predicted perfectly and are also introduced to the likelihood function such that the related nuisance parameters $\vec{\theta}$ vary the expected number of events, as already denoted in equation 6.2. Prior constraints on the nuisance parameters are added as additional factors to the likelihood function. In general those factors can follow any probability distribution. Here, they are chosen to be Gaussian distributions centred at zero and with unit width, which at the same time determines the scale of the nuisance parameters and how the impacts on λ_i need to be modelled.[2]

$$L(\vec{\mu}, \vec{\sigma}) = \prod_i P(k_i, \lambda_i(\vec{\mu}, \vec{\theta})) \cdot \prod_j G(\theta_j) \tag{6.3}$$

Some nuisance parameters simply scale a process, e.g. cross section uncertainties, such that the corresponding fraction of all bin contents can be scaled like $\lambda_{\mathrm{process}} \cdot \sigma_j^{\theta_j}$ where σ_j corresponds to a scale factor of one standard deviation. Other nuisances are more complex to model as they alter the shape of the histograms. In this case, all events are reprocessed with conditions varied up and down by one standard deviation of the nuisance in order to derive the effect of the nuisance. The fit model interpolates each bin between these "1σ-points" and the nominal bin content. Beyond, a linear extrapolation is applied. This approach is based on [79] but makes use of a sixth (not second) order polynomial for the interpolation in order to achieve a double differentiable transition from the interpolation to the linear extrapolation beyond the "1σ-points".

In principle, it is possible to introduce prior correlations between nuisance parameters to the likelihood function. However, this is not supported by the statistical framework used for CMS Higgs boson analyses for practical reasons. Instead, correlations are modelled by

---

[2]The likelihood that models the systematic uncertainties does in general not represent a probability in the frequentist sense because there is an exact systematic deviation. We just cannot determine it exactly. Therefore the likelihood rather models the degree of knowledge about the systematic variation. This is the case even if the uncertainty is based on a measurement with a large statistical component, unless this measurement would be repeated together with this analysis.

using a dedicated set of nuisance parameters: Typically, if two nuisances A and B have a certain correlation ρ, one common nuisance parameter θ_{AB} is introduced acting on both variations with a relative amplitude[3] of $\sqrt{|\rho|}$ and the relative direction depending on the sign of ρ. Two further nuisance parameters θ_A and θ_B are introduced acting only on one variation respectively with a relative amplitude of $\sqrt{1-|\rho|}$. The combined amplitude of each nuisance is the original one and the effective likelihood is the same as with two nuisance parameters correlated by ρ. This is not the only possible choice of parameters but it has the advantage that it is symmetric with respect to the to nuisances and in the case of a 50 % correlation all relative amplitudes are the same which keeps interpretations of postfit pulls and constraints simple.[4]

This scheme is used in this analysis in several cases. However, there is an exception that comes up with the usage of the τ-embedded samples. Several uncertainties related to corrections of selection efficiencies are treated as 50 % correlated between τ-embedded samples and fully simulated samples because the underlying scale factors have a shared numerator, which is the measured efficiency in observed data. With the above scheme, two nuisance parameters would be introduced acting on the fully simulated samples. But most analyses of different decay channels of the Higgs boson are not using τ-embedded samples and therefore have only one nuisance parameter for these uncertainties. In a combination of the analyses, this would lead to conflicts because common uncertainties would not have corresponding sets of nuisance parameters that could be merged. It would be necessary to a-posteriori split the nuisance in all other analyses the same way it is done in this analysis even though they do not use τ-embedded samples. Therefore, a different scheme is chosen in this analysis to cover correlations between τ-embedded and fully simulated samples:

One nuisance parameter acts on both types of samples, with full amplitude on fully simulated samples and a relative amplitude of 0.5 on τ-embedded samples. A second nuisance parameter acts only on τ-embedded samples with a relative amplitude of $\sqrt{1-0.5^2} \approx 0.87$. Again, the combined relative amplitudes are 1 in each case and it can be calculated with basic rules for covariances that the effective nuisances are 50 % correlated. Alternatively, it is quite comfortable to work with a vectorial representation like shown in figure 6.1. Details of this representation can be found in [80]. The essential facts are that uncertainties are represented by vectors with their length corresponding to the amplitude of the uncertainty and the correlation is related to the angle θ between vectors via $\rho = \cos(\theta)$. The combination of two uncertainties corresponds to the sum of the two vectors.

The presented modelling of the likelihood is the skeleton of the statistical model. The actual systematic uncertainties that are taken into account via this technique are listed in table 6.1. Most of them are related to the corrections described in section 4.4 and address efficiencies of selections, energy scales, reweighting to a more precise spectrum,

---

[3]Relative amplitude refers to the variation mediated by a nuisance parameter compared to the full variation that is intended for a nuisance.

[4]In case of a 50 % correlation, one additional advantage of this scheme is that it is intuitive. Note, that in practice, it is often sufficient to model these correlations only coarsely, such that one only distinguishes between this 50 % case, uncorrelated nuisances (independent nuisance parameters) and fully correlated nuisances (one shared nuisance parameter).
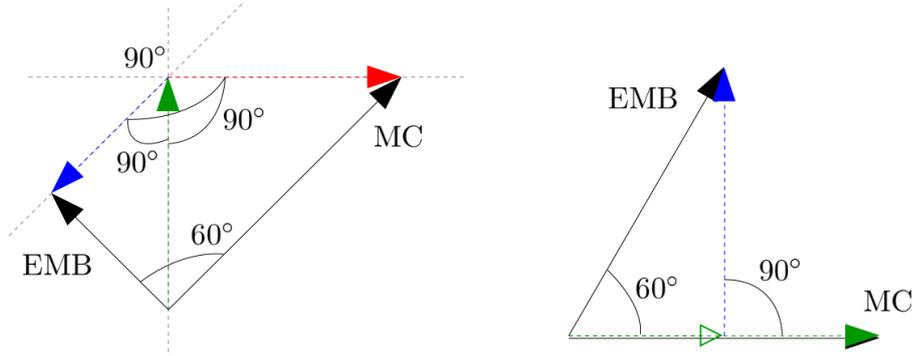
Figure 6.1: Illustration of the model for uncertainty correlations between τ-embedded and simulated samples. Left: Former and standard method with one nuisance parameter acting on both τ-embedded and simulated samples (green) with $1/\sqrt{2}$ of the target amplitude. Two additional nuisance parameters with the same amplitude act exclusively on τ-embedded (blue) or simulated (red) samples. An equilateral triangle is spanned with a 60° angle corresponding to a 50 % correlation between the resulting uncertainties modelled for the two sample types. Note that this extends over three dimensions. Right: Again one nuisance parameter acts on both samples (green) but on τ-embedded only with half the target amplitude. A second nuisance parameter acting exclusively on τ-embedded samples (blue) with $\sqrt{3}/2$ of the target amplitude. This again spans an equilateral triangle resulting in the same effective uncertainty model.

or the measurements of the extrapolation factors of the data driven background estimations. Uncertainties on theoretical cross sections and branching fractions are applied. For the signal, these are factorised into several levels of granularity. Besides the global components, there are additional components that describe the migration between bins of the STXS stage-1 scheme and a further set that addresses scale variations within each STXS bin without changing the yield of the total signal within the bin. However, the latter could potentially result in small yield changes within the analysis selection due to the experimental acceptance, but this was found to be at sub-percent level. In the measurements of cross sections, the uncertainties modifying the signal cross section at the level of granularity of the measurement or coarser are not included in the statistical inference. These are only included for the measurements of signal strengths.

A further important component of the statistical model, not listed in table 6.1, is the inclusion of statistical uncertainties due to the limited number of events used to create the model. This comprises both the event based modelling via simulation or τ-embedding and the data-driven methods where the generated histograms carry the statistical fluctuations from the data in the sidebands. For each histogram bin, a nuisance parameter that models the statistically allowed variation is introduced following the methods proposed in [81][79].

Table 6.1: List of systematic uncertainties

source	number of sub-sources (differential in)	variation per sub-source	affected processes	correlations ⊙	▽
luminosity	7 (source)	$< 2.2\,\%$	SIM	*	
prefiring 2016, 2017	1	$< 4\,\%$	SIM	$100\,\%$	
HLT eff. in $e\mu$	1	$2\,\%$	SIM,EMB	$0\,\%$	
HLT eff. in $e\tau_\mathrm{h}$	6 (trigger path, τ_h decay mode)	$2-7\,\%$	SIM,EMB	$0-50\,\%$	
HLT eff. in $\mu\tau_\mathrm{h}$	6 (trigger path, τ_h decay mode)	$2-7\,\%$	SIM,EMB	$0-50\,\%$	
HLT eff. in $\tau_\mathrm{h}\tau_\mathrm{h}$	4 (τ_h decay mode)	$\sim 7\,\%$	SIM,EMB	$0\,\%$	$50\,\%$
$\mu\mu$ HLT eff.	1	$4\,\%$	EMB	$0\,\%$	
e ID eff.	1	$2\,\%$	SIM,EMB	$100\,\%$	$50\,\%$
μ ID eff.	1	$2\,\%$	SIM,EMB	$100\,\%$	$50\,\%$
τ_h ID eff. in $e\tau_\mathrm{h}$, $\mu\tau_\mathrm{h}$	6 (source, $p_\mathrm{T}(\tau_\mathrm{h})$)	$2\,\%$	SIM,EMB	$0\,\%$	$50\,\%$
τ_h ID eff. in $\tau_\mathrm{h}\tau_\mathrm{h}$	5 (source, τ_h decay mode)	$2\,\%$	SIM,EMB	$0\,\%$	$50\,\%$
$e \rightsquigarrow \tau_\mathrm{h}$ eff.	2 ($|\eta|$)	$6-40\,\%$	$Z \to ll$	$0\,\%$	
$\mu \rightsquigarrow \tau_\mathrm{h}$ eff.	5 ($|\eta|$)	$8-60\,\%$	$Z \to ll$	$0\,\%$	
e energy	2 (scale, resolution)	$0.5-1.5\,\%^\dagger$	SIM,EMB	$100\,\%$	$0\,\%$
τ_h energy scale	4 (τ_h decay mode)	$0.2-1.1\,\%^\dagger$	SIM,EMB,F_F	$0\,\%$	$50\,\%$
jet energy	13 (scale per source, resolution)	$< 10\,\%^\dagger$	SIM	$0-100\,\%$	
$e \rightsquigarrow \tau_\mathrm{h}$ energy scale	4 (τ_h decay mode, $|\eta|$)	$0.8-6.6\,\%^\dagger$	$Z \to ll$	$0\,\%$	
$\mu \rightsquigarrow \tau_\mathrm{h}$ energy scale	2 (τ_h decay mode)	$1\,\%^\dagger$	$Z \to ll$	$0\,\%$	
$\not{E}_\mathrm{T}$ recoil	2 (scale, resolution)	$-3-5\,\%^\dagger$	signal, $Z \to ll$, W+jets	$0\,\%$	
$\not{E}_\mathrm{T}$ unclustered	1	$\sim 1\,\%^\dagger$	$t\bar{t}$, di-boson	$100\,\%$	
$\not{E}_\mathrm{T}$ τ-embedding	2($e\mu$, others)	$0.7-1.9\,\%^\dagger$	EMB	$100\,\%$	
b-tag eff.	2 (jet type)	$2-10\,\%$	SIM	$0\,\%$	
t quark p_T spectrum	1	$-10-10\,\%$	$t\bar{t}$		
Z p_T mass spectrum	1	$-2-10\,\%$	$Z \to ll$		
τ_h tracking eff.	2 (τ_h decay mode)	$1\,\%$	EMB	$50\,\%$	
$t\bar{t}$ contamination	1	$10\,\%^\ddagger$	EMB	$0\,\%$	
QCD raw extrap.	9 ($\#_\mathrm{jets}$, parameters)	$< 8\,\%$	QCD multijet	$0\,\%$	
QCD iso extrap.	1	$2-20\,\%$	QCD multijet	$100\,\%$	
F_F method	27 (channels, categories, $\#_\mathrm{jets}$, $\Delta R(e/\mu, \tau_\mathrm{h})$, source)	$< 10\,\%$	F_F	$0\,\%$	
$\sigma_\mathrm{di-boson,single-top}$	1	$5\,\%$	di-boson		
$\sigma_{t\bar{t}}$	1	$6\,\%$	$t\bar{t}$		
$\sigma_{W+\mathrm{jets}}$	1	$4\,\%$	W+jets		
σ_DY	1	$4\,\%$	$Z \to ll$		
$\mathcal{B}(H \to \tau\tau)$	3 (source)	$< 1.2\,\%$	$H \to \tau\tau$		
$\mathcal{B}(H \to WW)$	3 (source)	$< 1\,\%$	$H \to WW$		
PDF	4 (production mode)	$1.3-3.6\,\%$	H		
α_s	3 (production mode)	$0.8-8\,\%$	VH$_{lep}$,ttH, $H \to WW$		
$\alpha_\mathrm{s,STXS}$ ggH	9 (bin migration)	$< 18\,\%$	ggH		
$\alpha_\mathrm{s,STXS}$ qqH	10 (bin migration)	$< 18\,\%$	qqH		
STXS acceptance	11 (STXS bins)	$< 1\,\%$	ggH,qqH		

Abbreviations: SIM=simulation, EMB=τ-embedding, eff.=efficiency, extrap.=extrapolation
⊙ correlation between years of data-taking, ▽ correlation between simulated and τ-embedded samples
† acting on scale; ‡ acting on the $t\bar{t}$ fraction; given numbers indicate (local) yield changes otherwise
* $100\,\%$ for one source, $0\,\%$ else

6.3 Results

The nominal results of the measurement are obtained by maximising the likelihood and retrieving the signal parameters from the set of best-fit parameters $\{\hat{\vec{\mu}}, \hat{\vec{\theta}}\}$. Approximate estimates for the 68 % confidence intervals are obtained from a profiled likelihood [82]. This means that the dependence of the likelihood on a single parameter μ_i is scanned while all remaining signal and nuisance parameters are chosen to maximise the likelihood at the given values of the concerned parameter, denoted as $\{\breve{\vec{\mu}}_{\text{others}}, \breve{\vec{\theta}}\}$. The boundaries of the 68 % confidence interval are the values of the parameter μ_i where the likelihood ratio fulfils the following condition

$$- 2 \cdot \Delta \ln L(\mu_i) = 1, \text{ with } \Delta \ln L(\mu_i) := \ln \frac{L(\mu_i, \breve{\vec{\mu}}_{\text{others}}, \breve{\vec{\theta}})}{L(\hat{\vec{\mu}}, \hat{\vec{\theta}})} \tag{6.4}$$
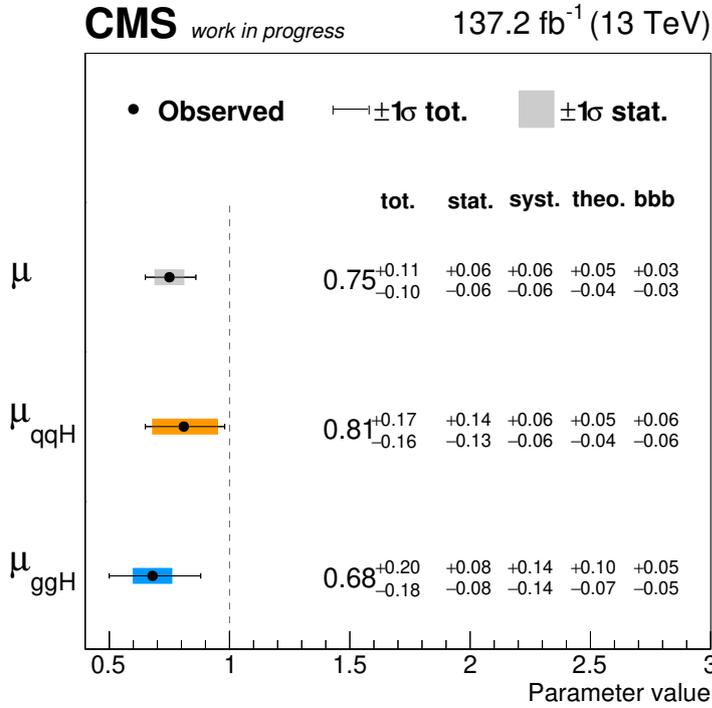
Figure 6.2: Results of the inclusive and STXS stage-0 signal strength measurements. The central values are shown together with the total 68 % confidence intervals. Furthermore, the uncertainties are provided as four fractions: finite statistics of observed data (stat.), systematic uncertainties from experimental sources (syst.), theory uncertainties (theo.), and finite statistics of the model (bbb).

The measurement of the inclusive $H \to \tau\tau$ signal is run on the STXS stage-0 categorisation presented in section 5.3.1. The measured signal strength μ, which reflects the ratio
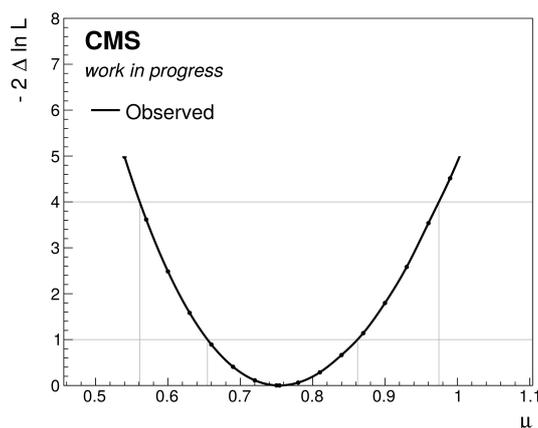
Figure 6.3: Profile likelihood scan of the inclusive measurement
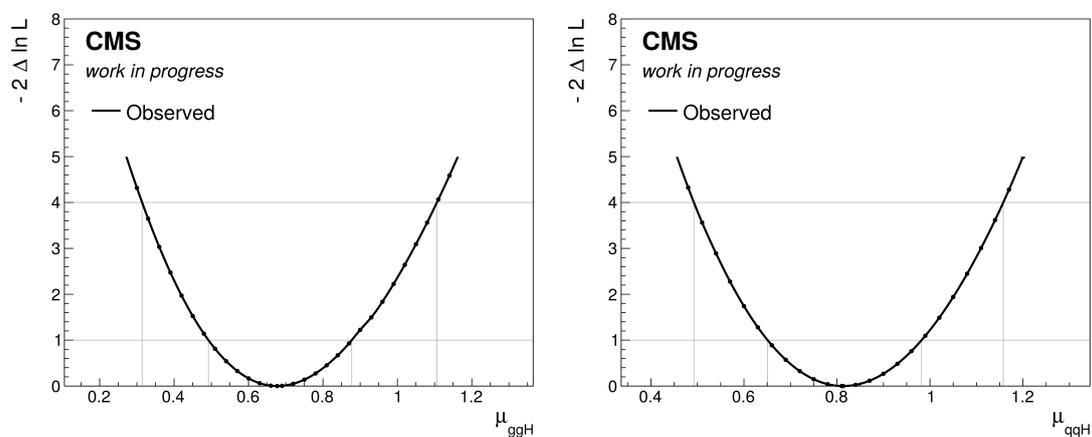
Figure 6.4: Profile likelihood scans of the STXS stage-0 measurement

of the observed cross section times branching fraction to the one predicted by the SM, is given in figure 6.2. It is compatible with the SM expectation within about two standard deviations. On the same categorisation, the simultaneous measurements of the signal strengths of ggH and qqH are performed. For these measurements, other production modes, which this analysis is not sensitive to, are set to the SM expectation and can only vary within given uncertainties. The results are again given in figure 6.2. Both of the results as well as the combined observation are compatible with the SM within less than two standard deviations. The measured correlation between the two parameters is -35%. Likelihood scans of the measured parameters are shown in figures 6.3 and 6.4. They illustrate how the measured numbers are determined and serve as a cross-check that the profiled likelihood has a smooth parabola-like shape.

The STXS stage-1 measurement is performed on the corresponding categorisation presented in section 5.3.2. The measured components of the signal correspond to the

partitioning given in figures 5.8 and 5.9. Again, minor signal contributions are fixed to the SM expectation. The measured central values and uncertainties are provided in figure 6.5. The correlations between the signal parameters are given in figure 6.6. Most of the correlations are smaller than 10 % and none exceed 50 %. Hence, the correlations are generally weak, which confirms the good separation between the single signal components achieved with the STXS stage-1 event categorisation. Likelihood scans for this measurement are provided in appendix D. Most of the measured parameters are compatible with the SM expectation within one or two standard deviations. However, some tension is observed for ggH events which do not include jets, which amounts to about three standard deviations each. This has been thoroughly checked for signs of mismodelling but seems to manifest as a down-fluctuation in the observed data or up-fluctuation in the background model. A negative signal strength is unlikely due to the signal model itself or the theory underlying it. Nevertheless, there can be a superposition of data fluctuations and a deviation of the theoretical predictions. Upper limits at 95 % confidence level, which are determined with the asymptotic approximation of the CL$_\mathrm{s}$ method [83][84][85], are additionally provided for the two concerned POIs. In both cases, they are larger than 70 % of the predicted signal strength, which only suggests a reduced signal but does not exclude it.

Comparing to measurements by CMS and ATLAS in the $H \to \gamma\gamma$ [17][18] and $H \to ZZ$ [23][24] channels, no such strong down-fluctuation in the 0-jet bins is observed. But it is worth mentioning that the result of the 0-jet bin with $p_\mathrm{T}^H < 10\,\mathrm{GeV}$ is smaller than 1 in all measurements and usually by more than what the 0-jet-bin with $p_\mathrm{T}^H > 10\,\mathrm{GeV}$ is larger than 1. E.g. 0.45 ± 0.3 is measured in the CMS $H \to \gamma\gamma$ analysis [17][5]. This promotes the assumption mentioned above, keeping in mind that up-fluctuations can also be involved in some of these measurements. A very common feature among these measurements is the trend along p_T^H. Among the 0-jet bins, there is always a larger signal strength measured for the high p_T^H bin than for the low p_T^H bin. Among the 1-jet bins, the bin with $p_\mathrm{T}^H < 60\,\mathrm{GeV}$ has always a signal strength below 1 and the smallest of all 1-jet bins. A larger signal strength is measured for the 1-jet bin with $60\,\mathrm{GeV} < p_\mathrm{T}^H < 120\,\mathrm{GeV}$ but the trend does usually not continue for the 1-jet bin with $120\,\mathrm{GeV} < p_\mathrm{T}^H < 200\,\mathrm{GeV}$. The 2-jet bins are difficult to compare due to different splittings used by the analyses. Other bins generally fluctuate in different directions. All this is only a qualitative observation and probably not yet significant with respect to the given uncertainties. Future combinations of these analyses will better reveal common trends and furthermore allow to quantify their significance.

The presented analysis and another one by CMS analyzing the same $H \to \tau\tau$ data [31] are cross-checking each other. Aside from several smaller differences in these analysis approaches, the main one is the event categorisation, which is based on cuts on single quantities and not on machine learning in the other analysis. The same trends are visible in the measurements of this analysis and both approaches deliver comparable results. Regarding the sensitivity however, the advantage of the presented multi-classification approach based on neural networks becomes visible in the STXS stage-1 results. While the sensitivity is similar for the more inclusive measurements, the presented analysis

---

[5]This value is read from a graphical representation and might vary by 0.05.
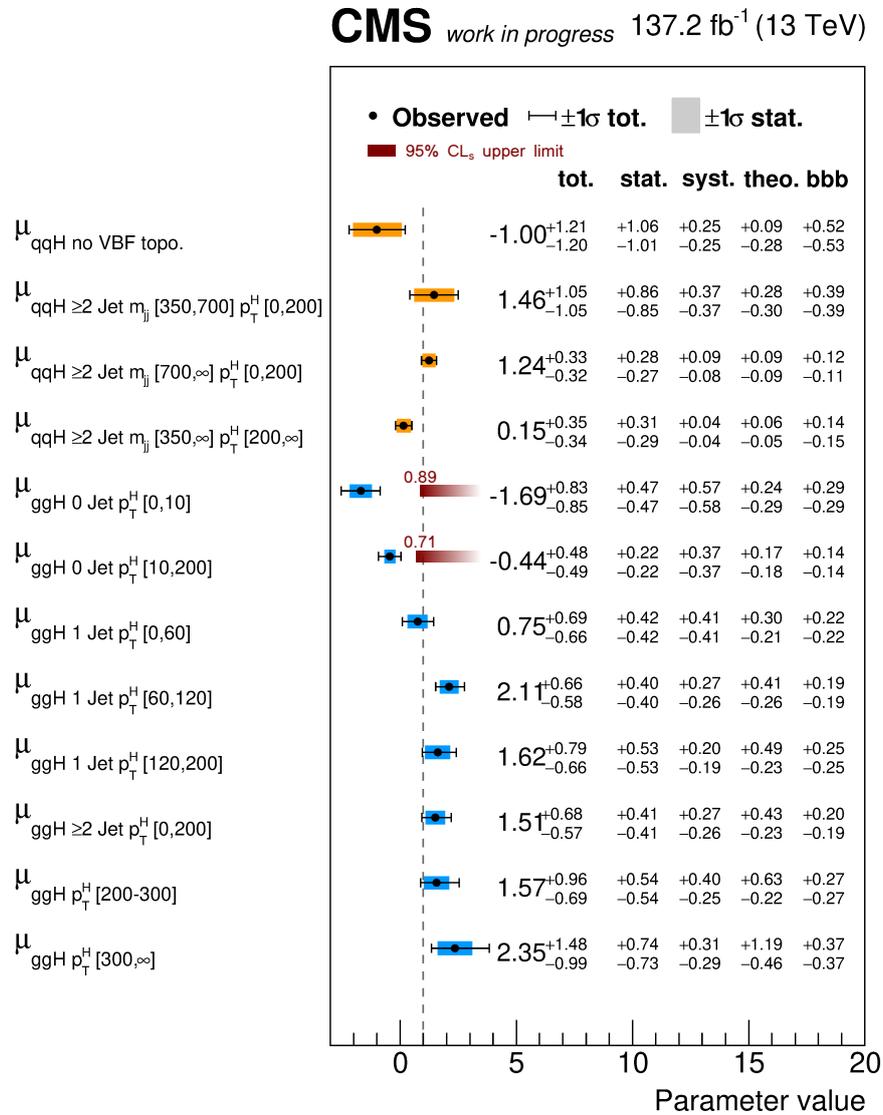
Figure 6.5: Results of the STXS stage-1 signal strength measurement. The central values are shown together with the total 68 % confidence intervals. Furthermore, the uncertainties are provided as four fractions: finite statistics of observed data (stat.), systematic uncertainties from experimental sources (syst.), theory uncertainties (theo.), and finite statistics of the model (bbb).
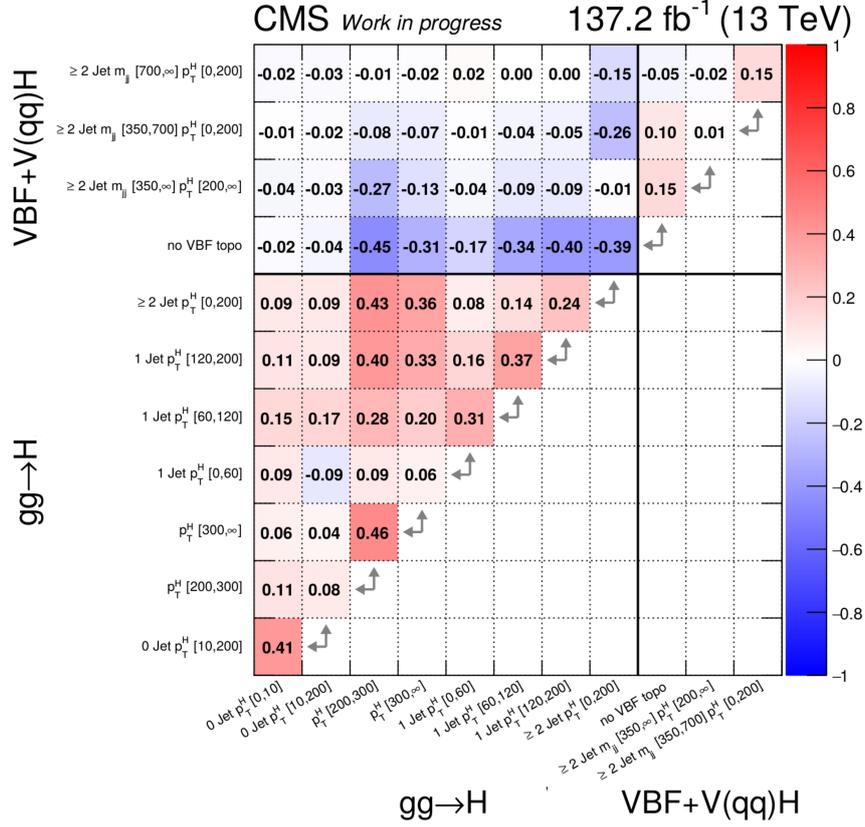
Figure 6.6: Correlation matrix of the STXS stage-1 measurement. Signal components of the same production mode cover exclusive phase-space regions of the final-state. In contrast, signal components of ggH and qqH can populate the same regions, which is why they are typically anti-correlated. Components of the same production mode are mostly correlated, which is mediated via shared overlapping components of the other production mode.

reaches a clearly enhanced sensitivity of the STXS stage-1 measurement, even achieving half the uncertainty for some of the signal parameters. Disentangling the large number of signal components is a complex task and the multi-classification is a powerful tool for this. This became visible in the context of the shown boundary effects and is confirmed by the results of the measurement.

So far, the results were presented as a measurement of the signal strength. Measurements of the cross section times branching fraction are provided in figure 6.7 for all STXS stages. In contrast to the measurements of the signal strength, the theoretical uncertainties that modify the signal yields at the level of granularity of the respective measurement or coarser are not included in the measurement. In the figure, they are shown separately and with respect to the SM prediction.

Finally, contours of two-dimensional scans of the profiled likelihood are provided in figure 6.8. They address the question about differences in the couplings of the Higgs boson to fermions and gauge bosons. One scan scales the signal strength, in which case the production modes are grouped according to the coupling of the Higgs boson that is involved in the production. In the other scan, the couplings to either fermions κ_F or vector bosons κ_V are modified directly, which involves the decay of Higgs boson. In both cases, the SM prediction is contained in the 95 % confidence interval of the measurement.
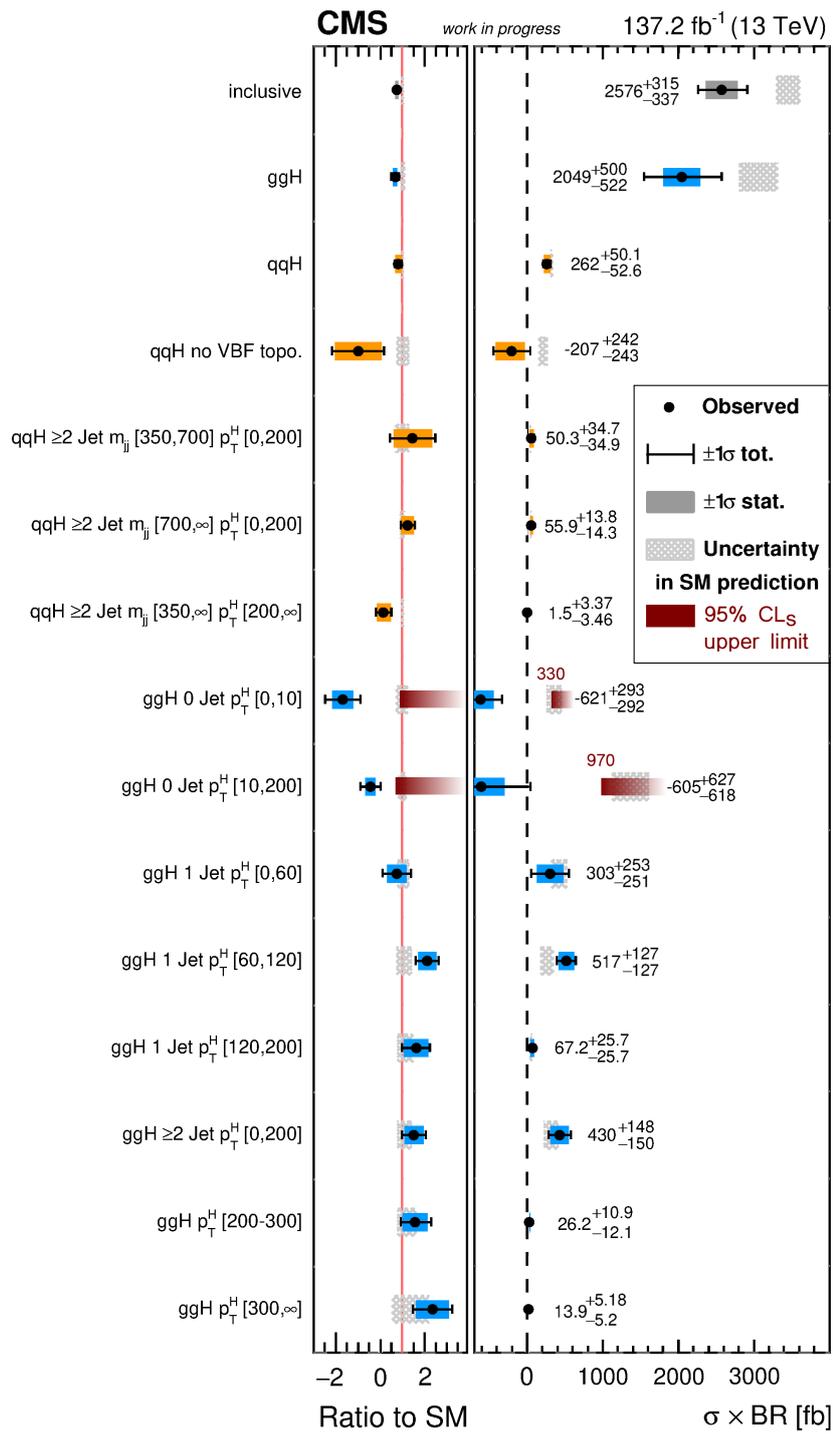
Figure 6.7: Measured cross sections times branching fraction ($\sigma \times \mathrm{BR}$)
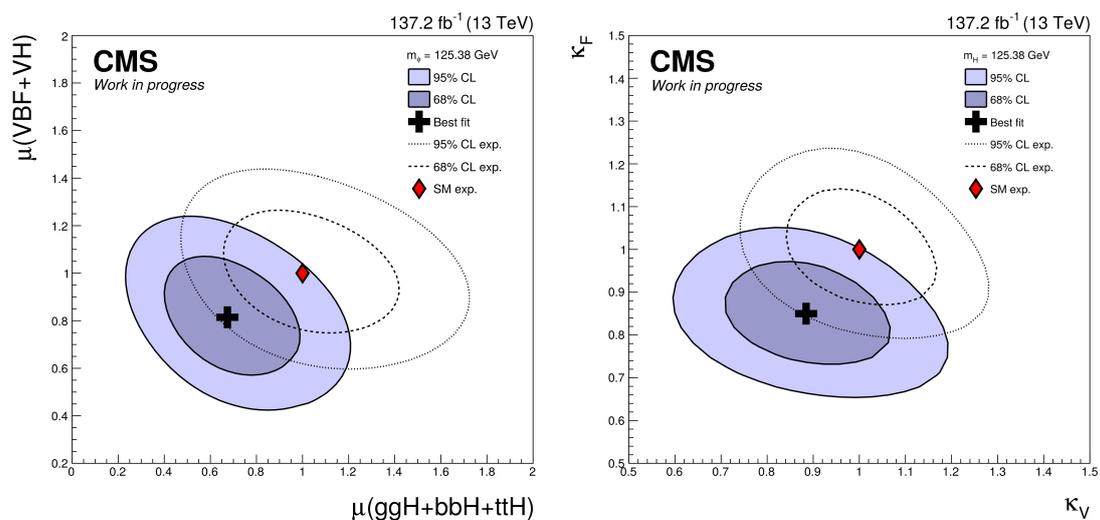
Figure 6.8: Contour plots of 2D likelihood scans with respect to Higgs boson coulings to fermions and gauge bosons. Left: The signal strength is scaled independently for the production modes in which the Higgs boson couples to gauge bosons and the ones in which the Higgs boson couples to fermions. Right: The couplings of the Higgs boson to fermions and gauge bosons are modified independently. This involves the decay of the Higgs boson. The minor contribution from $H \rightarrow WW$ events that enters this analysis is included as signal in this scan.

Conclusion and Outlook

Run-2 of the operation of the Large Hadron Collider (LHC) is a period of verifying the decay of the Higgs boson in multiple channels and making differential measurements of the signal rates with increasing precision. The differential measurements are based on the Simplified Template Cross Sections (STXS) scheme, which is shared among analyses of the ATLAS and CMS experiments. After the confirmation of the decay of the Higgs boson into two τ leptons, which was based on the data taken in 2016, the analysis of this decay channel proceeds to differential measurements as well.

In this context, the topic of this thesis is the measurement of differential cross sections in the $H \rightarrow \tau\tau$ channel with $137.2\,\mathrm{fb}^{-1}$ of data taken by the CMS experiment during Run-2. This is done by applying modern analysis techniques in order to obtain a solid data model and high sensitivity despite the numerous concurring background processes and signal components.

Tau pairs are reconstructed in the four most sensitive di-τ final states comprising $93.8\,\%$ of the $H \rightarrow \tau\tau$ decay channel. Up to $95\,\%$ of the background is modelled with data driven methods, which reduces the dependency on simulation and takes better into account details of the beam conditions and detector performance. Background with genuine τ pairs in the final state is modelled with the τ-embedding technique, which is based on lepton universality and transforms observed di-μ events into di-τ events such that only the τ pair and its signature in the detector are simulated. Backgrounds due to the misidentification of jets induced by gluons or quarks are modelled via extrapolation techniques where observed data from sideband regions is used to predict the contributions of these backgrounds in the region of the phase space that is used for the actual measurement.

The selected collision events are assigned to event categories using multi-classification based on neural networks. This classification is trained to identify events according to the underlying physics process or even sub-components as for the STXS stage-1 signal. The motivation behind this is to optimally separate signal from background and furthermore disentangle single signal components in order to be sensitive to each one of these. Within the event categories, the output scores of the classifiers are used for further discrimination between the considered processes.

The objective of the classification task is strongly related to the measurement and in

particular the level of granularity that defines the signal processes. Therefore, different classification approaches are followed for the STXS stage-1 measurement, which comprises twelve signal components in its version used for this analysis, and for the measurements of the ggH and qqH production modes or the inclusive $H \to \tau\tau$ signal.

The classifiers, which are trained for each of the analysed di-τ final states, are provided with a set of 15 input quantities that describe the kinematic properties of the τ candidates and yield information about associated jets and the Missing Transverse Energy. The quality of the modelling of these input quantities is thoroughly checked involving goodness of fit tests of the marginal distributions themselves and all pairwise correlations, altogether 1440 tests, as well as closer inspection in several cases. This is important for achieving a properly controlled modelling of the output quantities of the classifiers such that the risk of model-induced biases in the final measurement is minimised.

Cross sections and their ratios to the Standard Model of Particle Physics (SM) prediction are measured for three levels of granularity of the signal: Inclusive $H \to \tau\tau$, STXS stage-0, which means a simultaneous measurement of ggH and qqH, and STXS stage-1 with twelve kinematic signal bins. The results are compatible with the prediction of the SM. Largest tensions are observed for the two STXS bins of ggH which comprise events without jets. The tensions are at the order of three standard deviations but the measured signal rates are negative, which is hardly explainable via the signal process. Certainly, there is a contribution from statistical fluctuations in the data or the background model. A correlation between the signal rate and the transverse momentum of the Higgs boson is perceivable, which similarly appears in analyses of $H \to \gamma\gamma$ or $H \to ZZ$ events by the ATLAS and CMS experiments. It is not significant in the standalone measurements reported so far but might become more evident in future combinations of these measurements.

The comparison with a version of the analysis that does not use a machine learning driven event classification demonstrates the potential of the multi-classification approach for the STXS stage-1 measurement. Constraints that are improved by factors up to two would otherwise not be achieved before end of LHC Run-3.

With LHC Run-2 being completed, the resulting measurements of properties of the Higgs boson including the results shown in this analysis will stand for several years. Yet to come is the combination of measurements, which takes into account the measurements of multiple decay channels within each experiment and finally both ATLAS and CMS. Given the larger amount of data and having different coupling types in both the analysed production modes and decay channels, these combinations will have a higher sensitivity on the measurement of the actual couplings of the Higgs boson, be it the coupling to bosons, fermions or more particular types.

Control plots of neural network input quantities

The distributions of the input quantities for the neural networks and $\not{E}_T$ in addition are shown in the following for each year of data-taking and di-τ final state.

- 2016:
    - $e\mu$:   fig. A.1, A.2, A.3
    - $e\tau_h$:   fig. A.4, A.5, A.6
    - $\mu\tau_h$:   fig. A.7, A.8, A.9
    - $\tau_h\tau_h$: fig. A.10, A.11, A.12

- 2017:
    - $e\mu$:   fig. A.13, A.14, A.15
    - $e\tau_h$:   fig. A.16, A.17, A.18
    - $\mu\tau_h$:   fig. A.19, A.20, A.21
    - $\tau_h\tau_h$: fig. A.22, A.23, A.24

- 2018:
    - $e\mu$:   fig. A.25, A.26, A.27
    - $e\tau_h$:   fig. A.28, A.29, A.30
    - $\mu\tau_h$:   fig. A.31, A.32, A.33
    - $\tau_h\tau_h$: fig. A.34, A.35, A.36
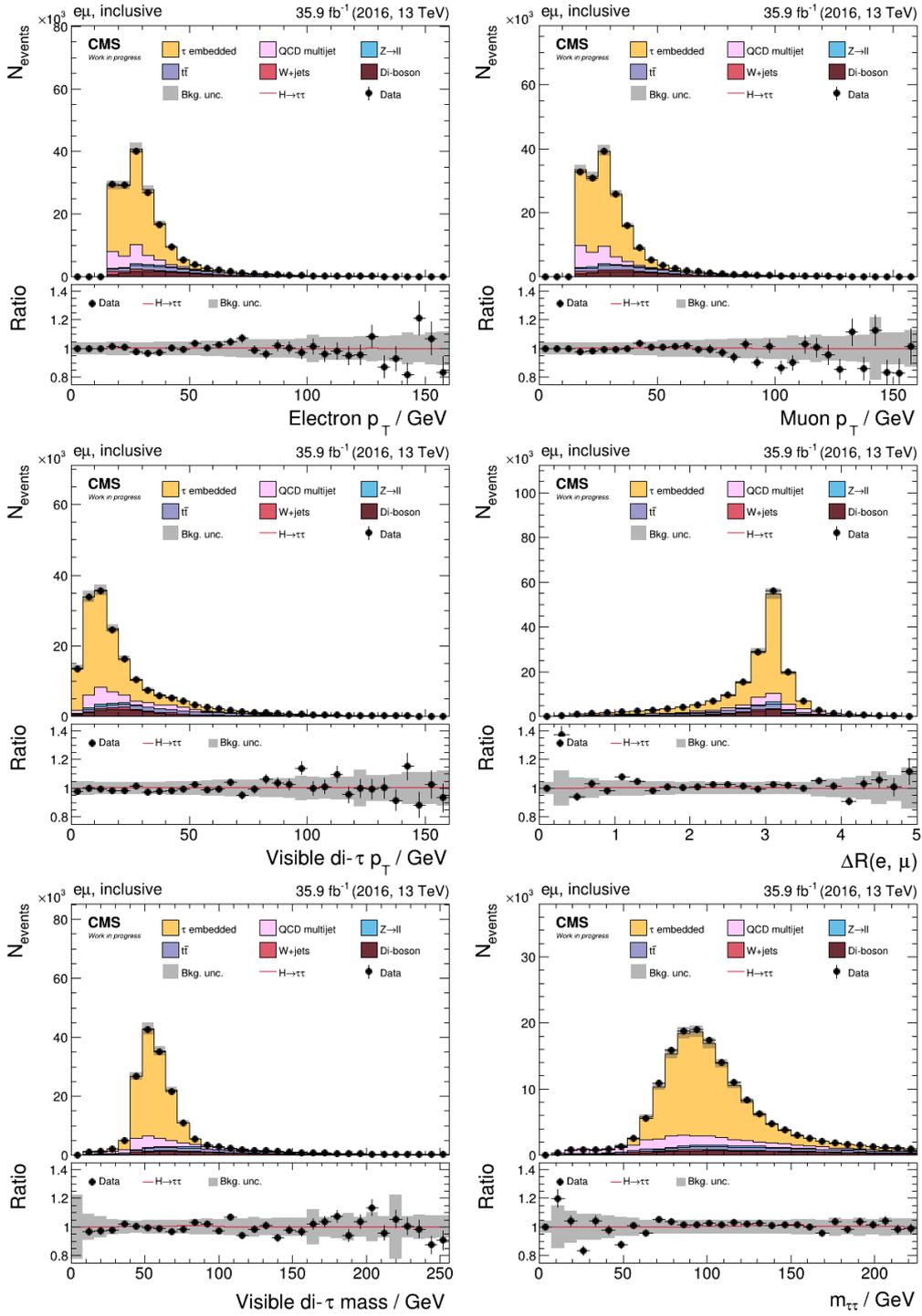
Figure A.1: Distributions of the input quantities in the 2016 *eμ* channel (I)

Figure A.2: Distributions of the input quantities in the 2016 $e\mu$ channel (II)
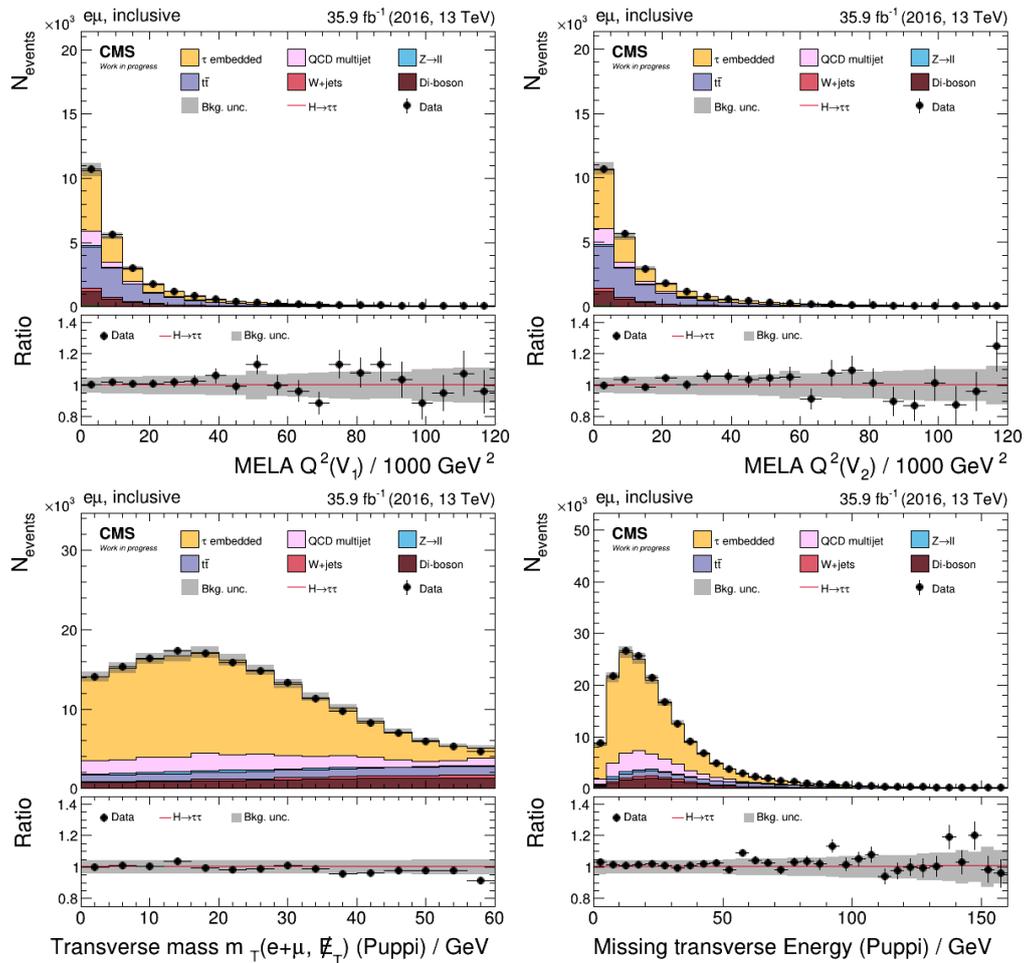
Figure A.3: Distributions of the input quantities in the 2016 *eμ* channel (III)
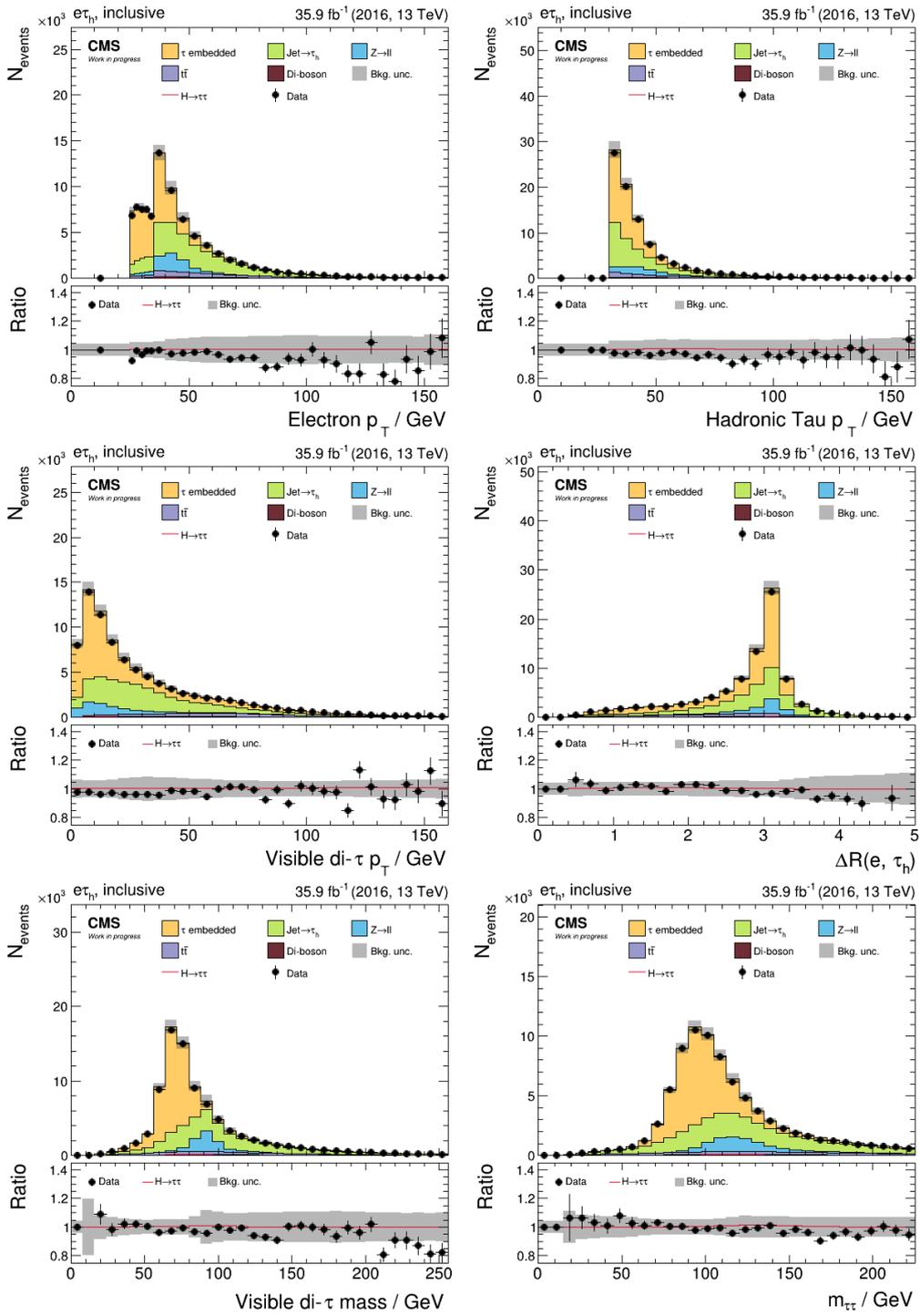
Figure A.4: Distributions of the input quantities in the 2016 $e\tau_{\mathrm{h}}$ channel (I)
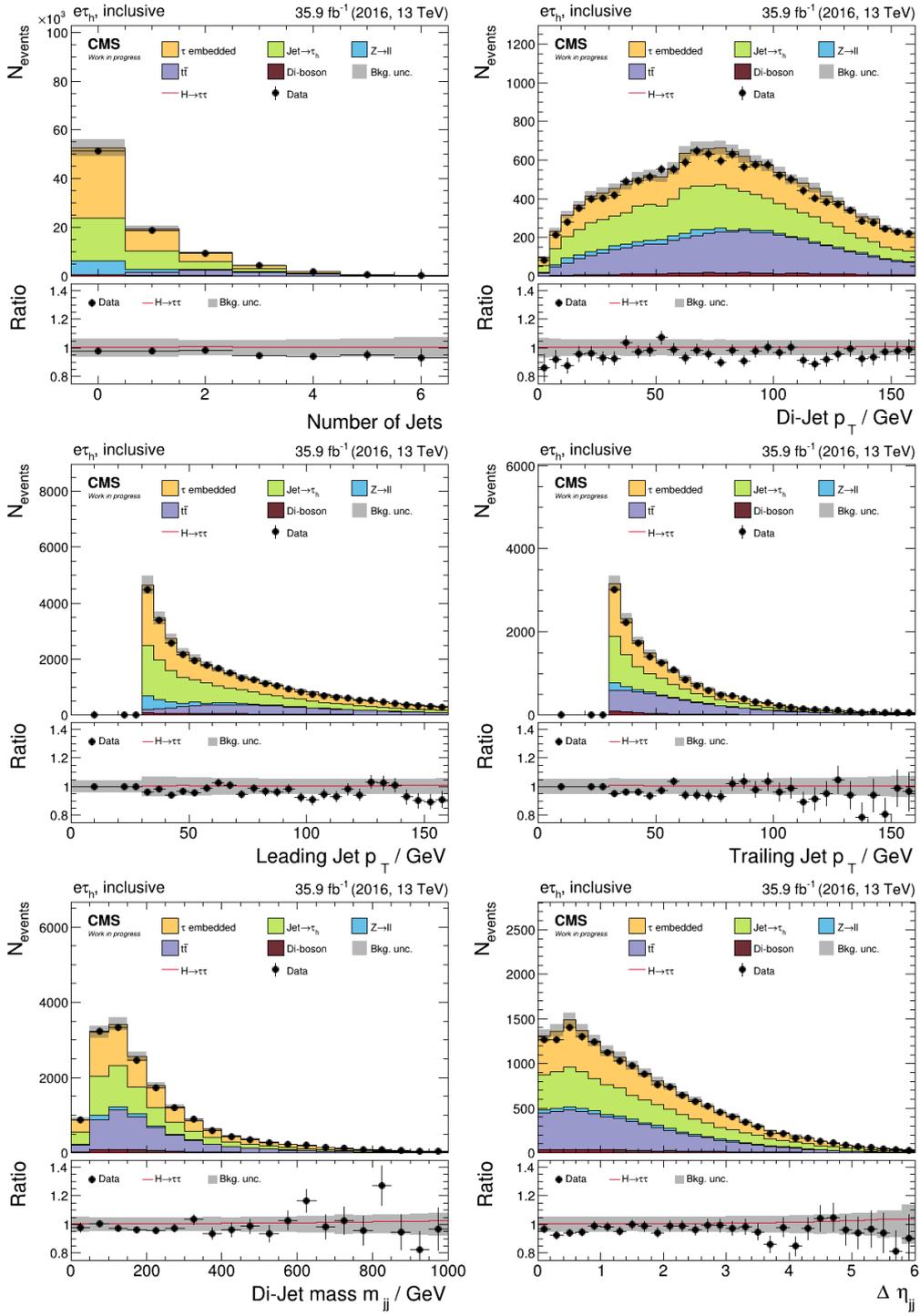
Figure A.5: Distributions of the input quantities in the 2016 $e\tau_{\mathrm{h}}$ channel (II)
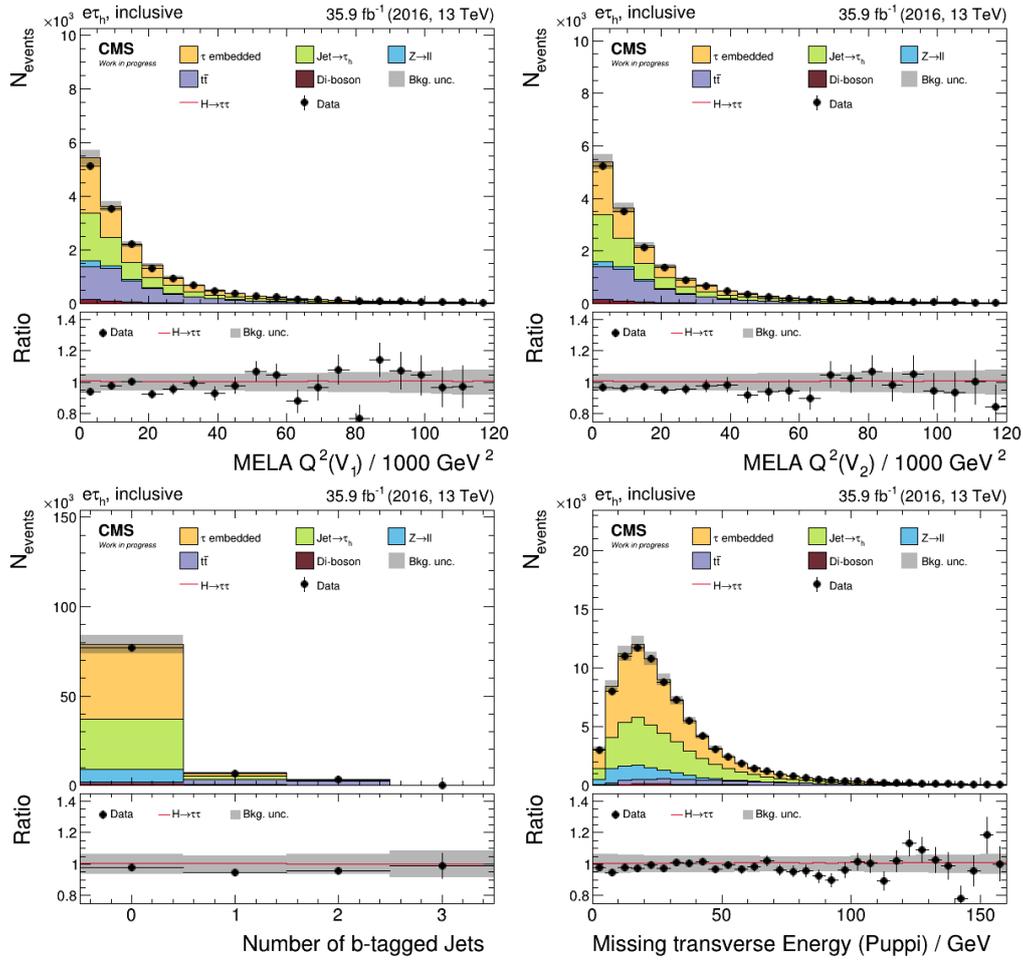
Figure A.6: Distributions of the input quantities in the 2016 $e\tau_\mathrm{h}$ channel (III)
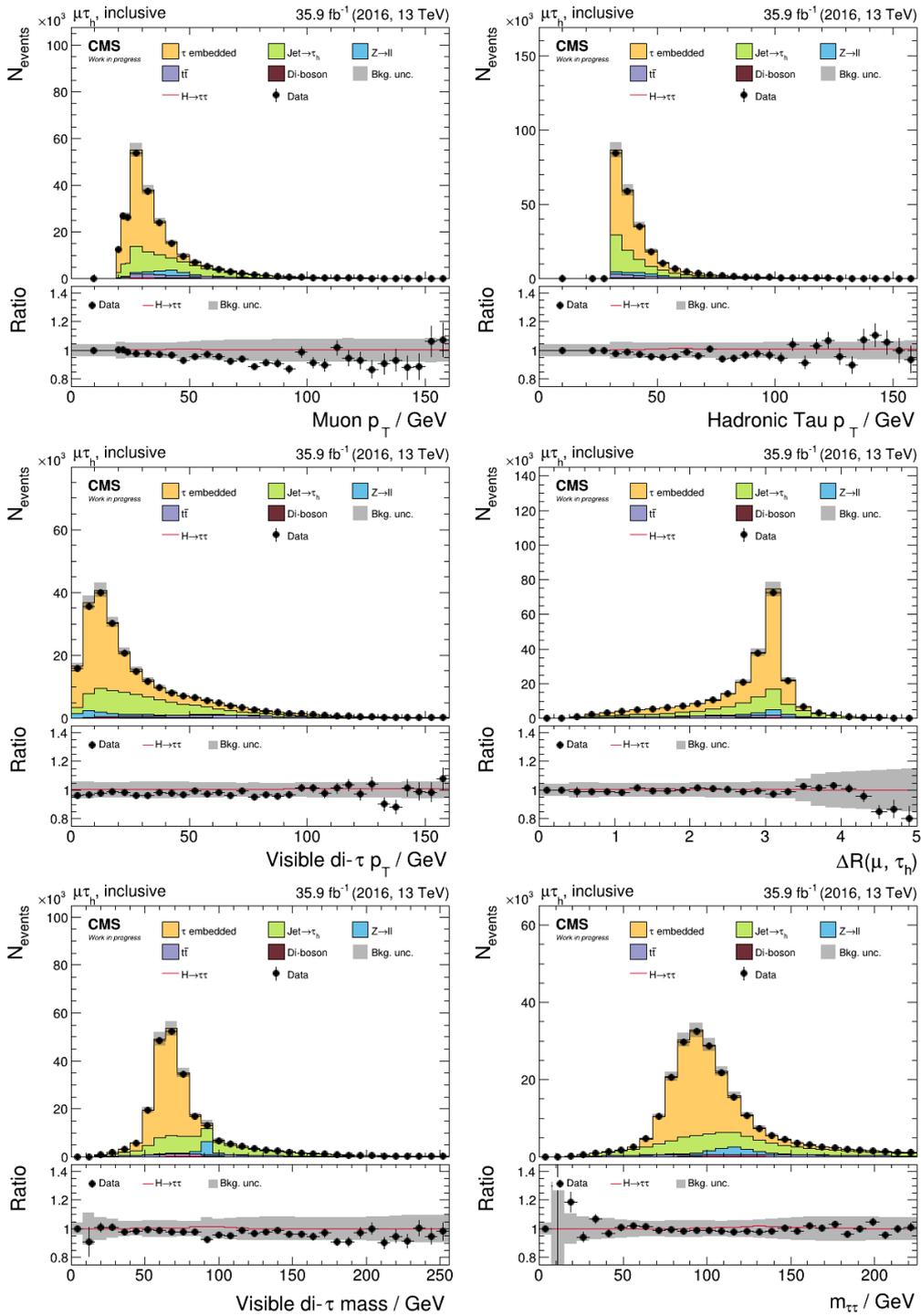
Figure A.7: Distributions of the input quantities in the 2016 $\mu\tau_{\mathrm{h}}$ channel (I)
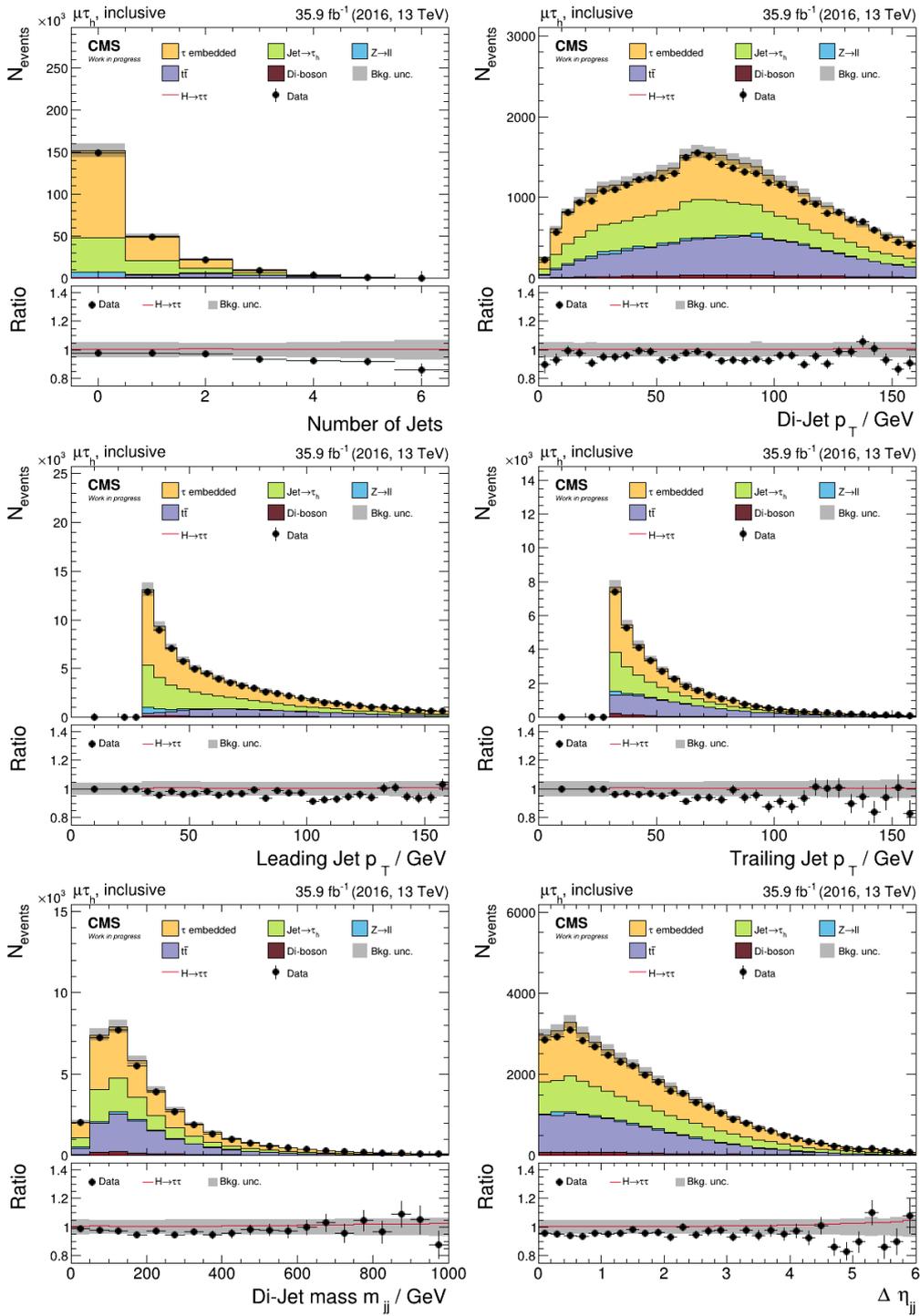
Figure A.8: Distributions of the input quantities in the 2016 $\mu\tau_\mathrm{h}$ channel (II)
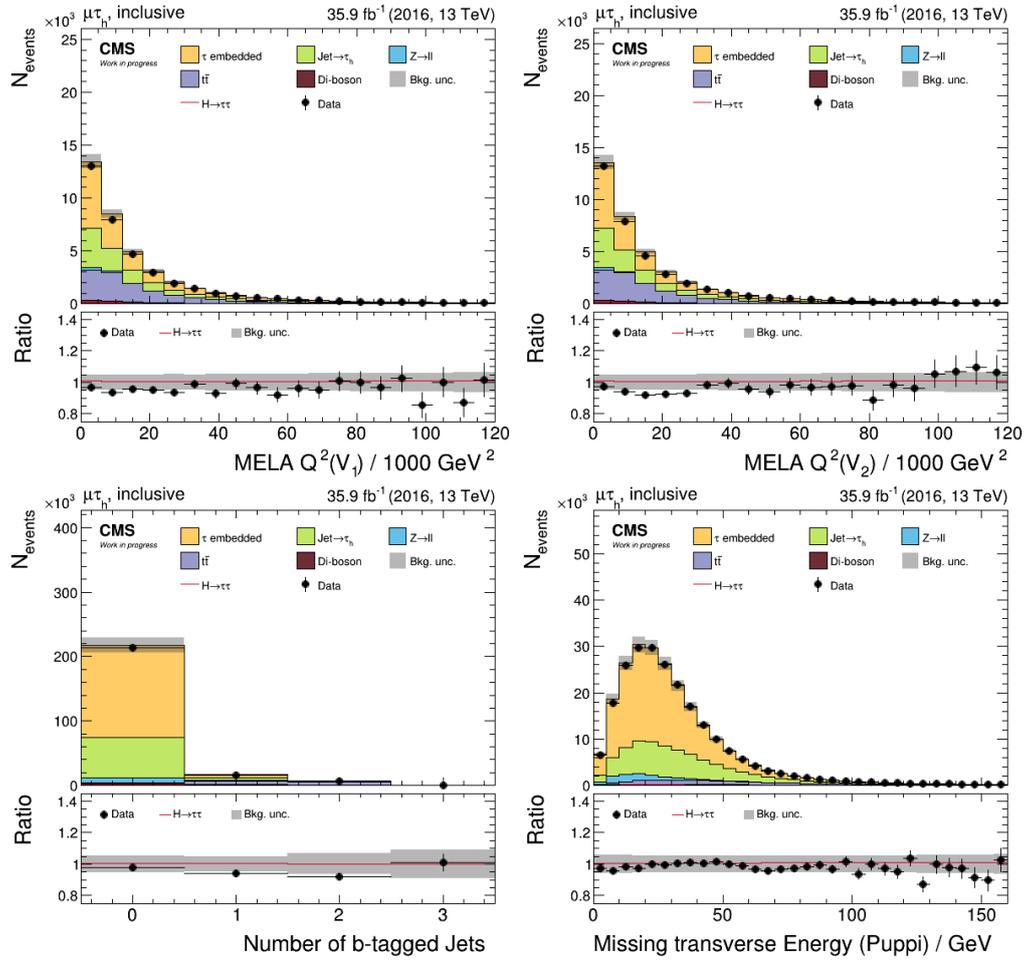
Figure A.9: Distributions of the input quantities in the 2016 $\mu\tau_h$ channel (III)
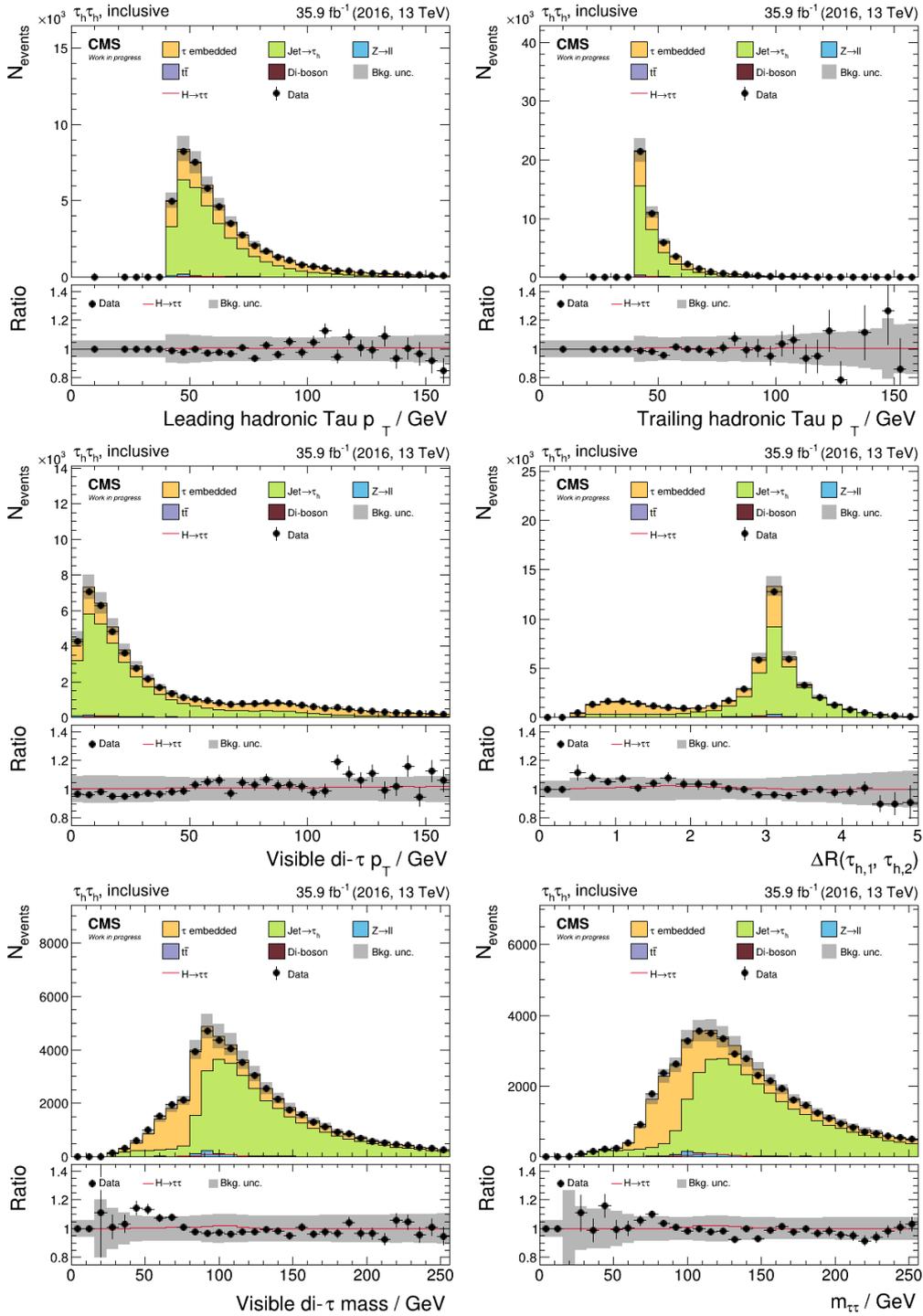
Figure A.10: Distributions of the input quantities in the 2016 $\tau_h\tau_h$ channel (I)
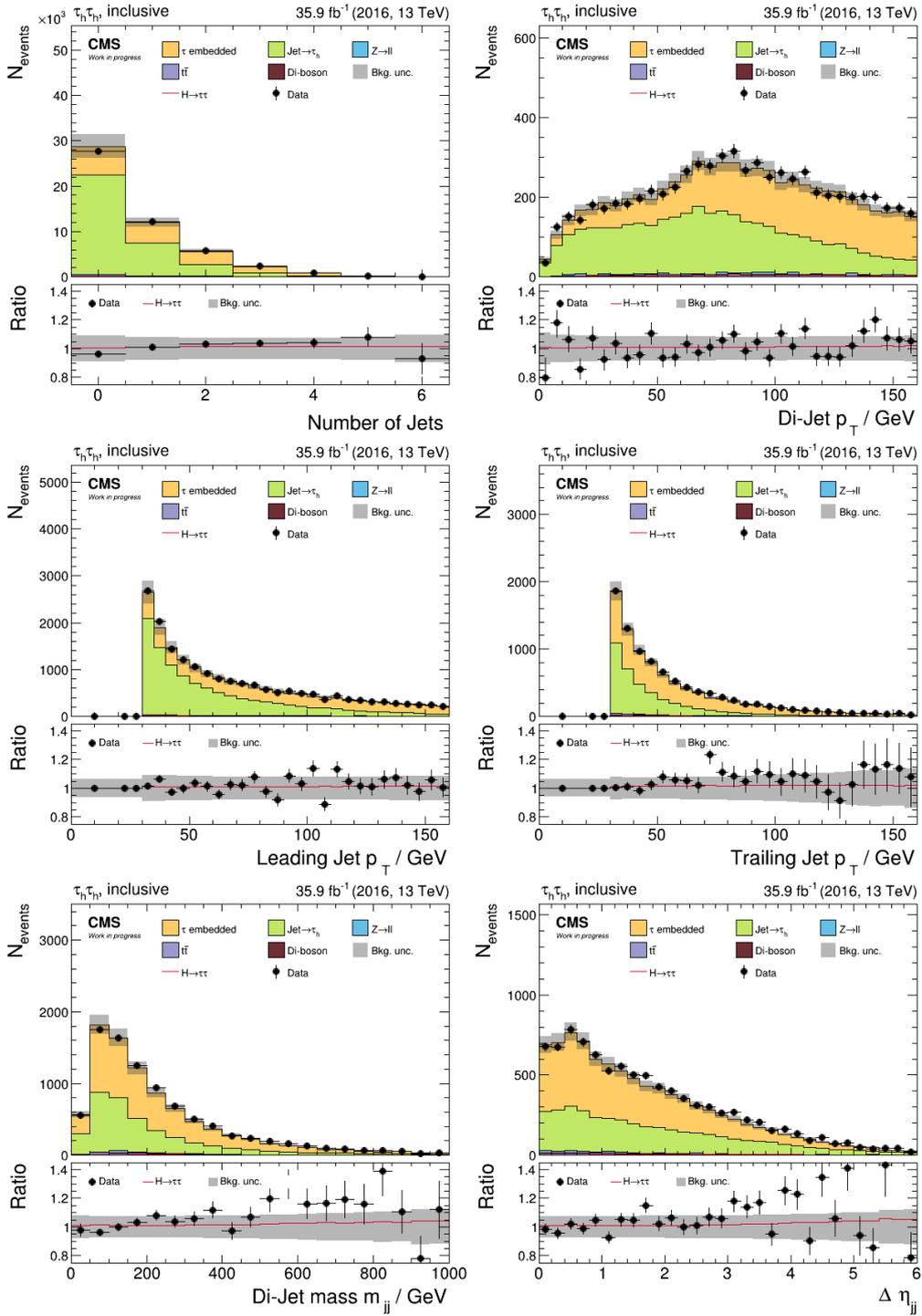
Figure A.11: Distributions of the input quantities in the 2016 $\tau_h\tau_h$ channel (II)
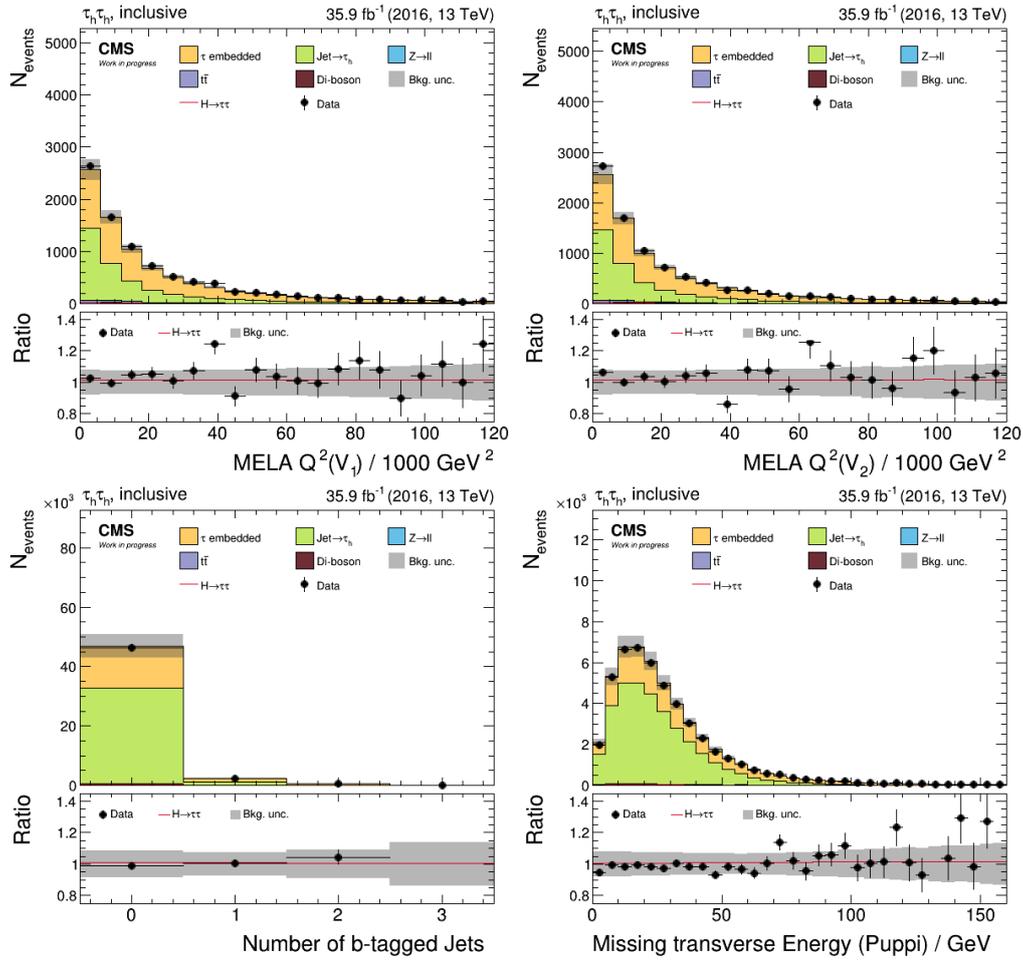
Figure A.12: Distributions of the input quantities in the 2016 $\tau_\mathrm{h}\tau_\mathrm{h}$ channel (III)
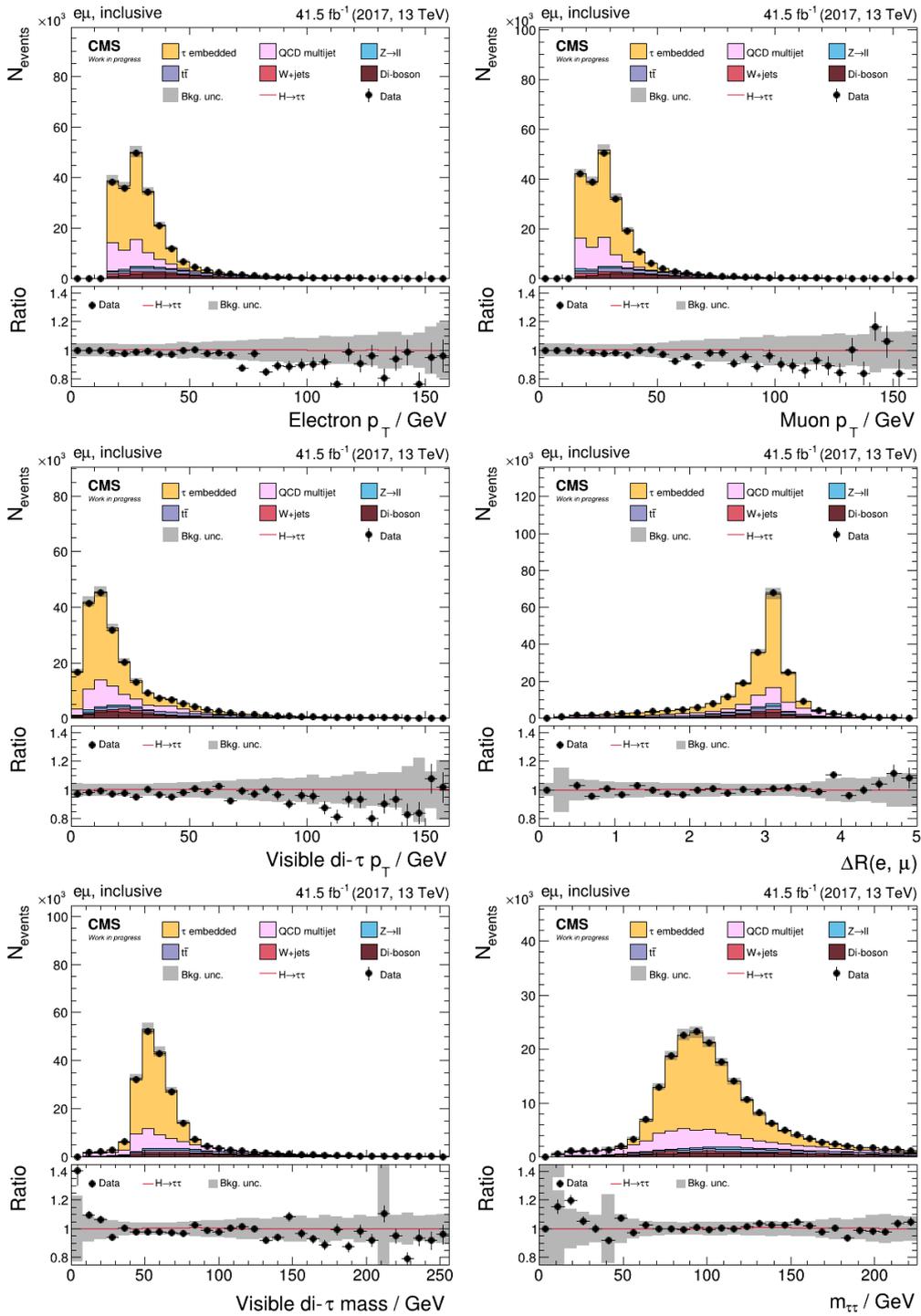
Figure A.13: Distributions of the input quantities in the 2017 $e\mu$ channel (I)
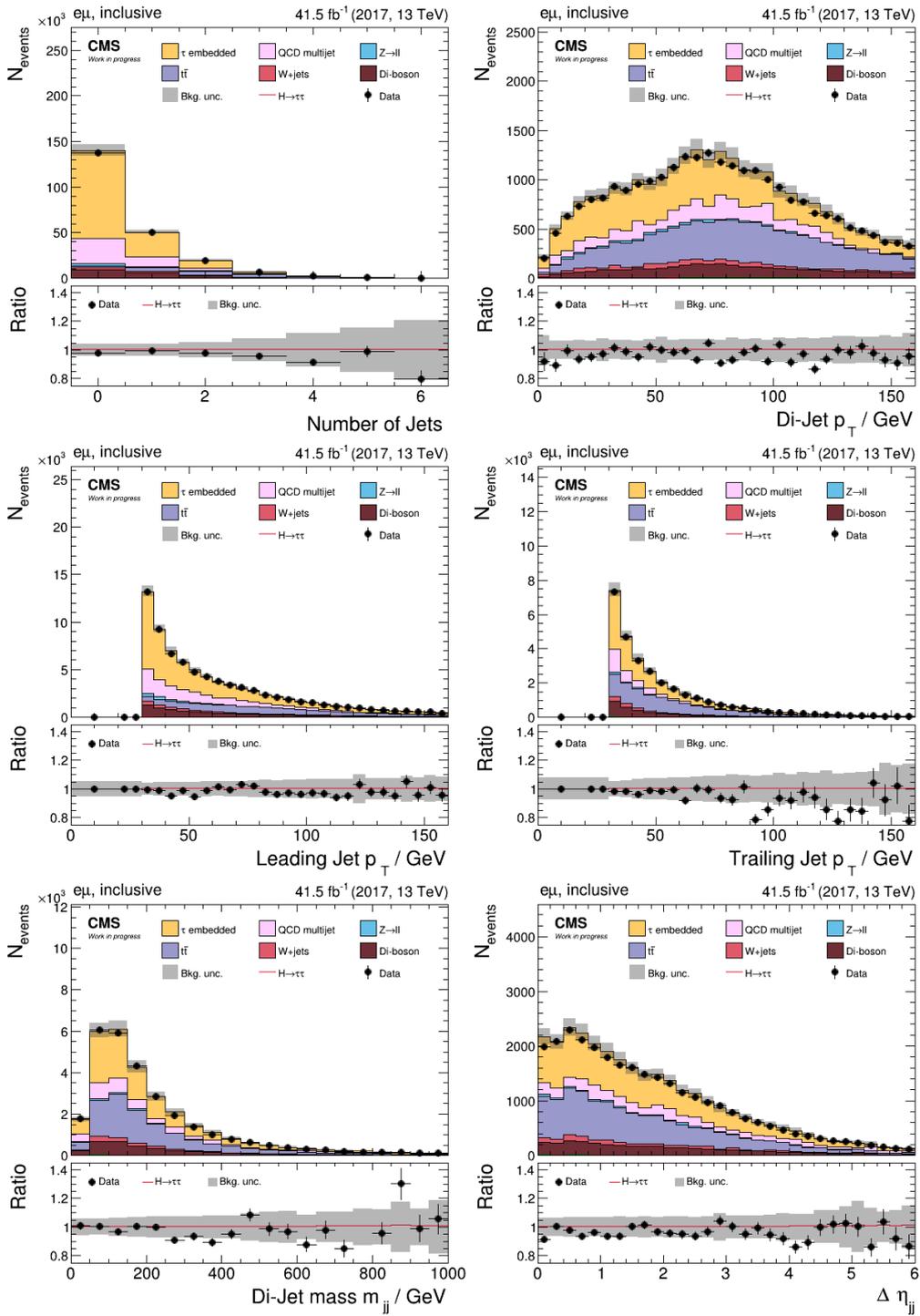
Figure A.14: Distributions of the input quantities in the 2017 $e\mu$ channel (II)
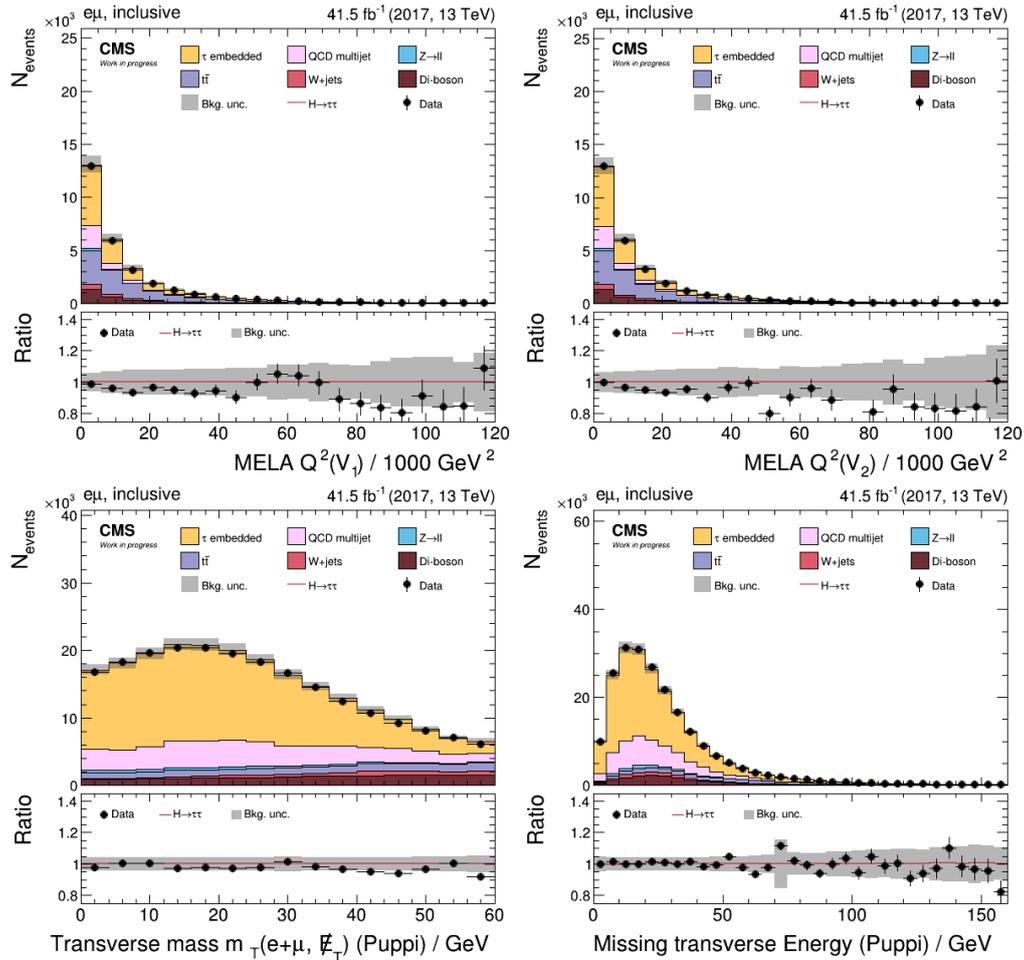
Figure A.15: Distributions of the input quantities in the 2017 $e\mu$ channel (III)
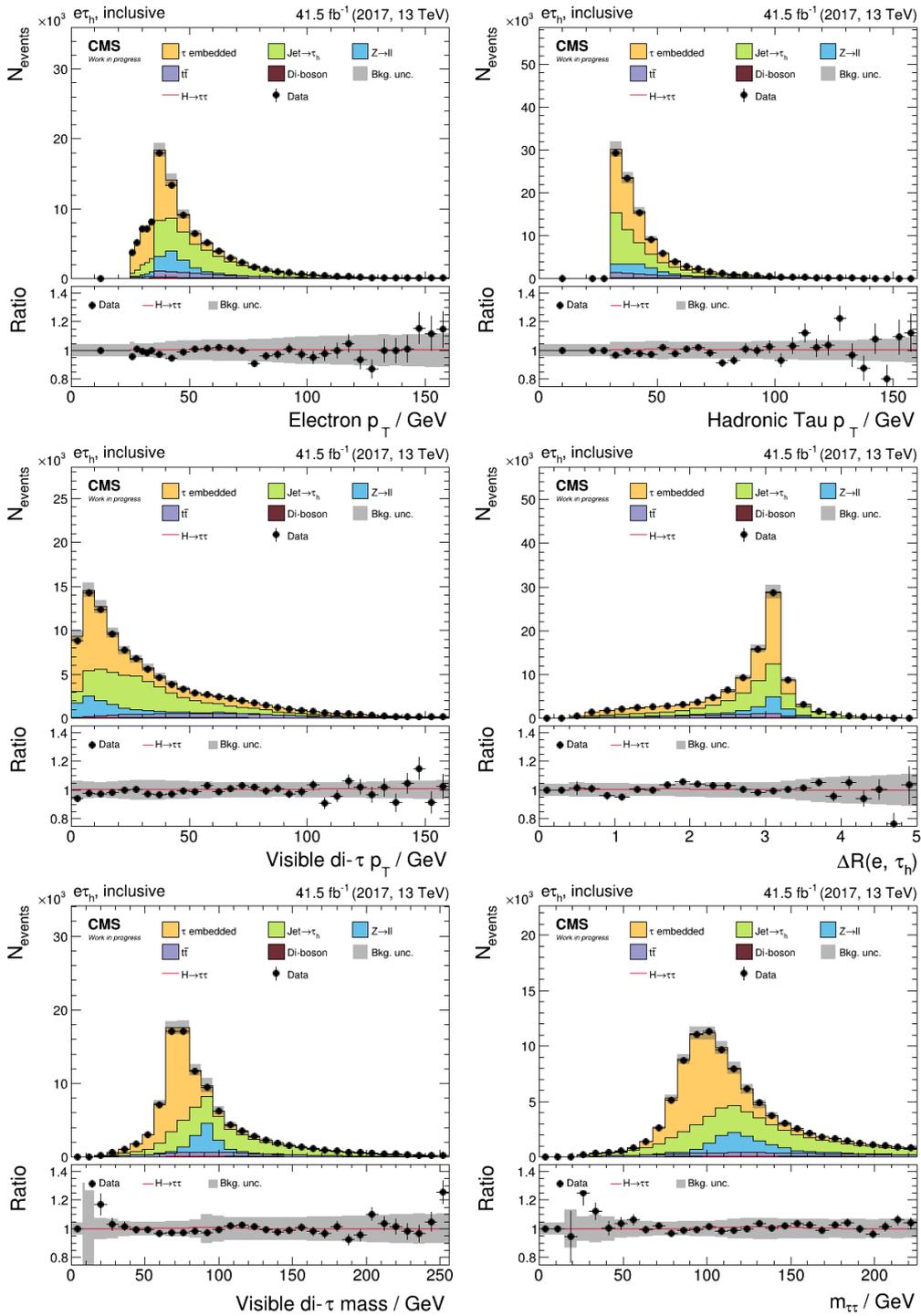
Figure A.16: Distributions of the input quantities in the 2017 $e\tau_\mathrm{h}$ channel (I)
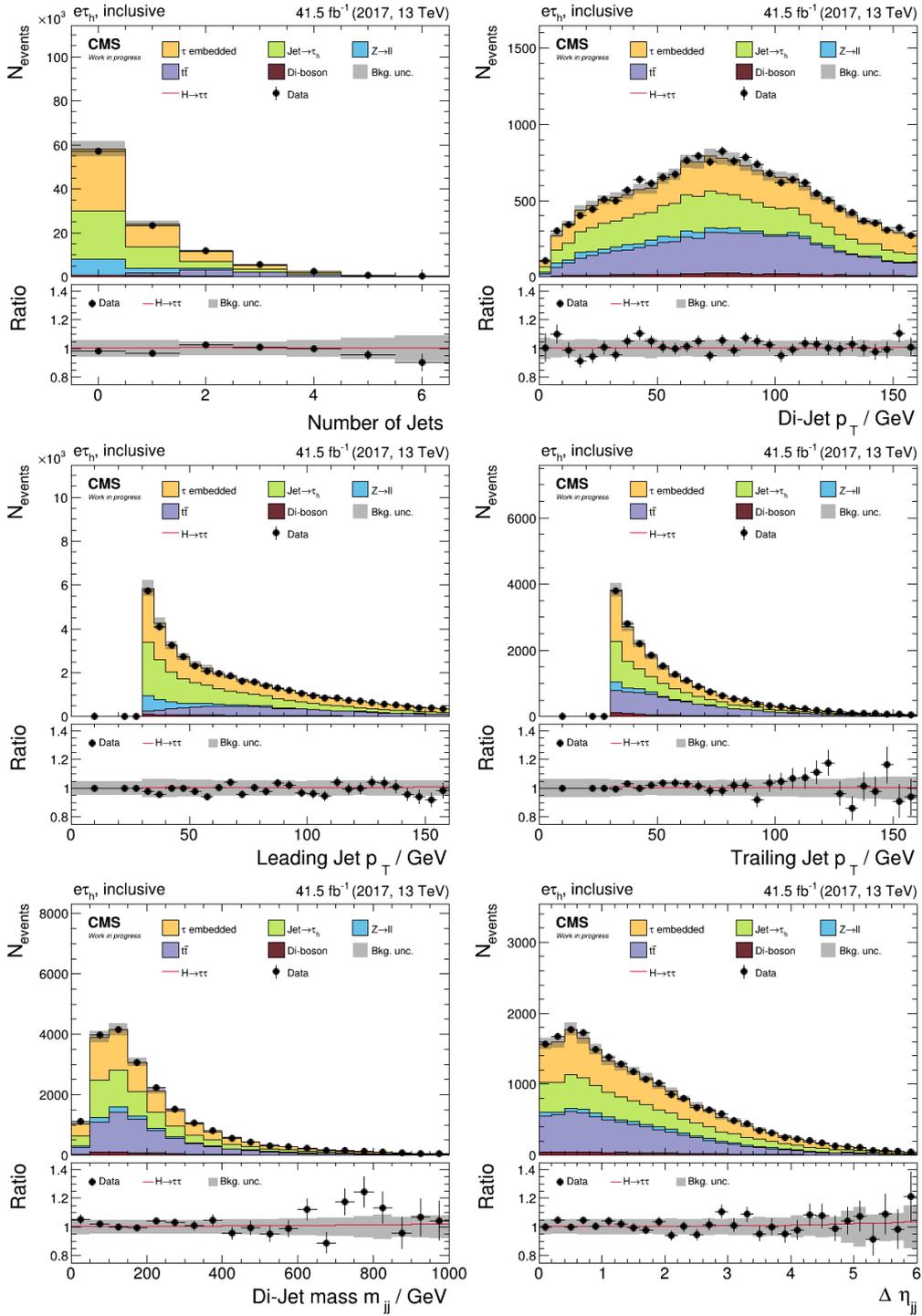
Figure A.17: Distributions of the input quantities in the 2017 $e\tau_\mathrm{h}$ channel (II)
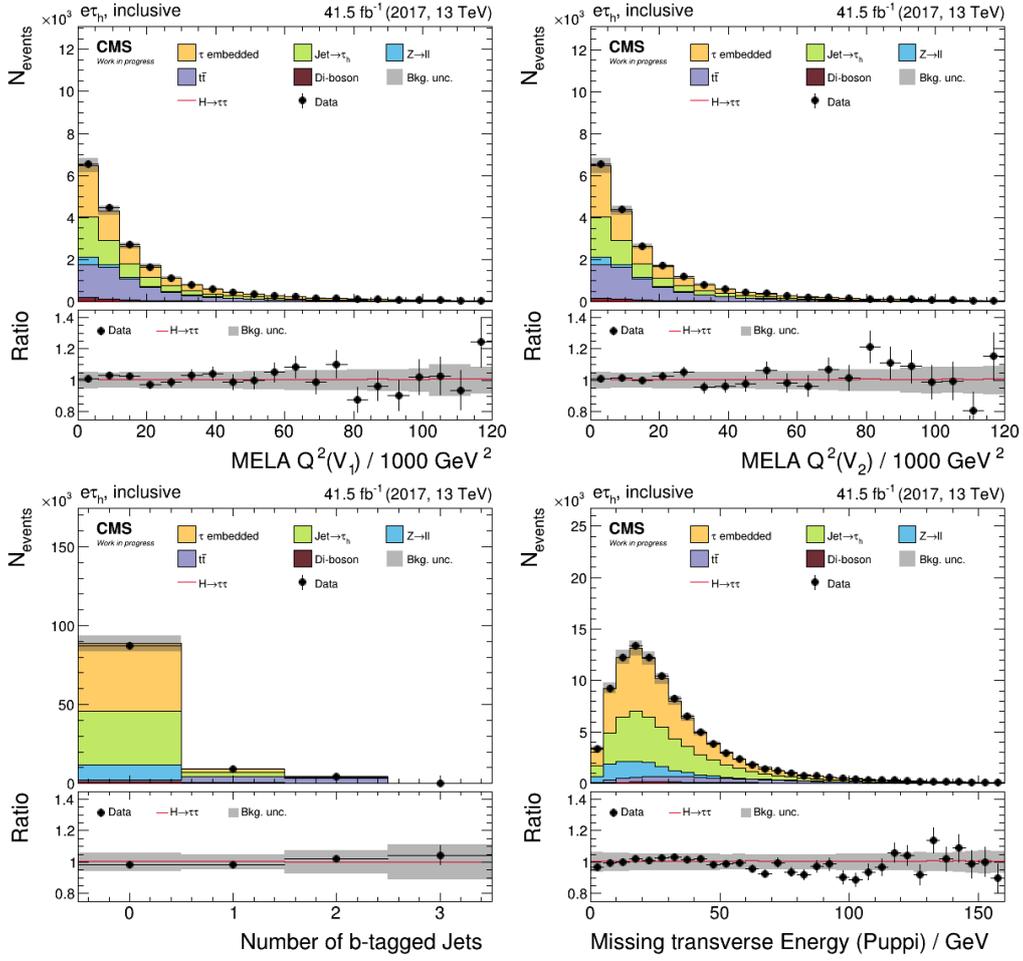
Figure A.18: Distributions of the input quantities in the 2017 $e\tau_h$ channel (III)
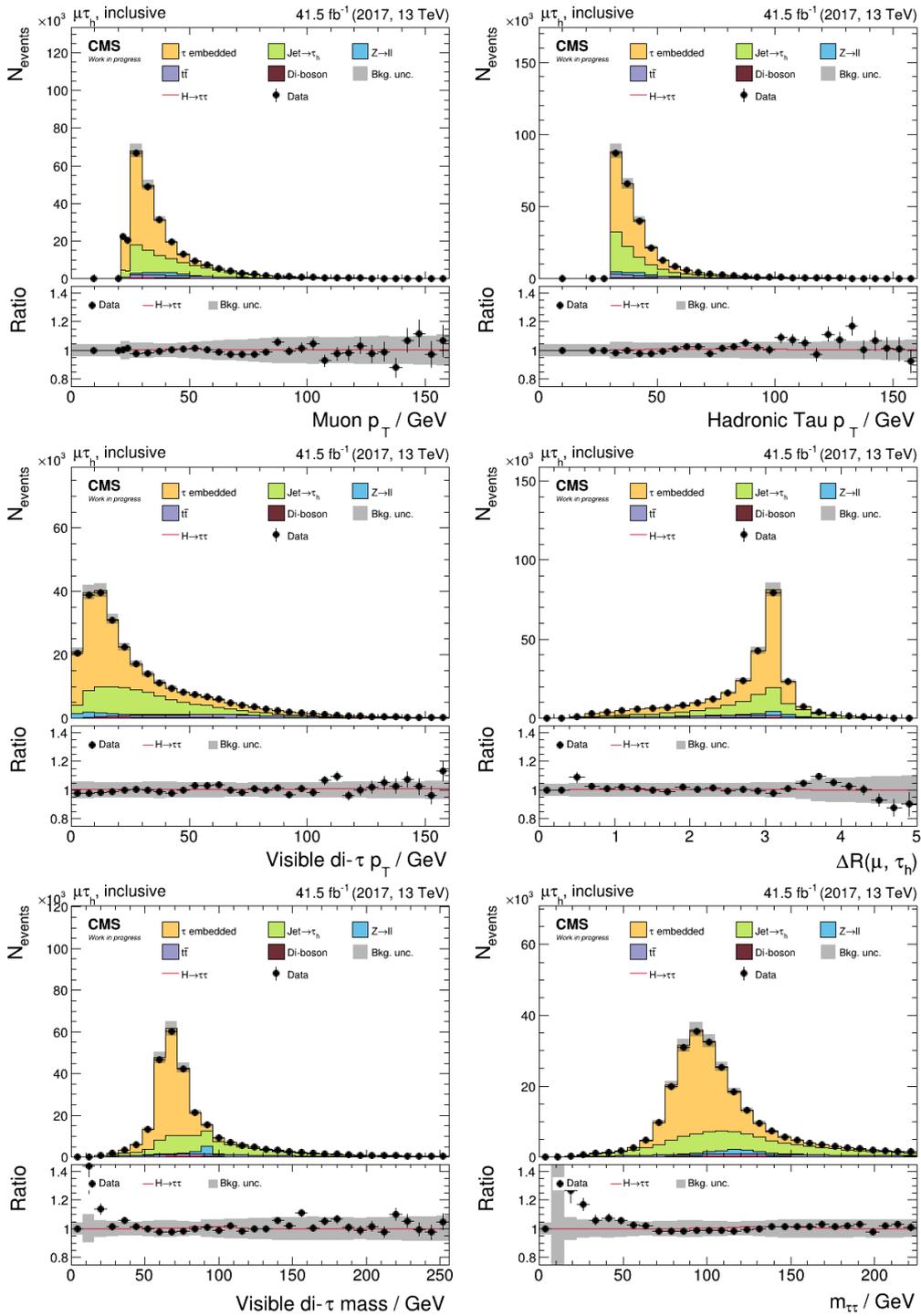
Figure A.19: Distributions of the input quantities in the 2017 $e\mu$ channel (I)
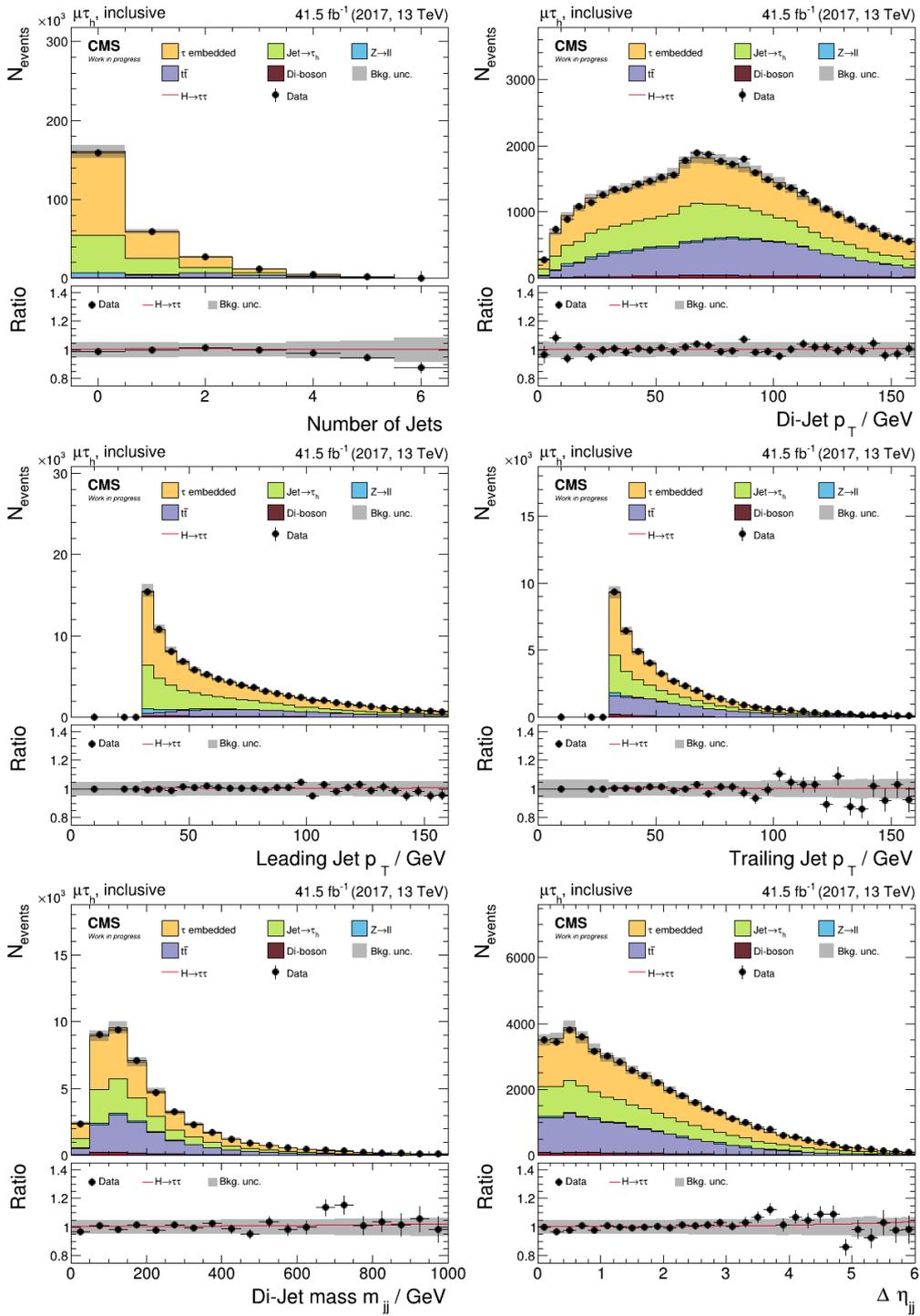
Figure A.20: Distributions of the input quantities in the 2017 $e\mu$ channel (II)
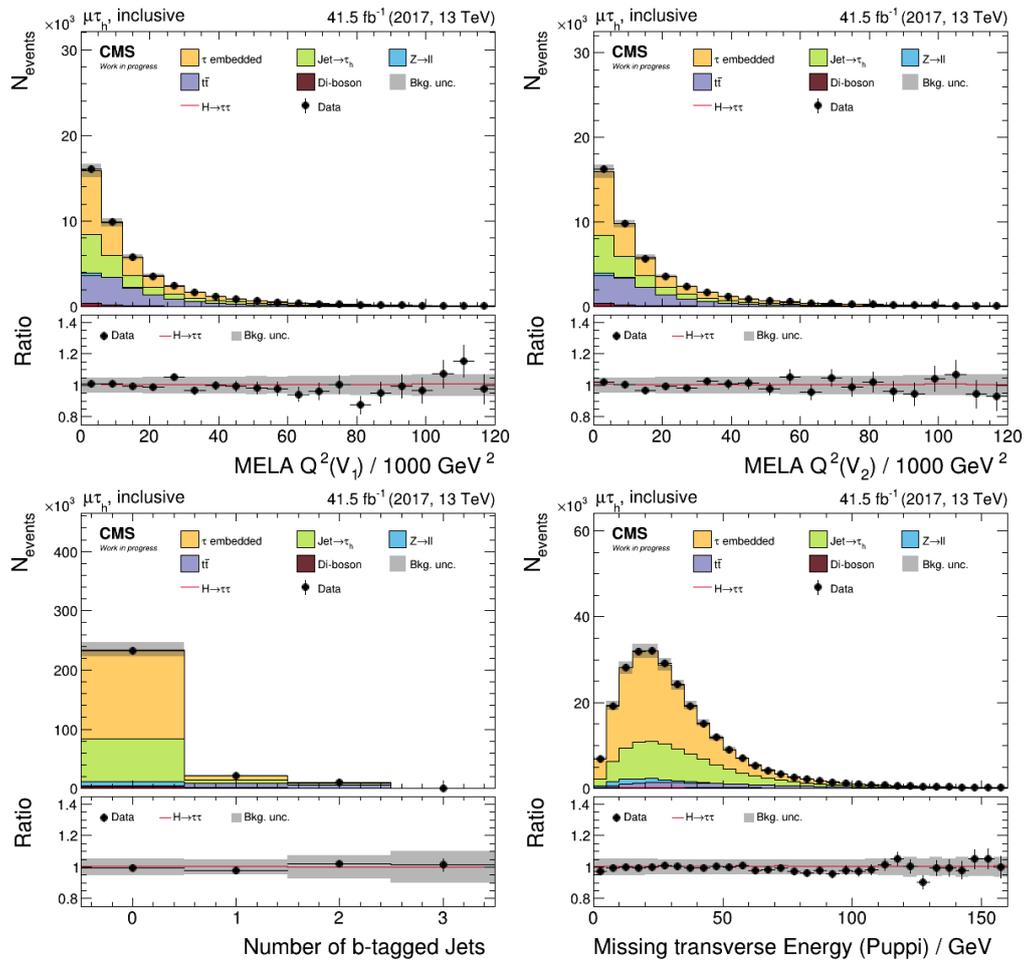
Figure A.21: Distributions of the input quantities in the 2017 *eμ* channel (III)
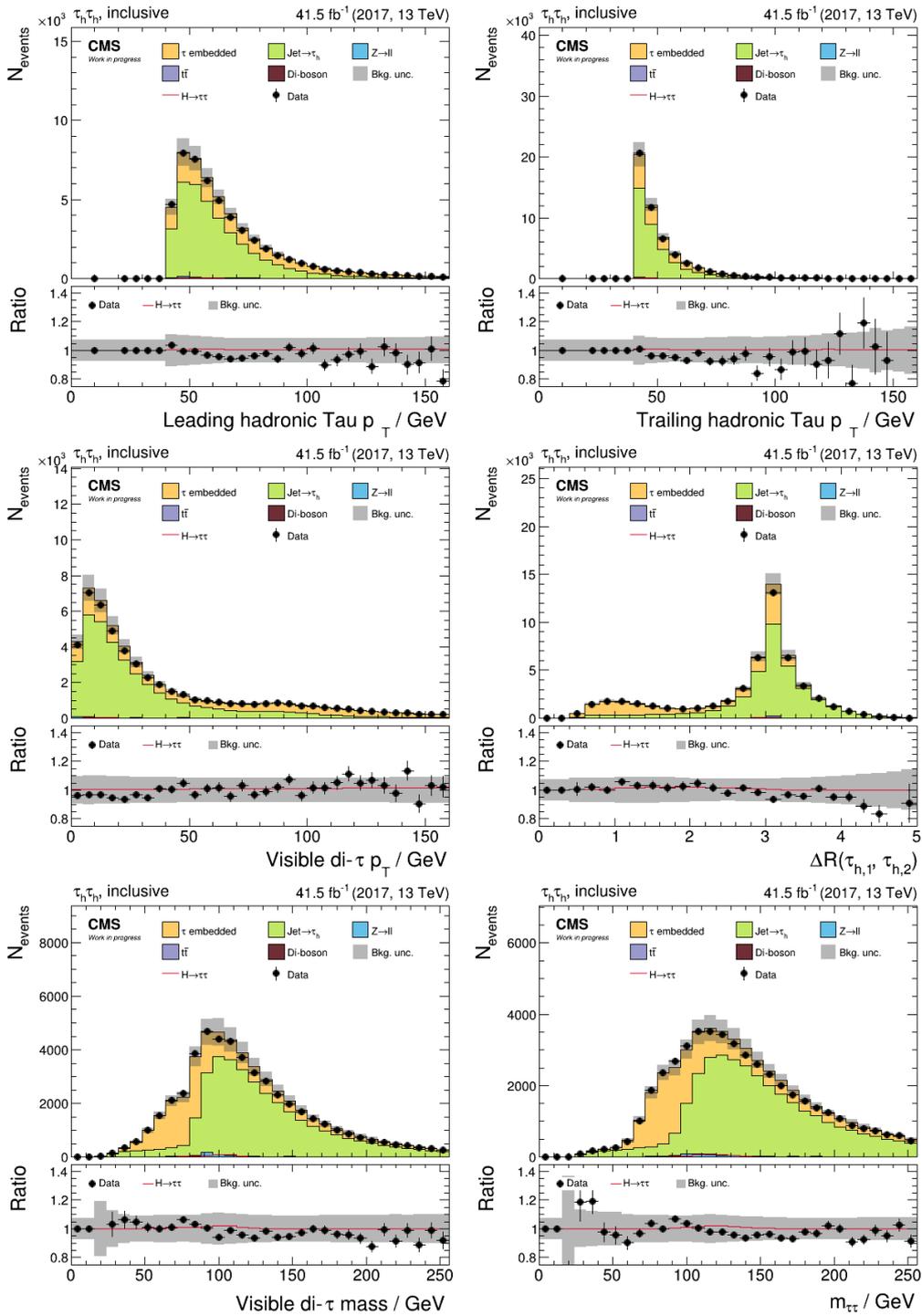
Figure A.22: Distributions of the input quantities in the 2017 $e\mu$ channel (I)
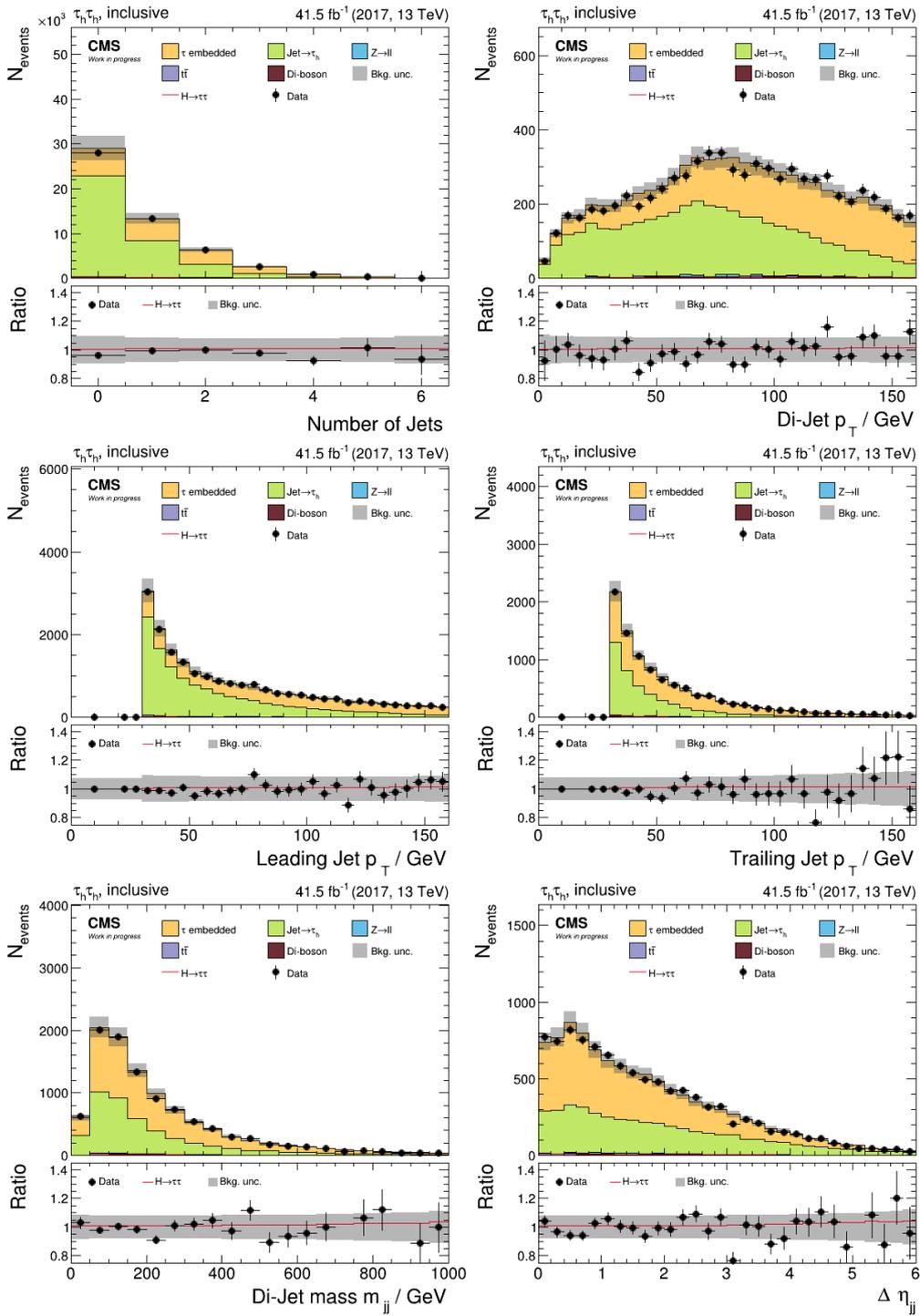
Figure A.23: Distributions of the input quantities in the 2017 $e\mu$ channel (II)
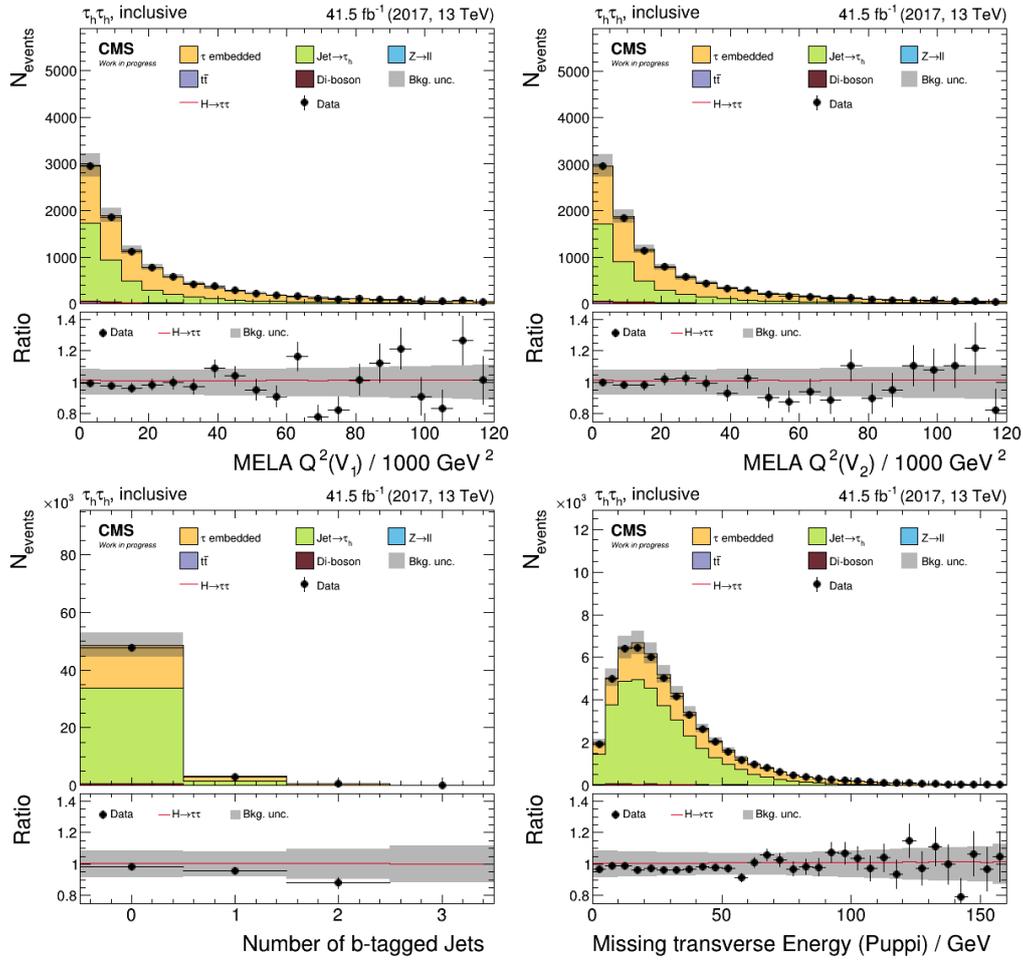
Figure A.24: Distributions of the input quantities in the 2017 $e\mu$ channel (III)
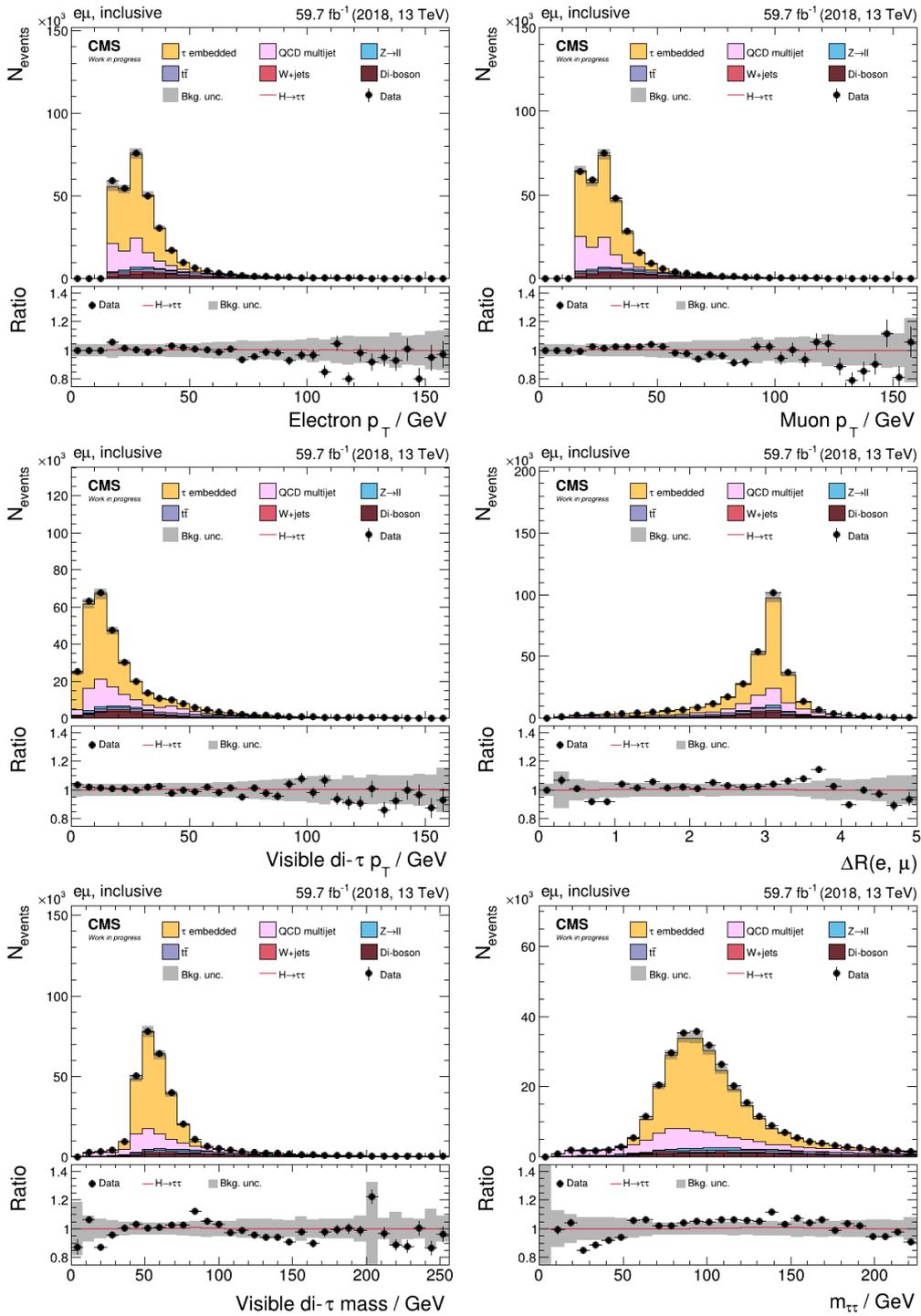
Figure A.25: Distributions of the input quantities in the 2018 $e\mu$ channel (I)
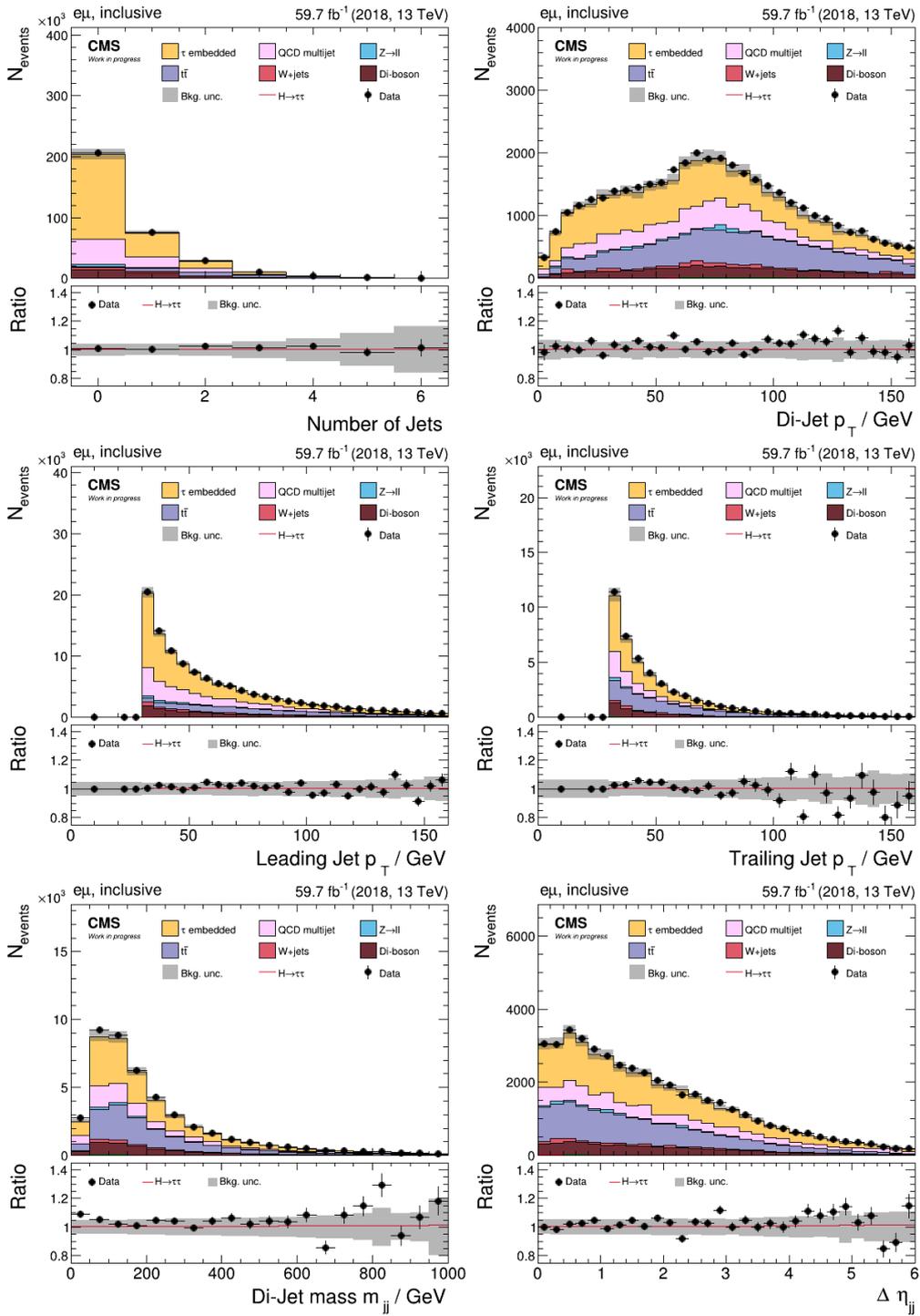
Figure A.26: Distributions of the input quantities in the 2018 $e\mu$ channel (II)
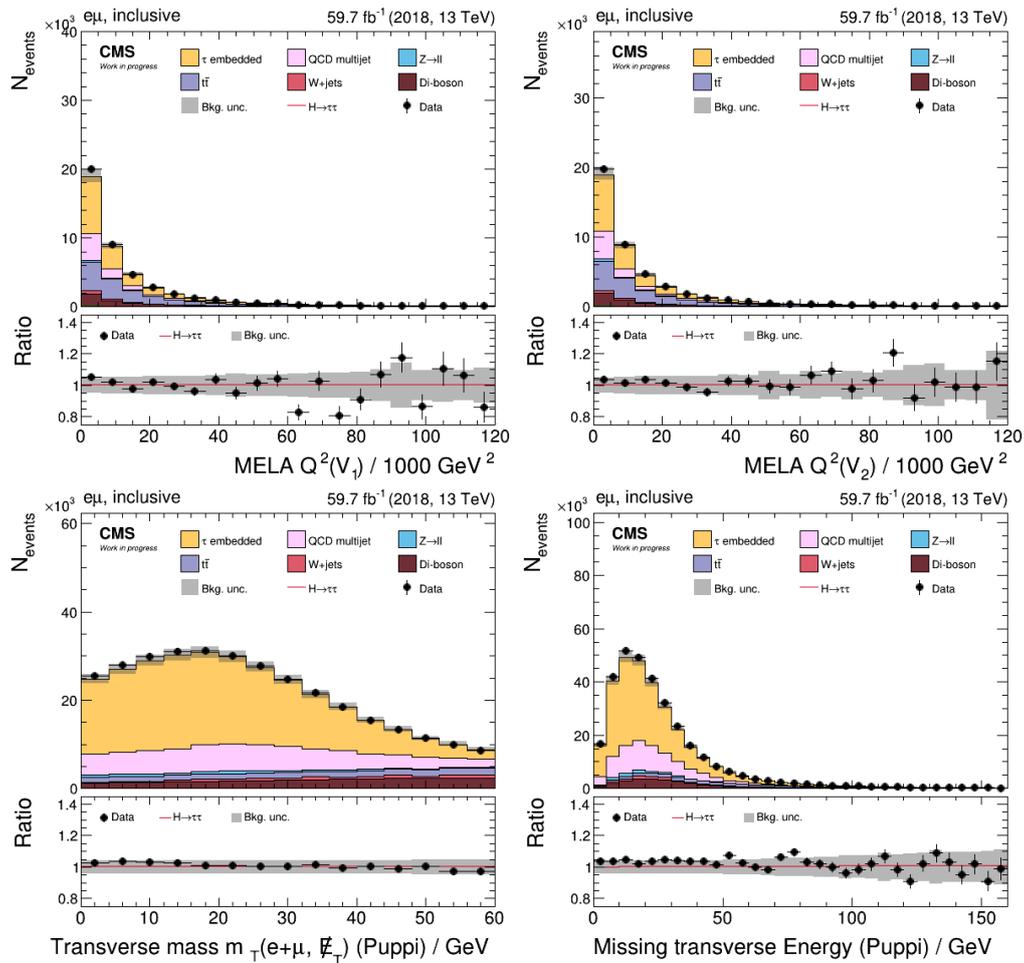
Figure A.27: Distributions of the input quantities in the 2018 $e\mu$ channel (III)
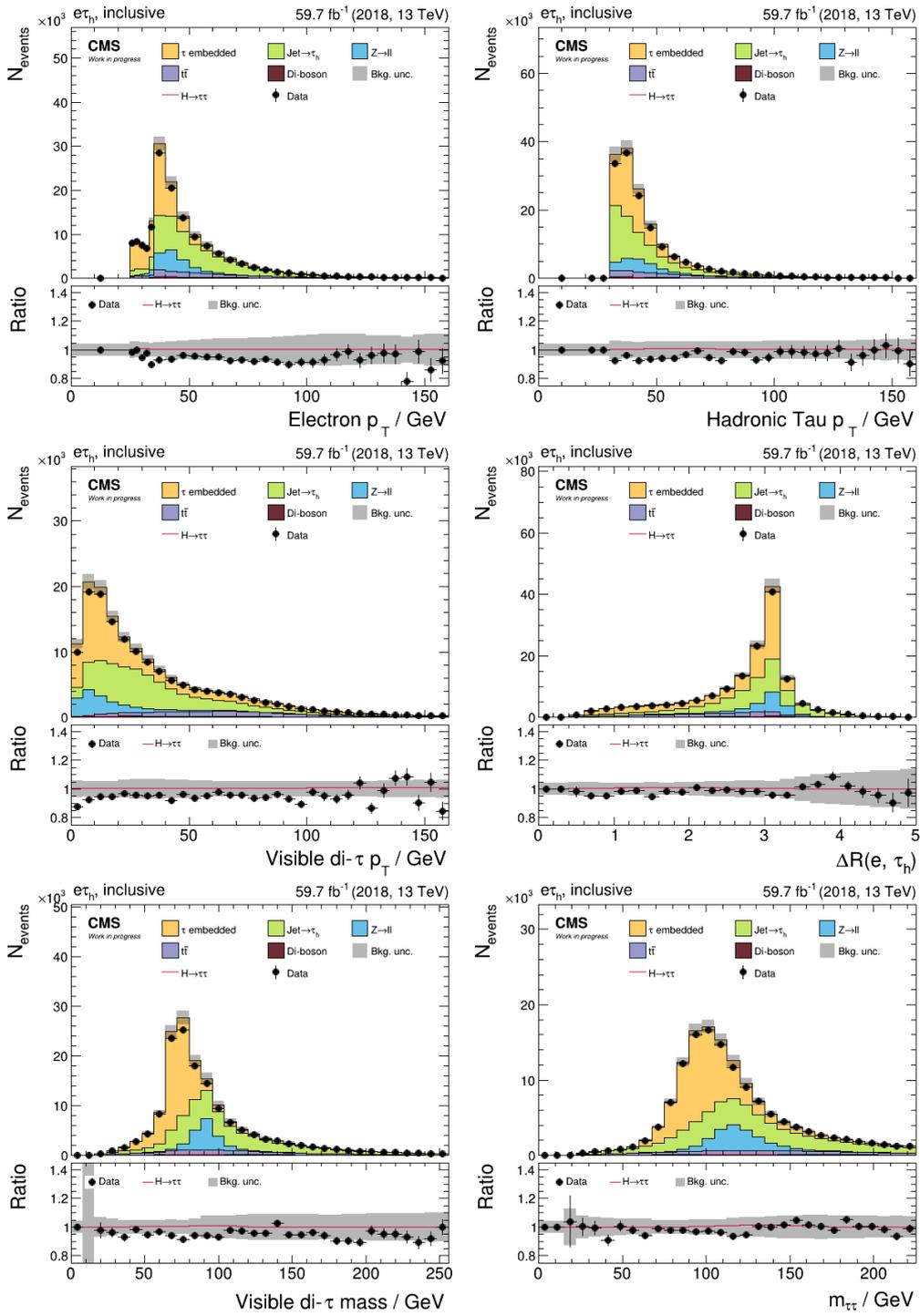
Figure A.28: Distributions of the input quantities in the 2018 $e\tau_{\mathrm{h}}$ channel (I)
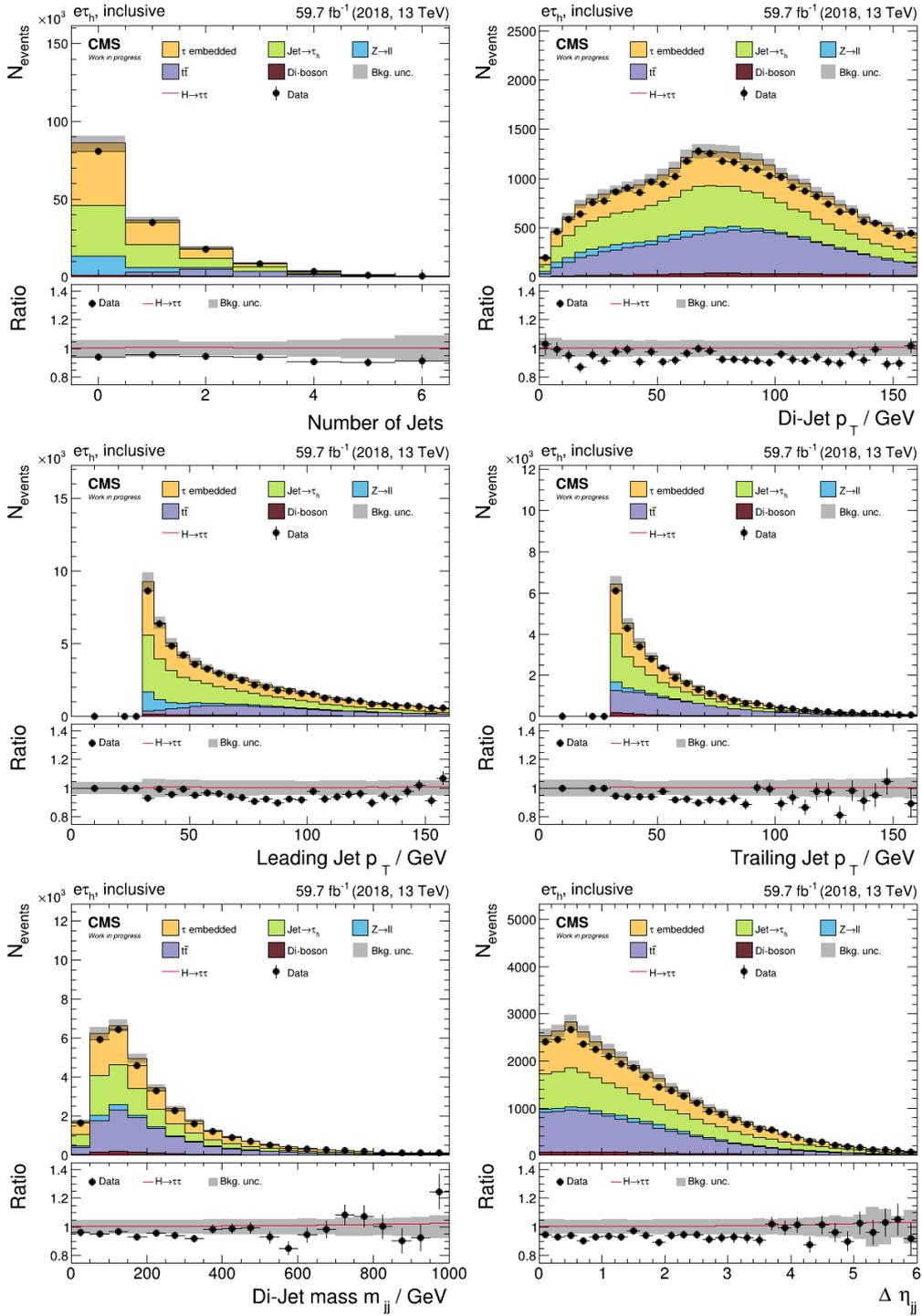
Figure A.29: Distributions of the input quantities in the 2018 $e\tau_{\mathrm{h}}$ channel (II)
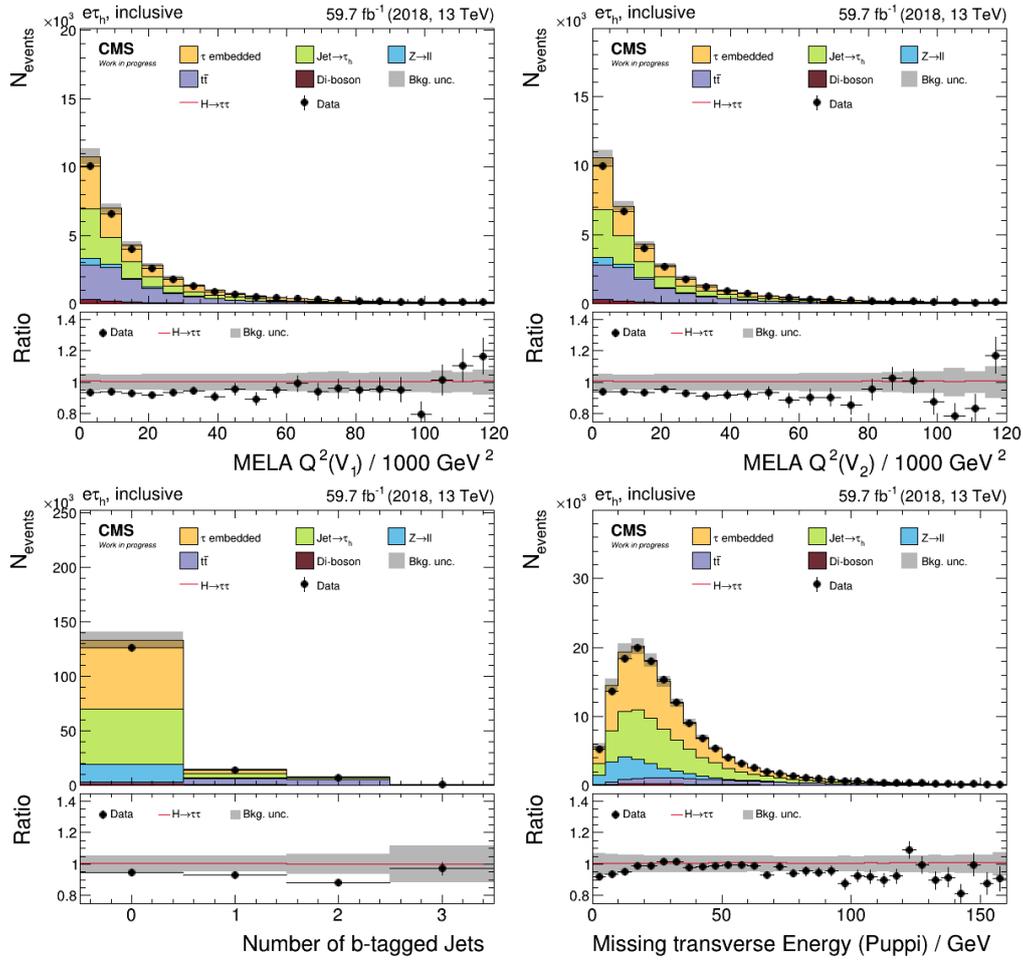
Figure A.30: Distributions of the input quantities in the 2018 $e\tau_\mathrm{h}$ channel (III)
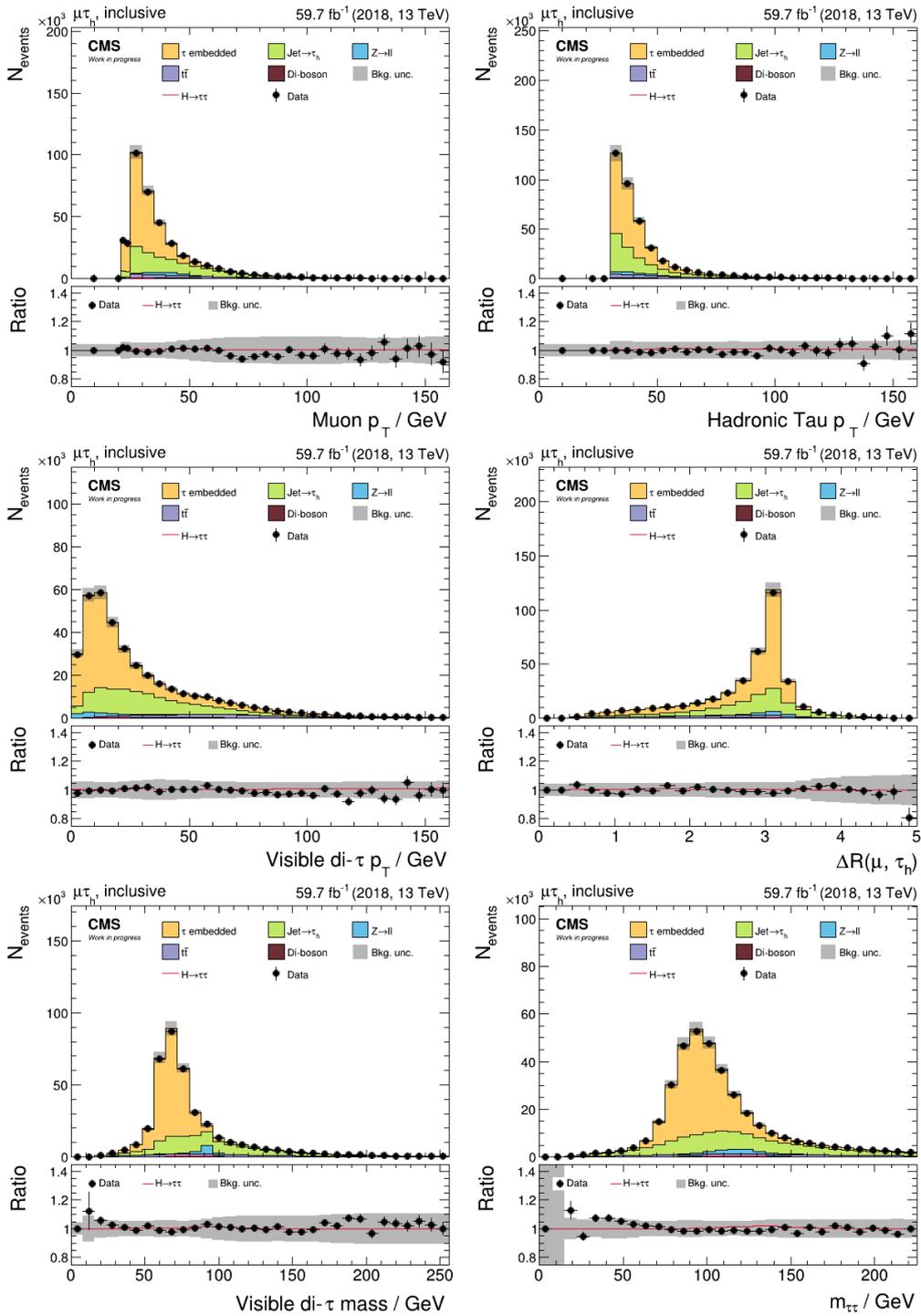
Figure A.31: Distributions of the input quantities in the 2018 $\mu\tau_h$ channel (I)
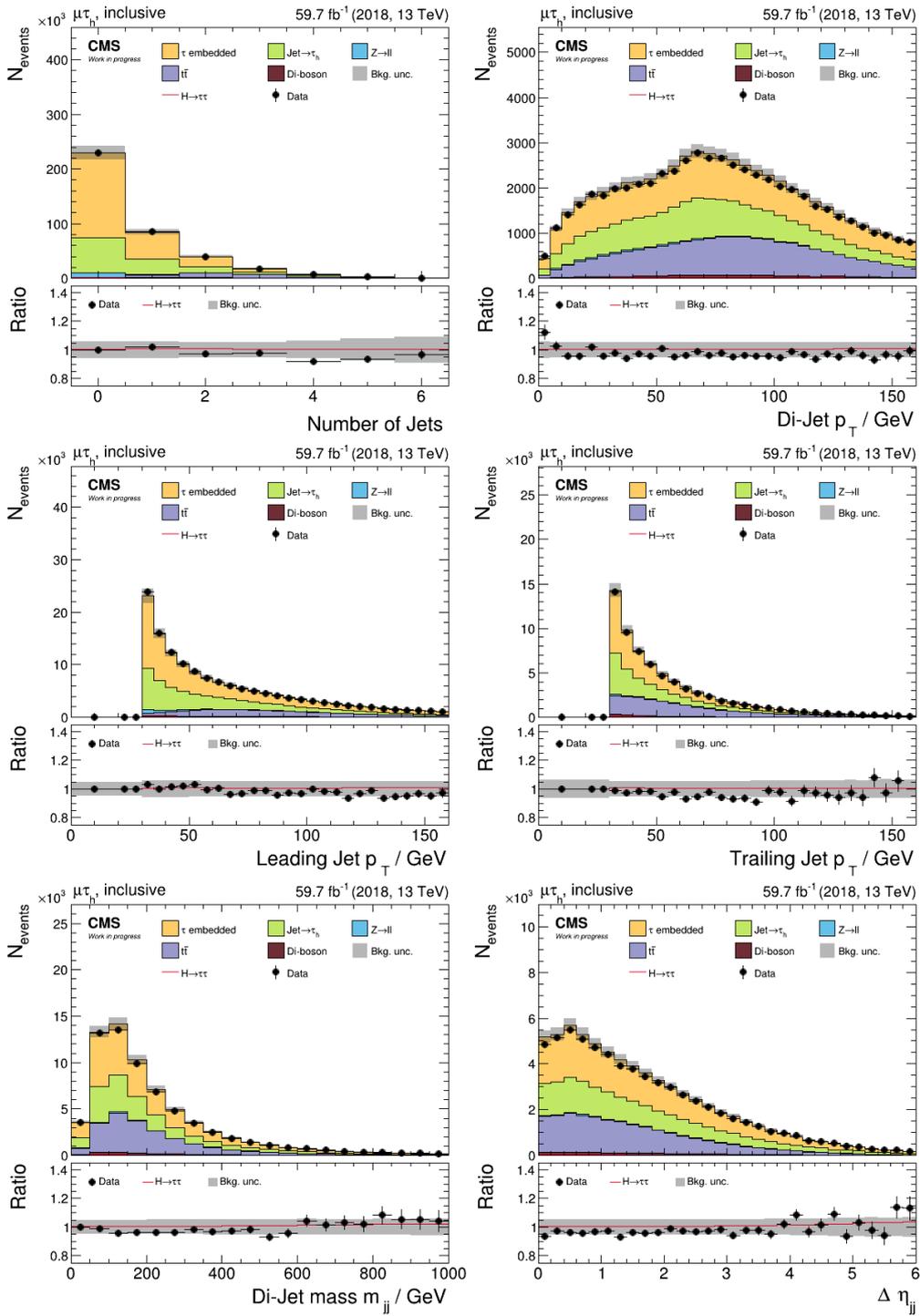
Figure A.32: Distributions of the input quantities in the 2018 $\mu\tau_h$ channel (II)
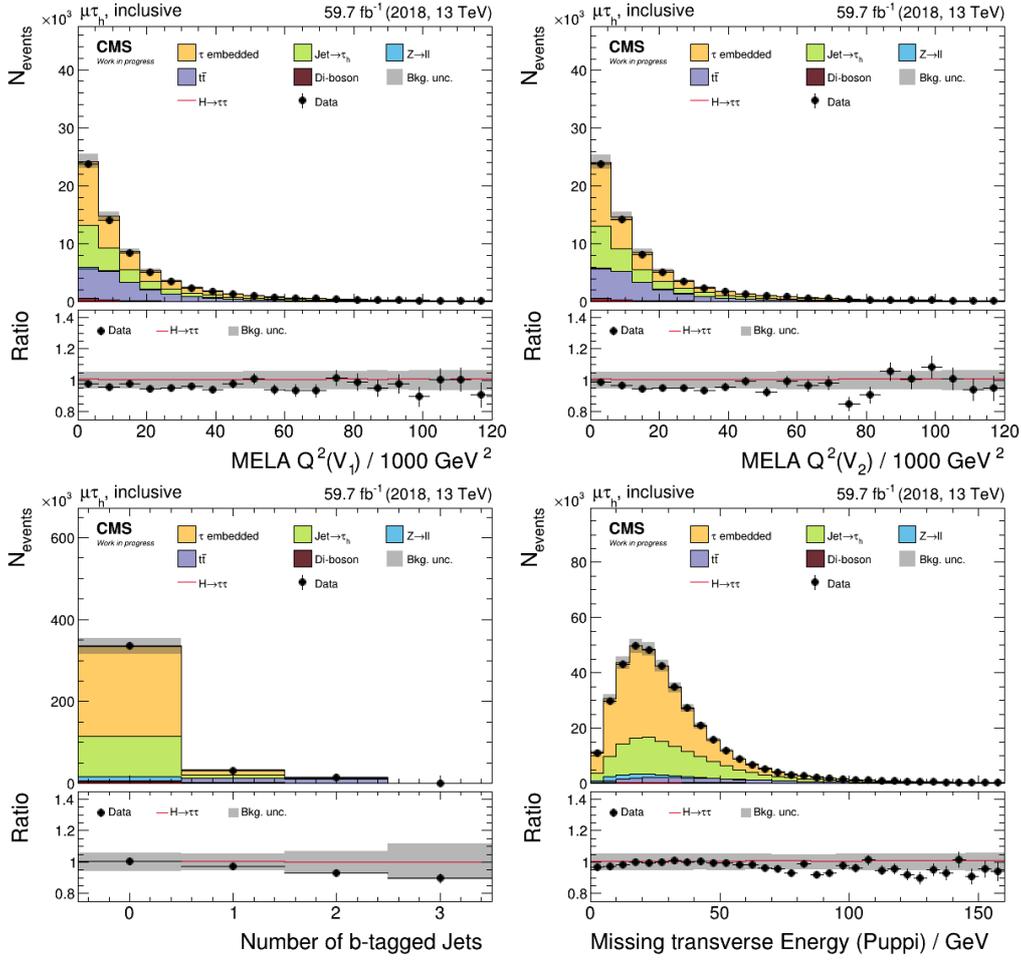
Figure A.33: Distributions of the input quantities in the 2018 $\mu\tau_h$ channel (III)
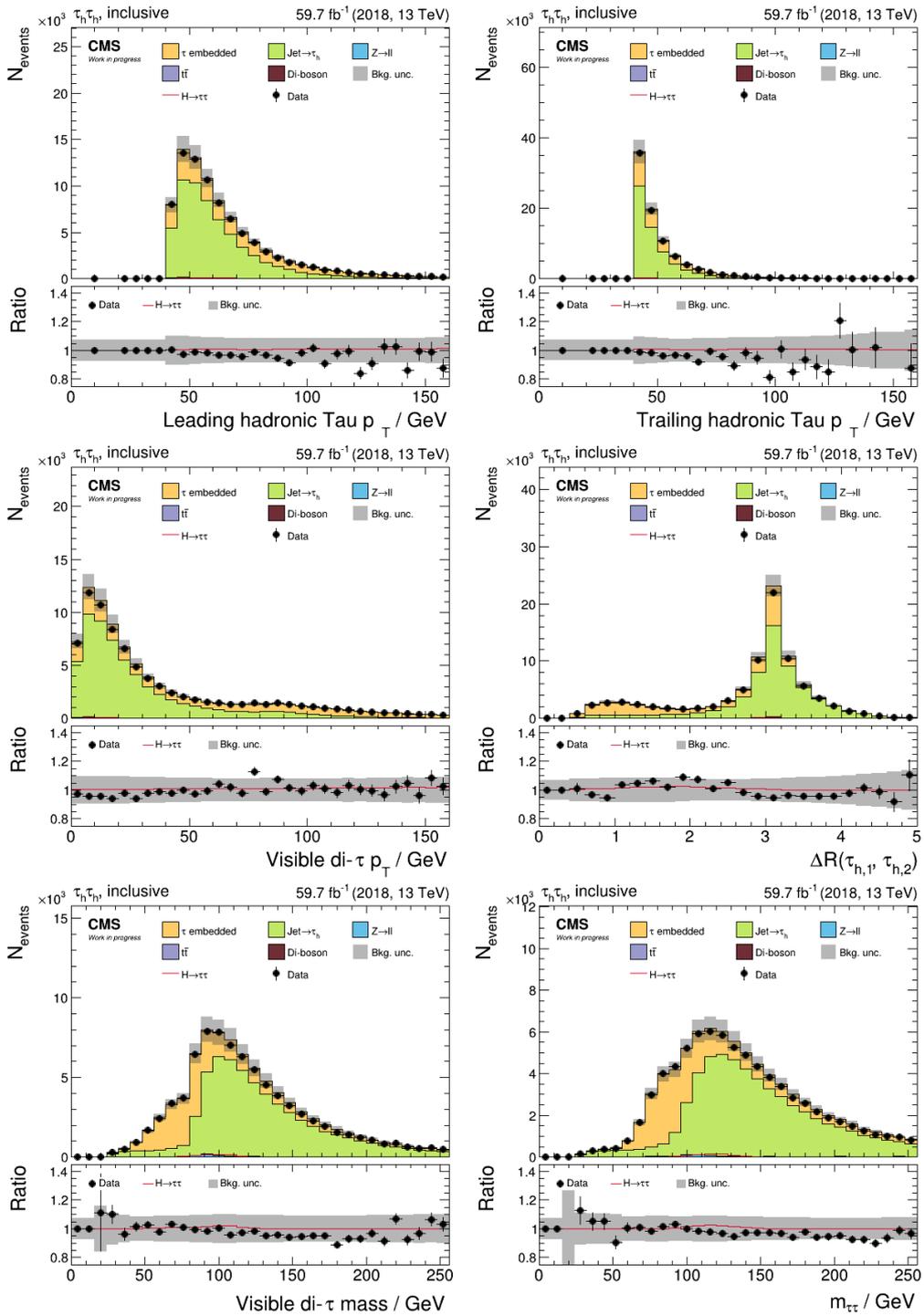
Figure A.34: Distributions of the input quantities in the 2018 $\tau_h\tau_h$ channel (I)
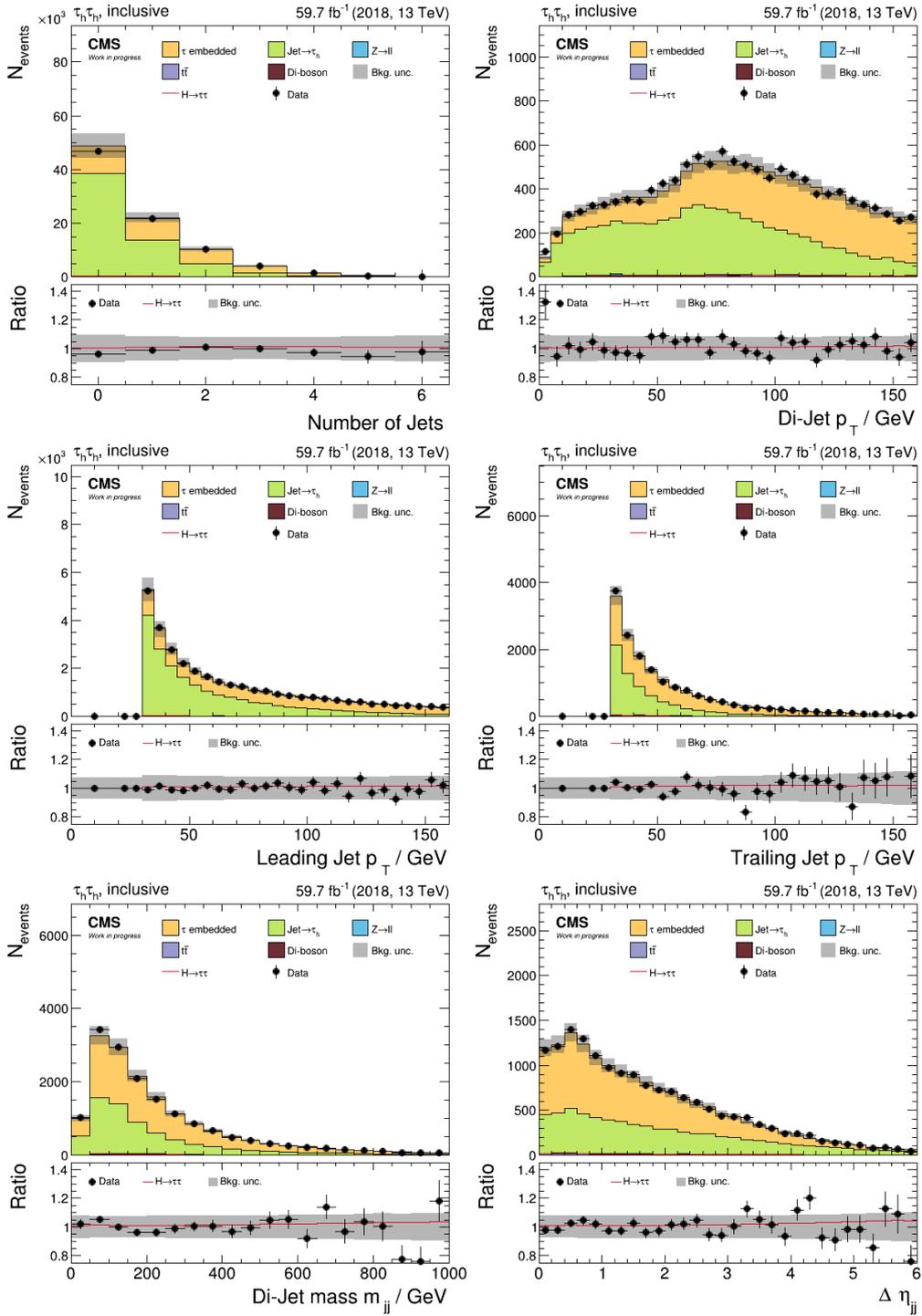
Figure A.35: Distributions of the input quantities in the 2018 $\tau_h\tau_h$ channel (II)
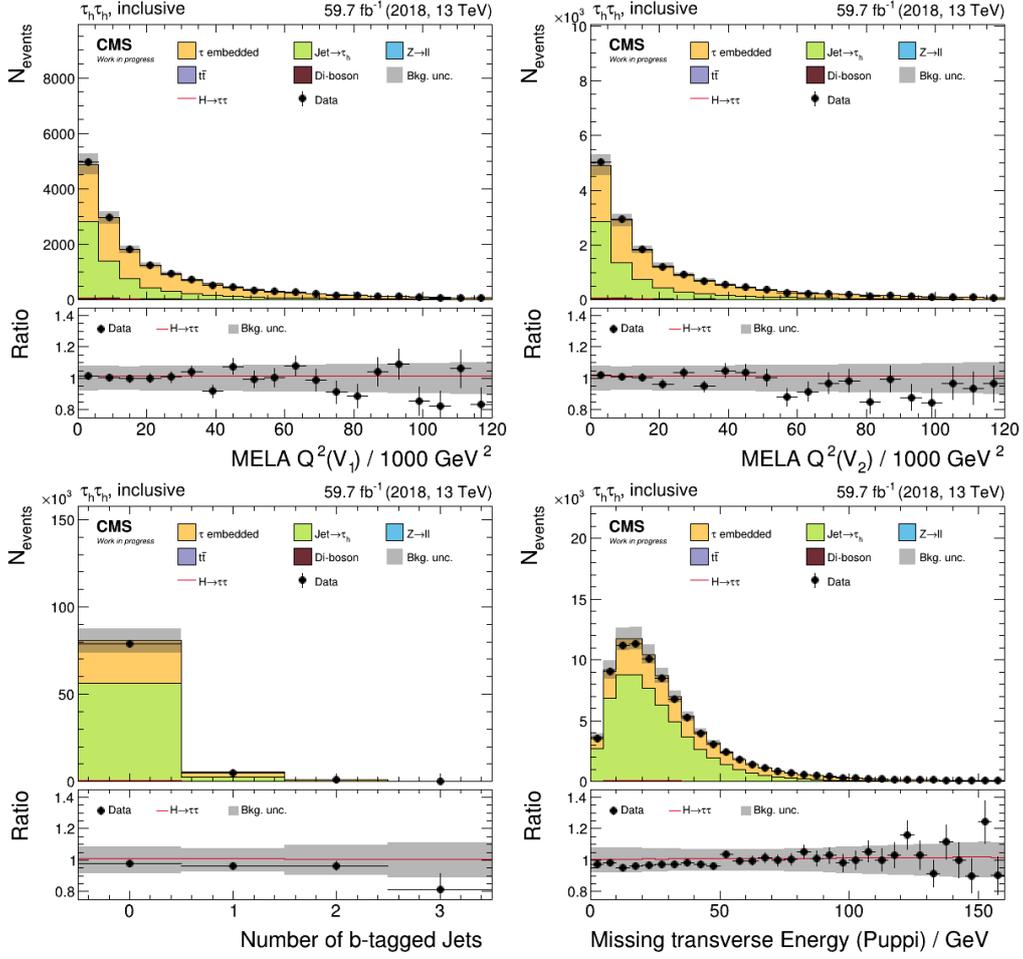
Figure A.36: Distributions of the input quantities in the 2018 $\tau_\mathrm{h}\tau_\mathrm{h}$ channel (III)

Event distributions in the neural network based categories

The distributions of the discriminators based on the NN scores are shown in the following figures. For the sake of a compact presentation, the respective distributions of different years of data-taking, which are very similar, are added together. Note however, that the years of data-taking are resolved in the statistical inference. More information about the discriminators is provided in section 5.3.

B.1 Discriminators for the inclusive and STXS stage-0 measurements
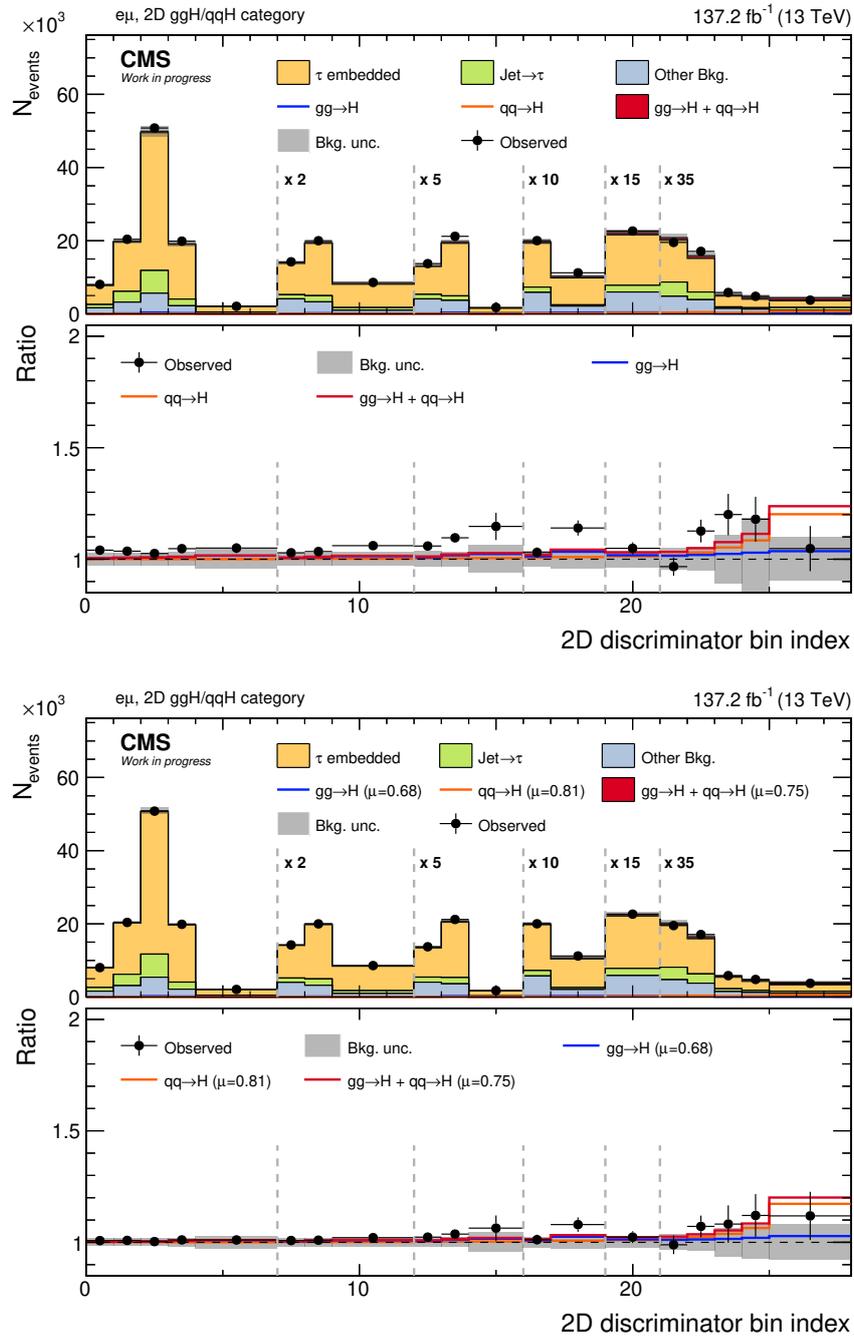
Figure B.1: STXS stage-0 signal category in the *eμ* channel. Prefit distributions are shown in the upper two panels, postfit distributions in the lower two panels.
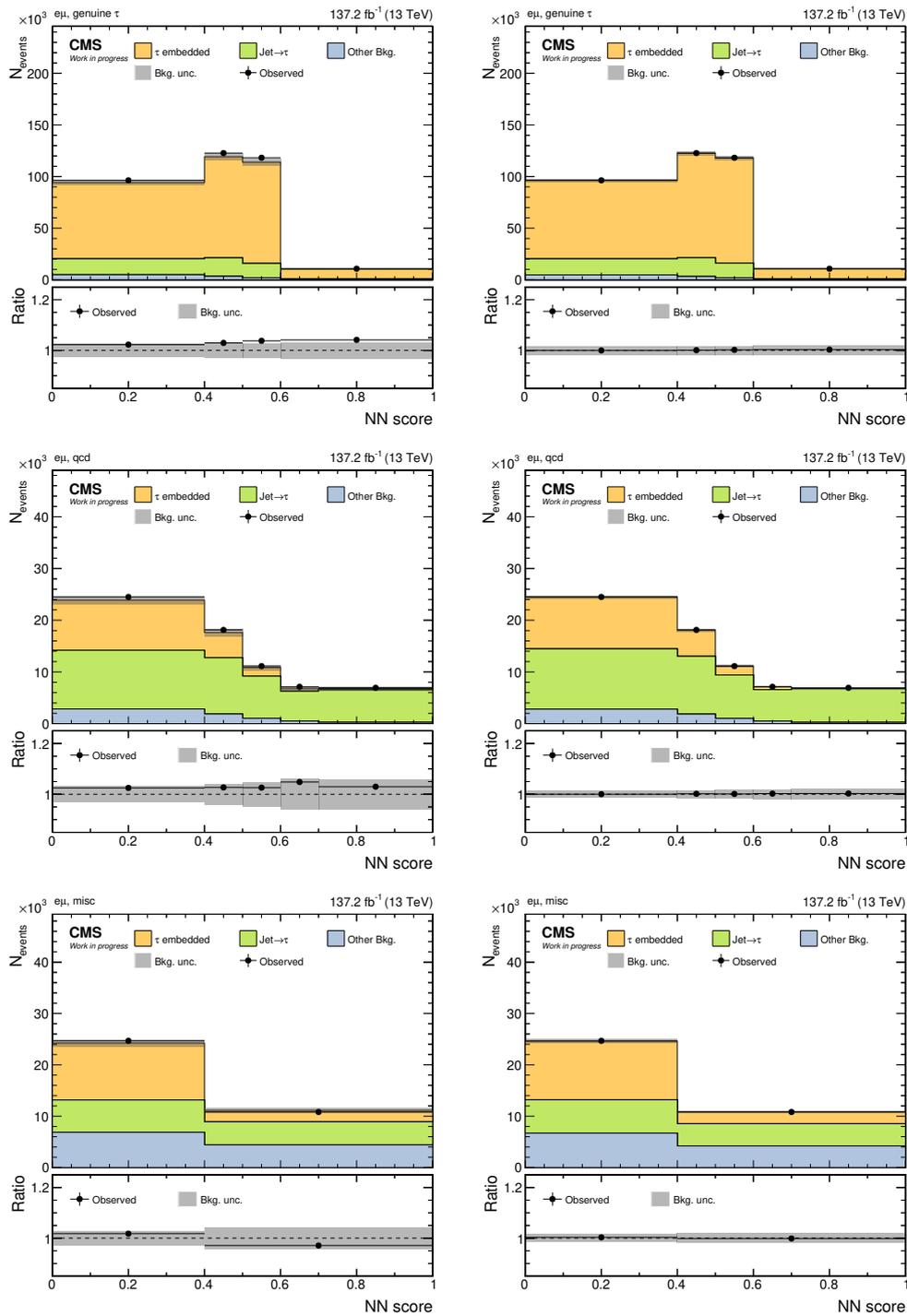
Figure B.2: STXS stage-0 background categories in the $e\mu$ channel (I). Prefit distributions are shown on the left, postfit distributions on the right.
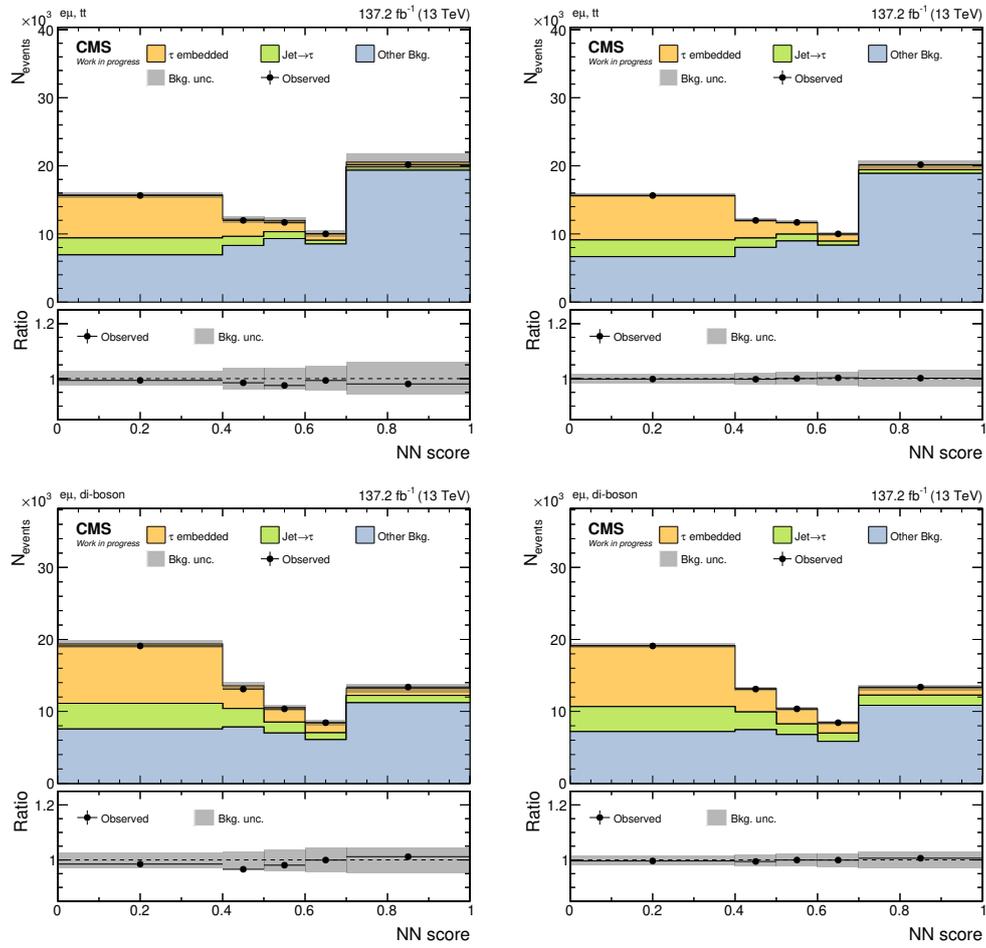
Figure B.3: STXS stage-0 background categories in the $e\mu$ channel (II). Prefit distributions are shown on the left, postfit distributions on the right.
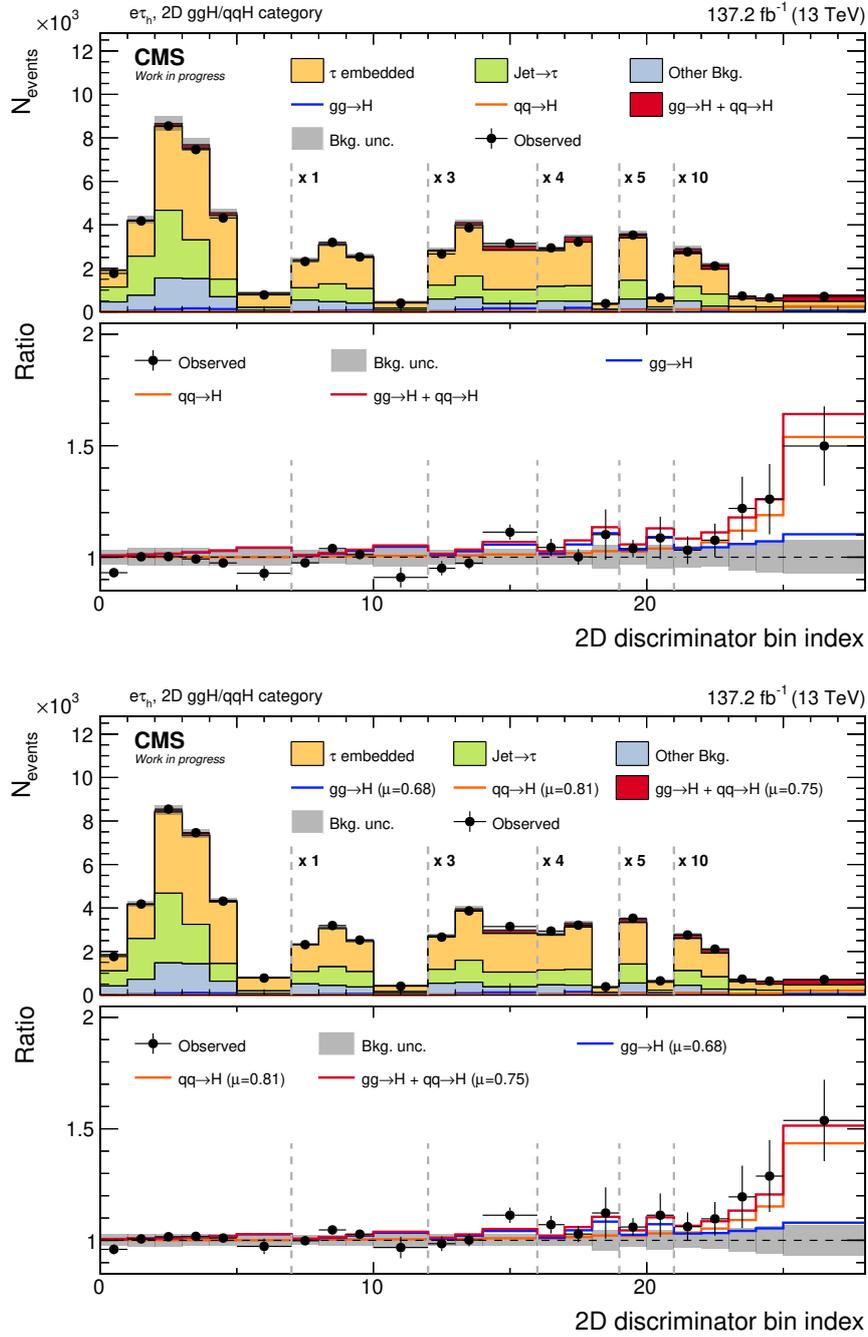
Figure B.4: STXS stage-0 signal category in the $e\tau_\mathrm{h}$ channel. Prefit distributions are shown in the upper two panels, postfit distributions in the lower two panels.
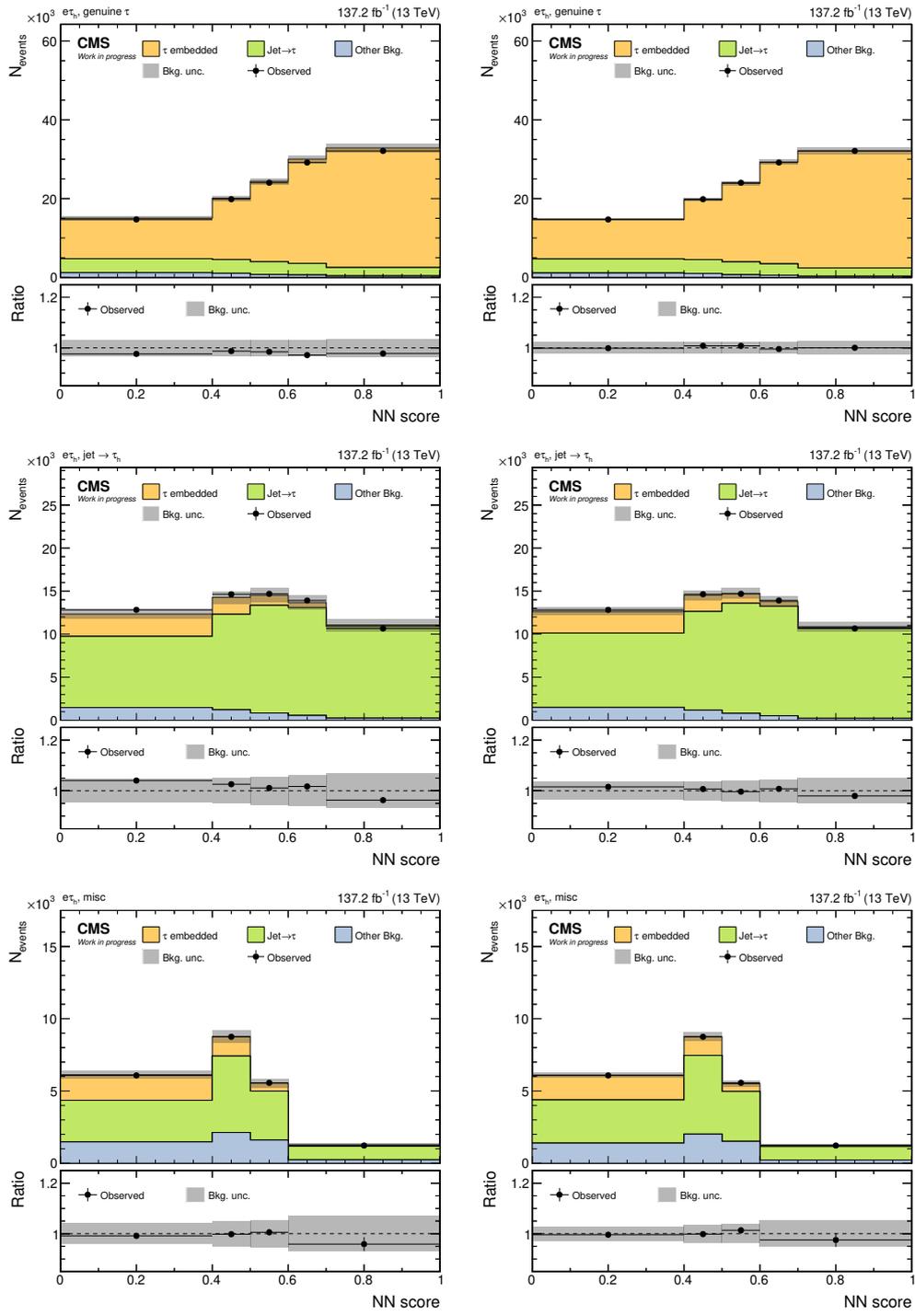
Figure B.5: STXS stage-0 background categories in the $e\tau_\mathrm{h}$ channel (I). Prefit distributions are shown on the left, postfit distributions on the right.
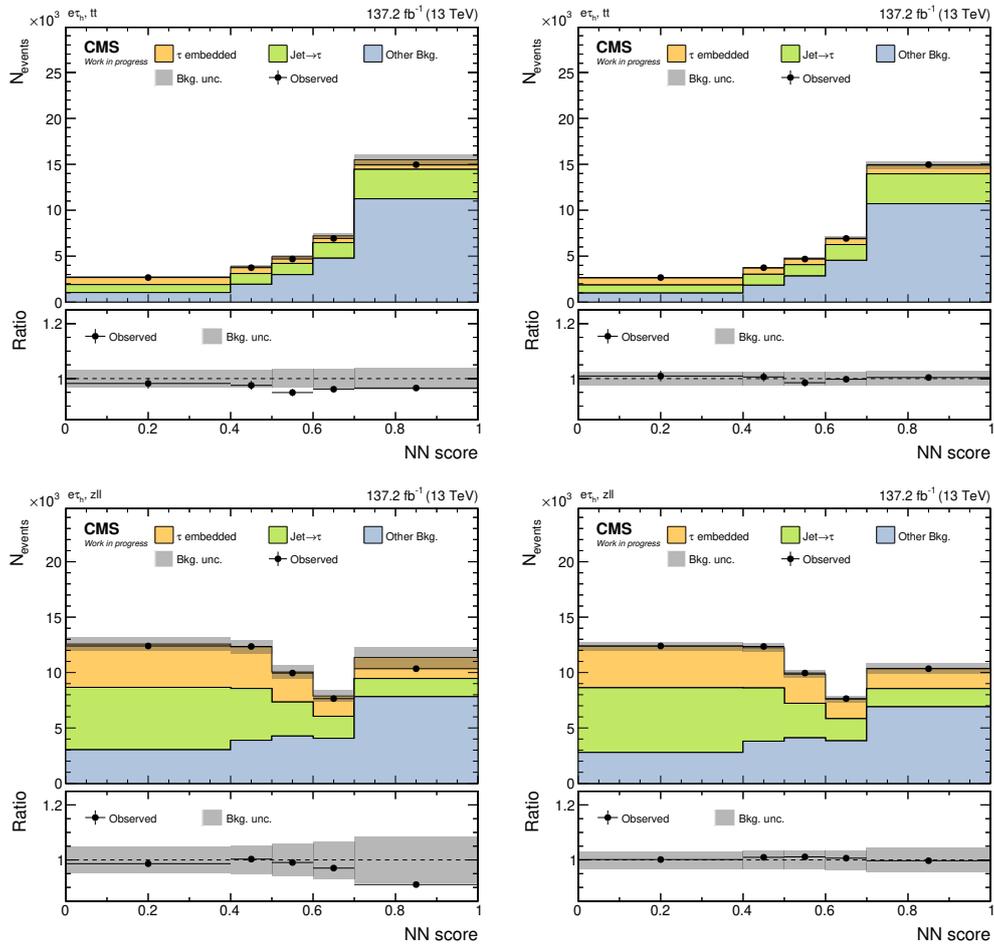
Figure B.6: STXS stage-0 background categories in the $e\tau_h$ channel (II). Prefit distributions are shown on the left, postfit distributions on the right.
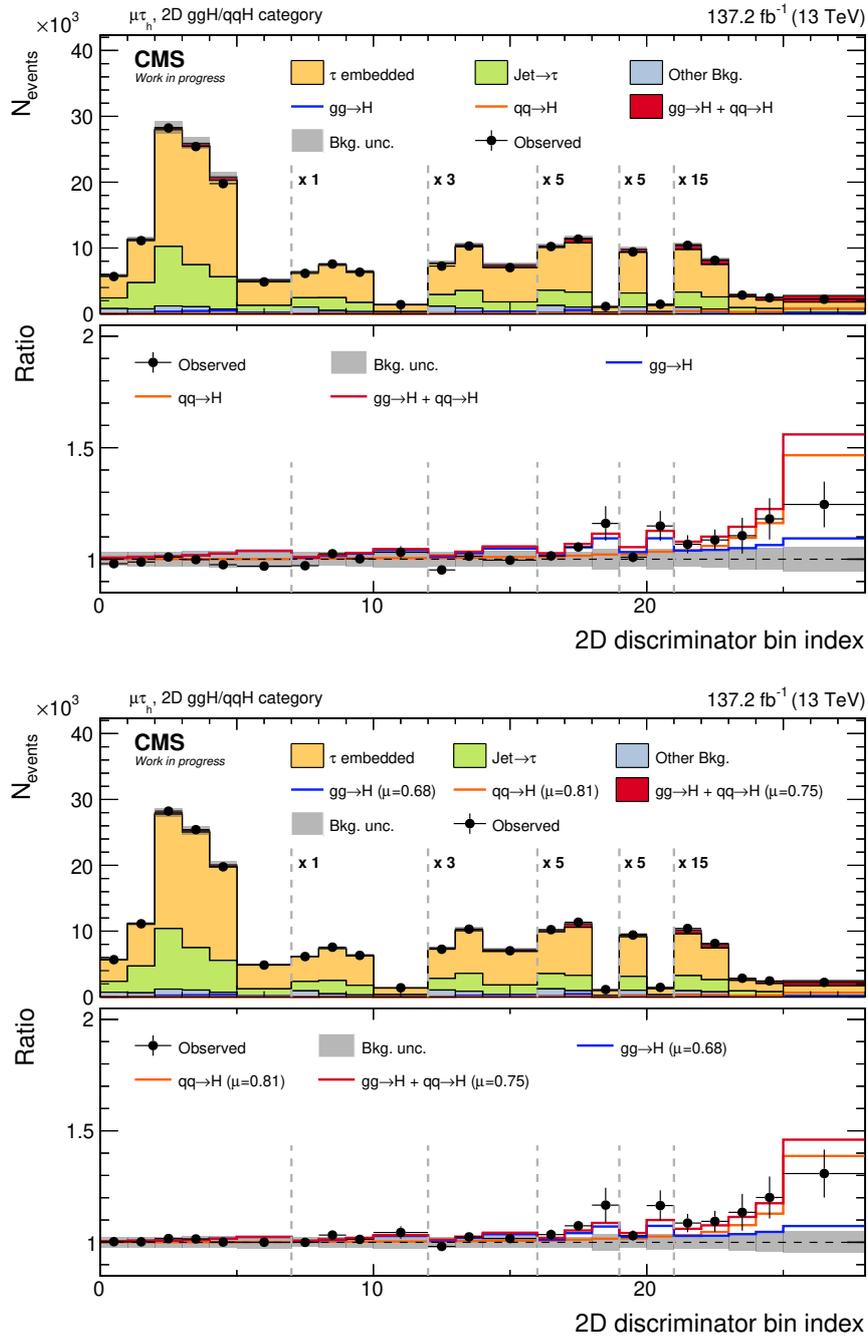
Figure B.7: STXS stage-0 signal category in the $\mu\tau_\mathrm{h}$ channel. Prefit distributions are shown in the upper two panels, postfit distributions in the lower two panels.
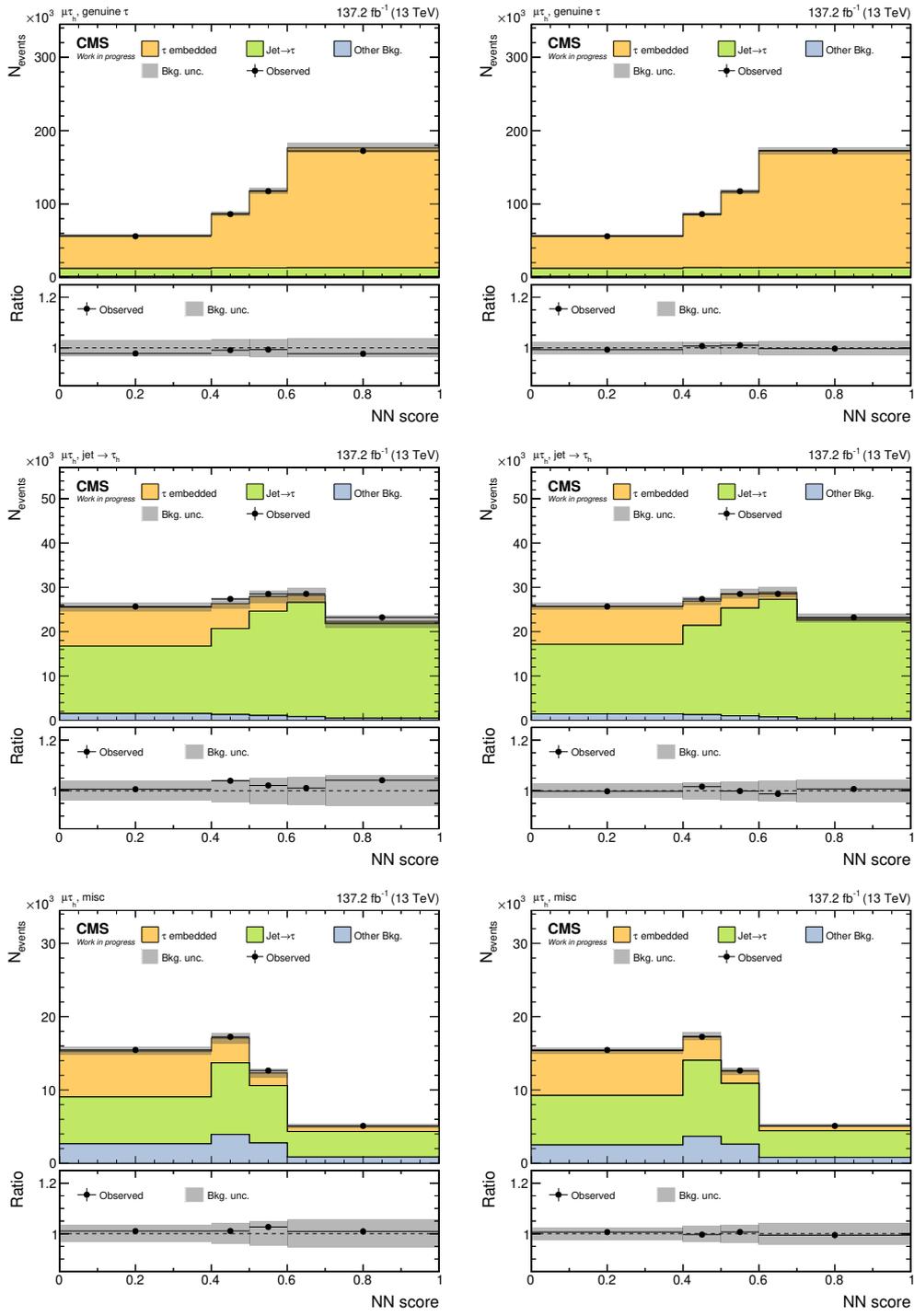
Figure B.8: STXS stage-0 background categories in the $\mu\tau_\mathrm{h}$ channel (I). Prefit distributions are shown on the left, postfit distributions on the right.
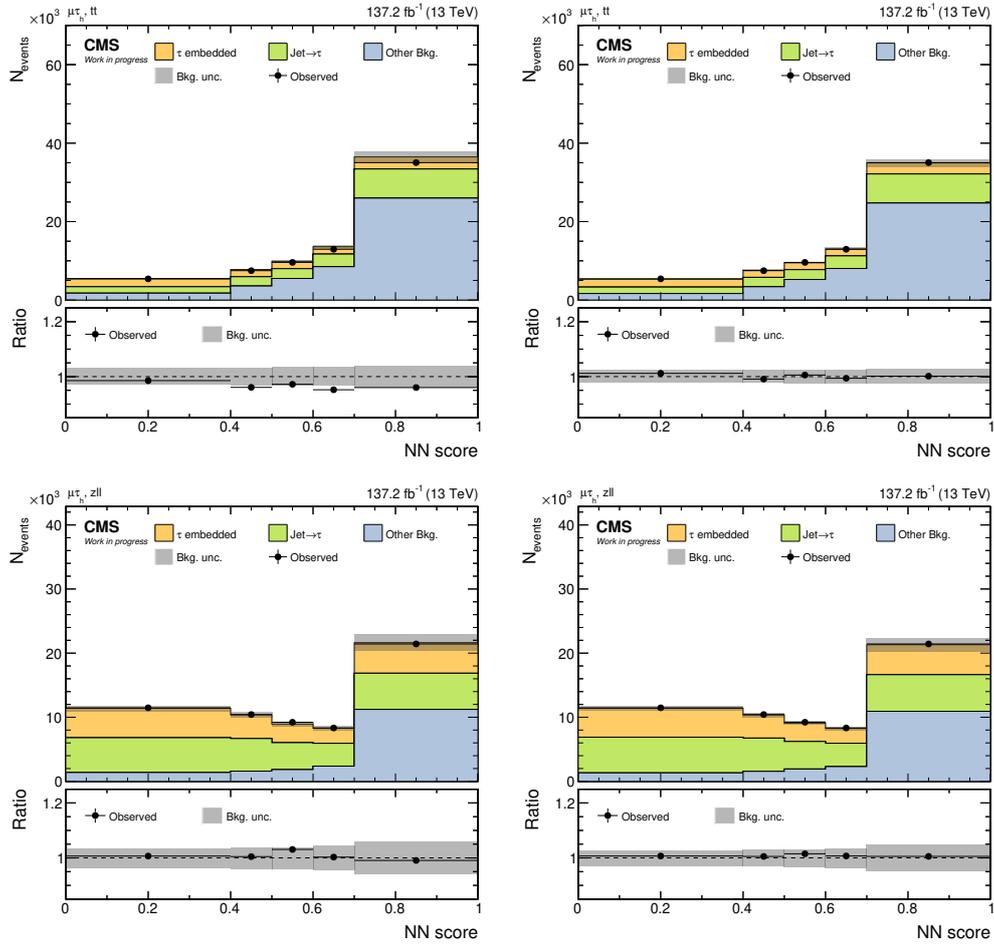
Figure B.9: STXS stage-0 background categories in the $\mu\tau_\text{h}$ channel (II). Prefit distributions are shown on the left, postfit distributions on the right.
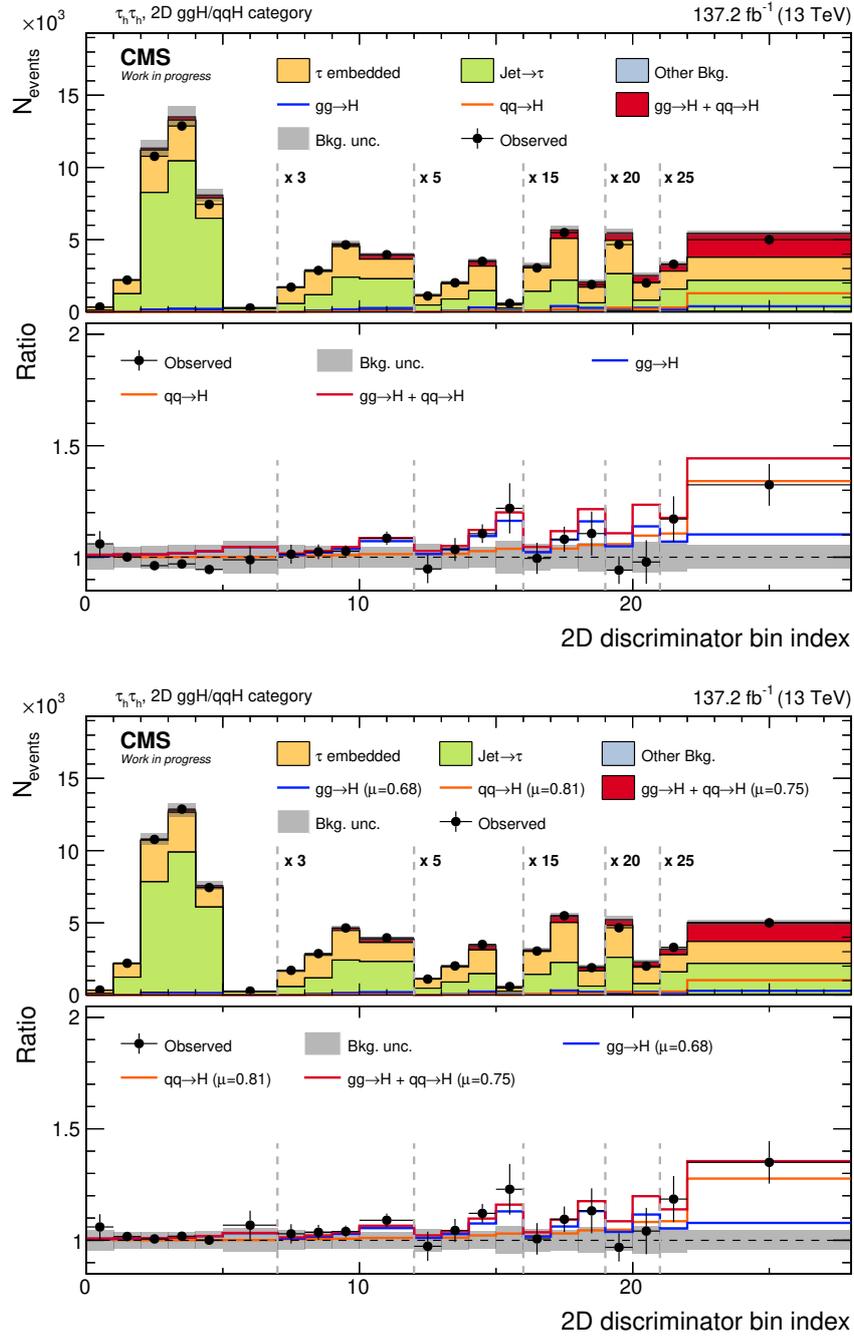
Figure B.10: STXS stage-0 signal category in the $\tau_h\tau_h$ channel. Prefit distributions are shown in the upper two panels, postfit distributions in the lower two panels.
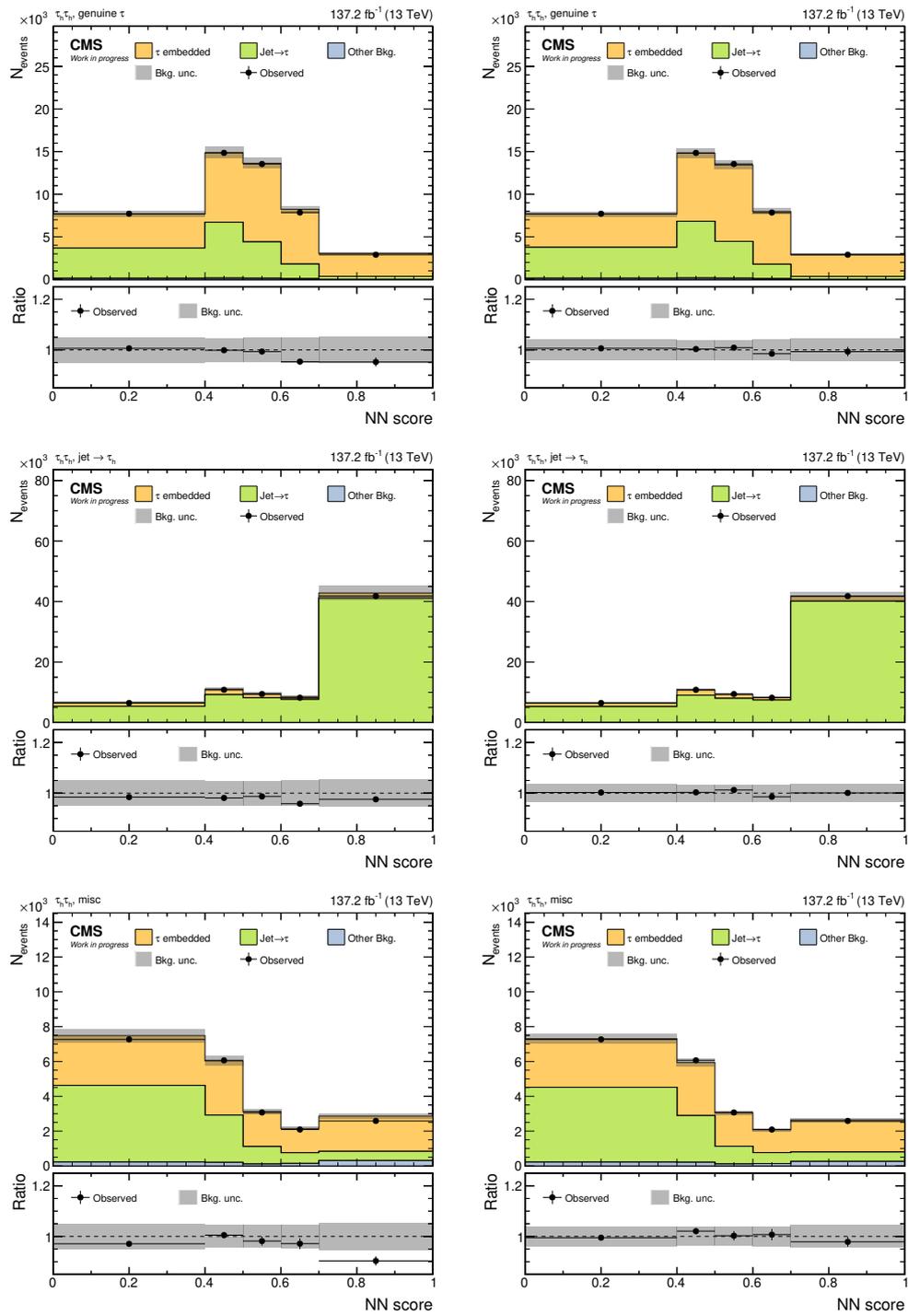
Figure B.11: STXS stage-0 background categories in the $\tau_h\tau_h$ channel. Prefit distributions are shown on the left, postfit distributions on the right.

B.2 Discriminators for the STXS stage-1 measurements
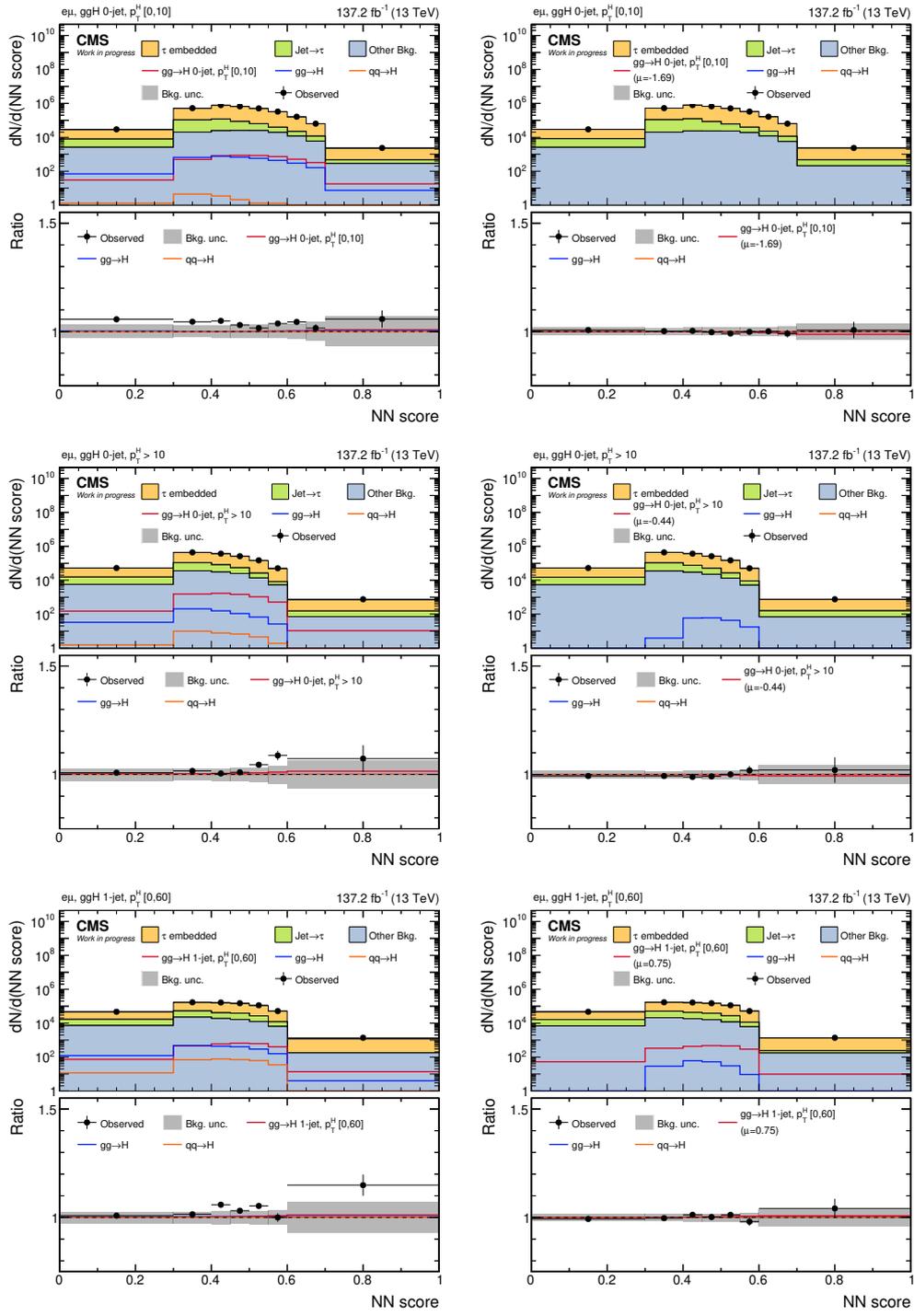
Figure B.12: STXS stage-1 categories in the $e\mu$ channel (I). Prefit distributions are shown on the left, postfit distributions on the right.
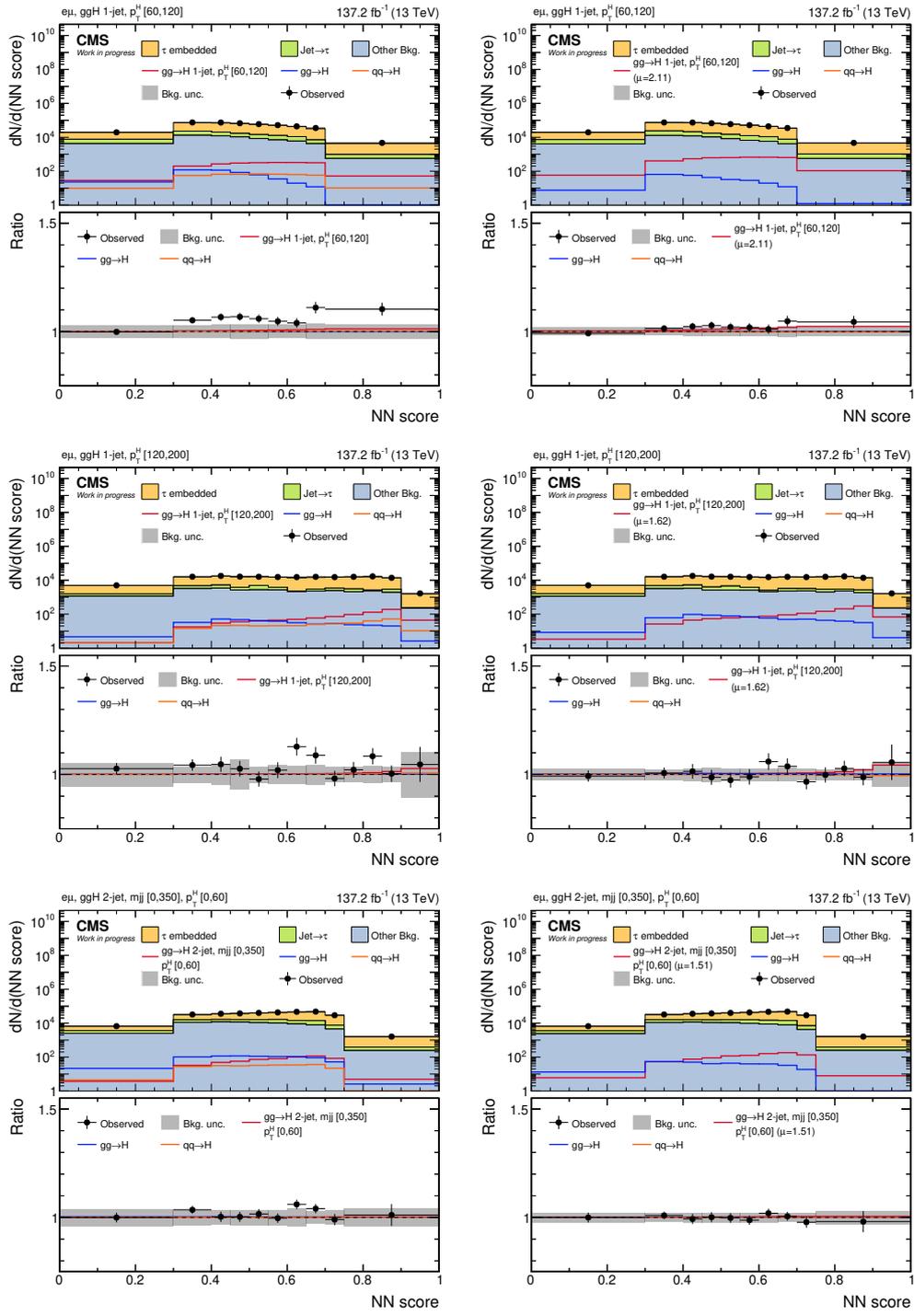
Figure B.13: STXS stage-1 categories in the $e\mu$ channel (II). Prefit distributions are shown on the left, postfit distributions on the right.
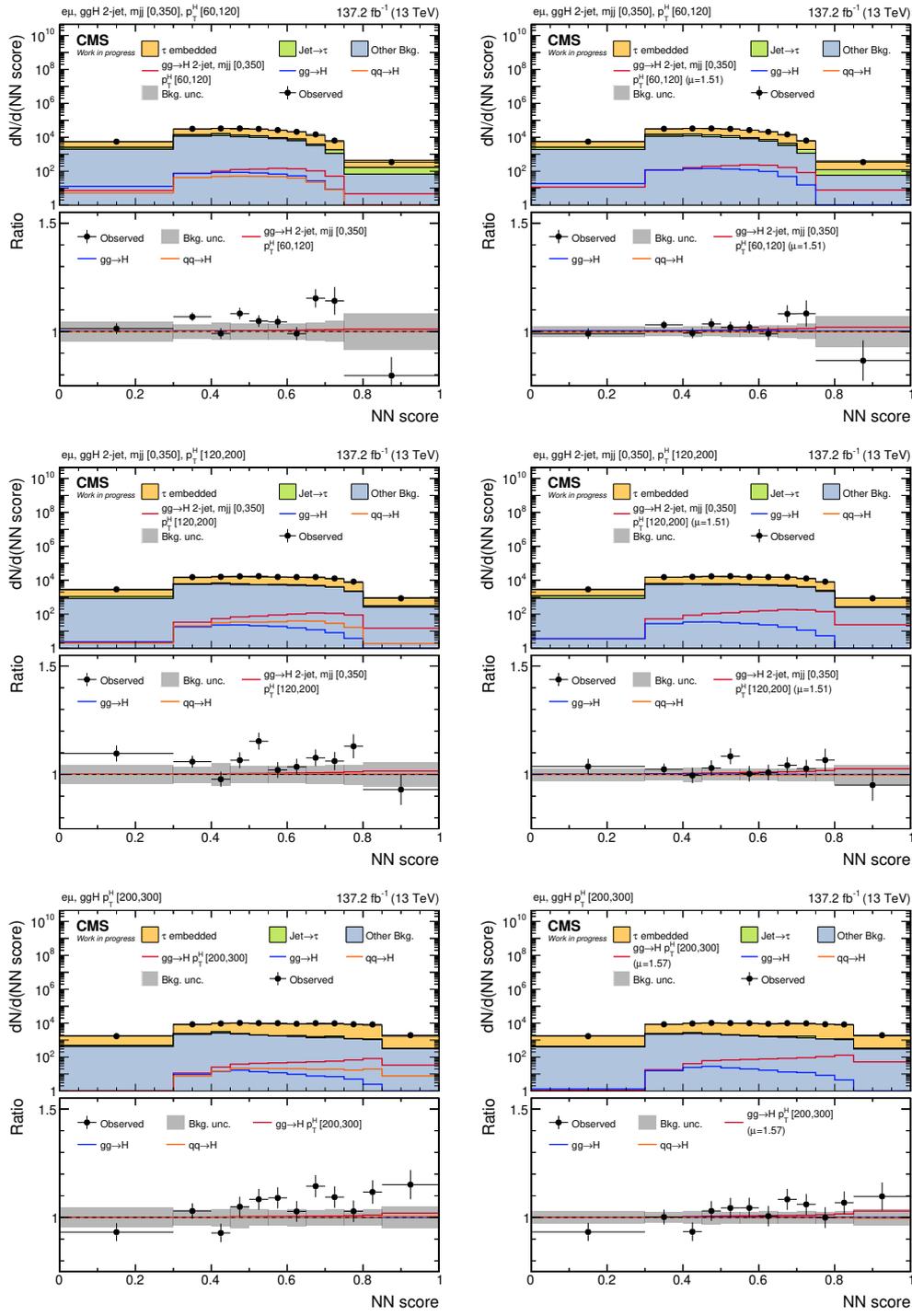
Figure B.14: STXS stage-1 categories in the *eμ* channel (III). Prefit distributions are shown on the left, postfit distributions on the right.

Figure B.15: STXS stage-1 categories in the $e\mu$ channel (IV). Prefit distributions are shown on the left, postfit distributions on the right.
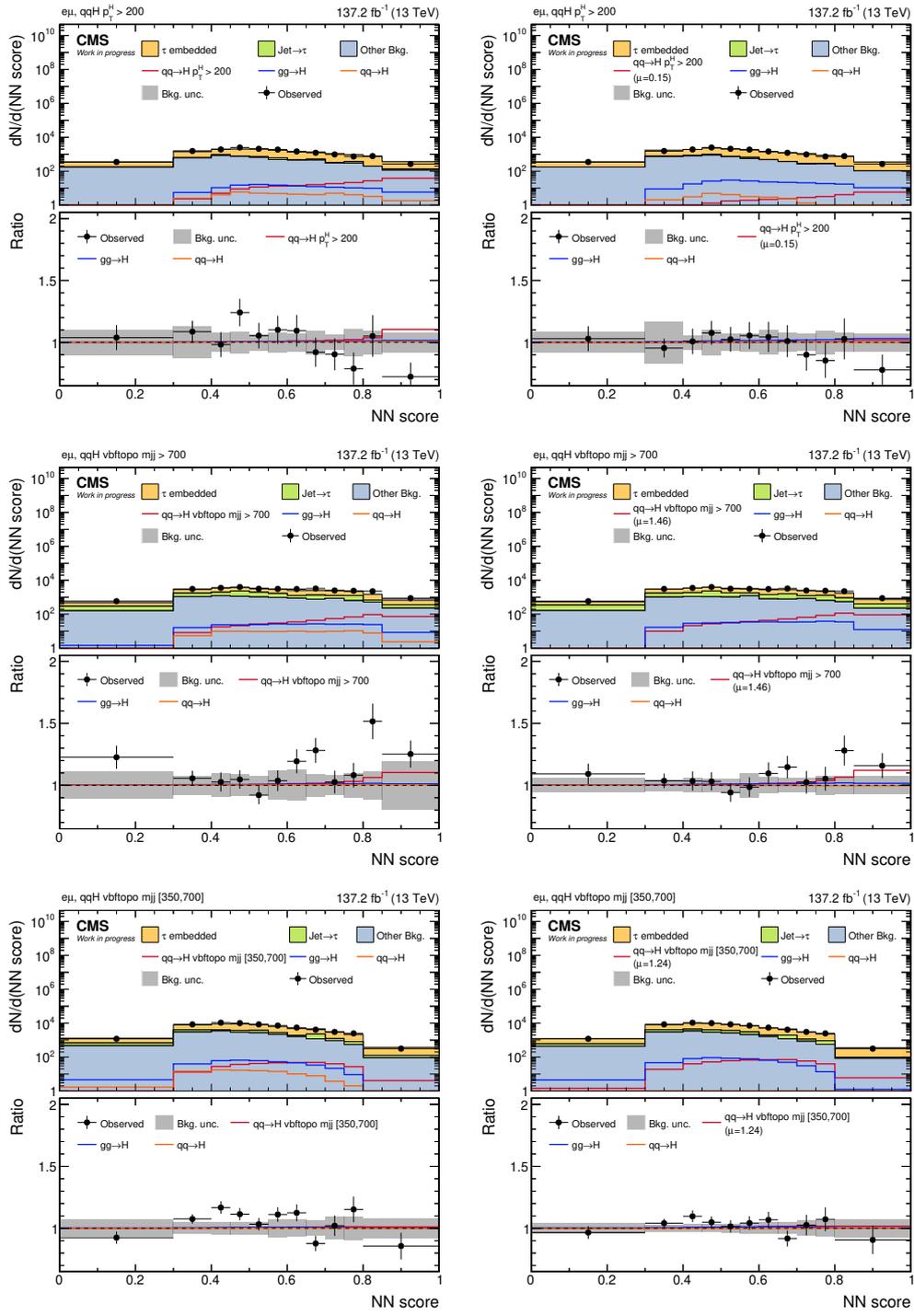
Figure B.16: STXS stage-1 categories in the $e\mu$ channel (V). Prefit distributions are shown on the left, postfit distributions on the right.
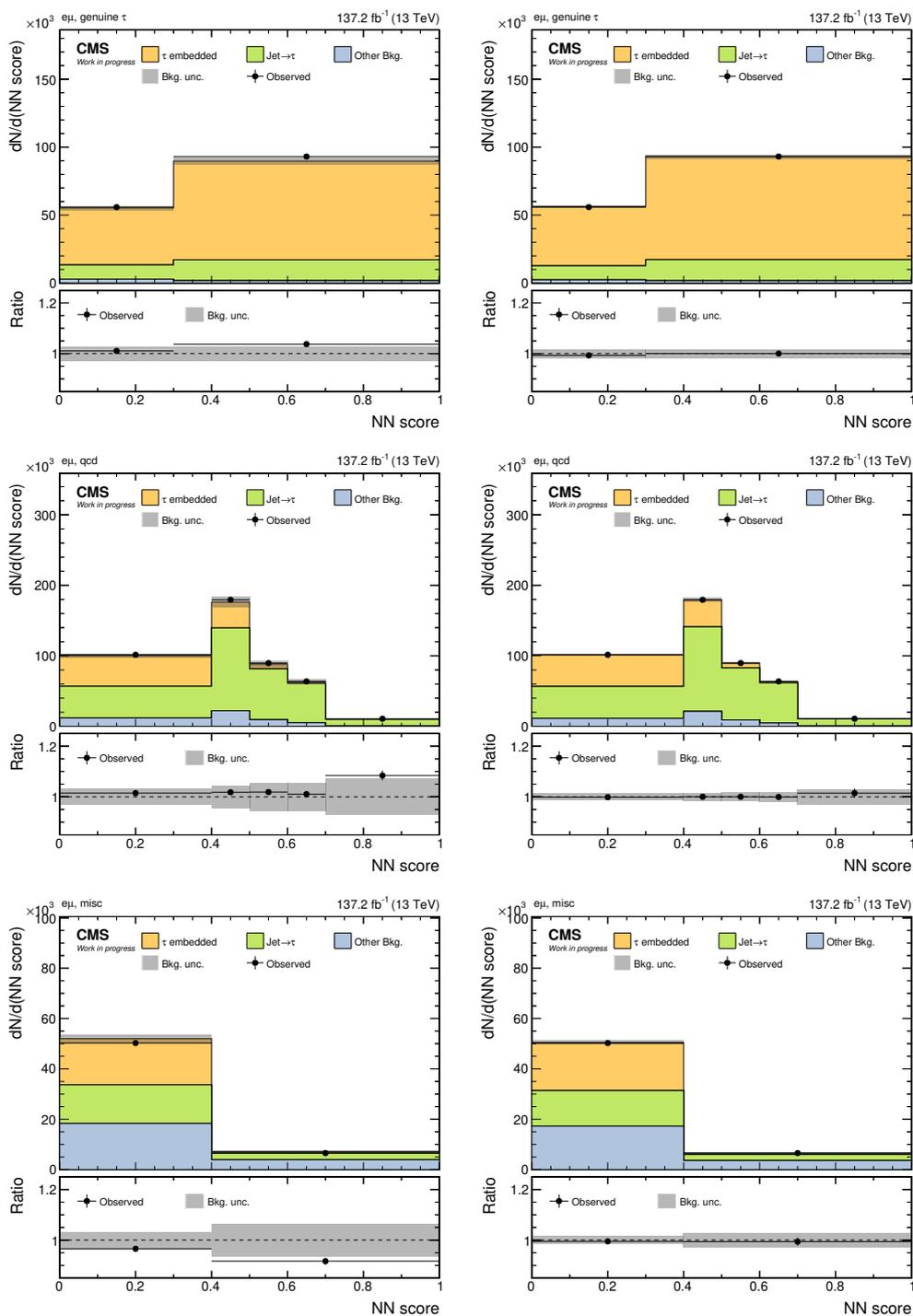
Figure B.17: STXS stage-1 categories in the $e\mu$ channel (VI). Prefit distributions are shown on the left, postfit distributions on the right.
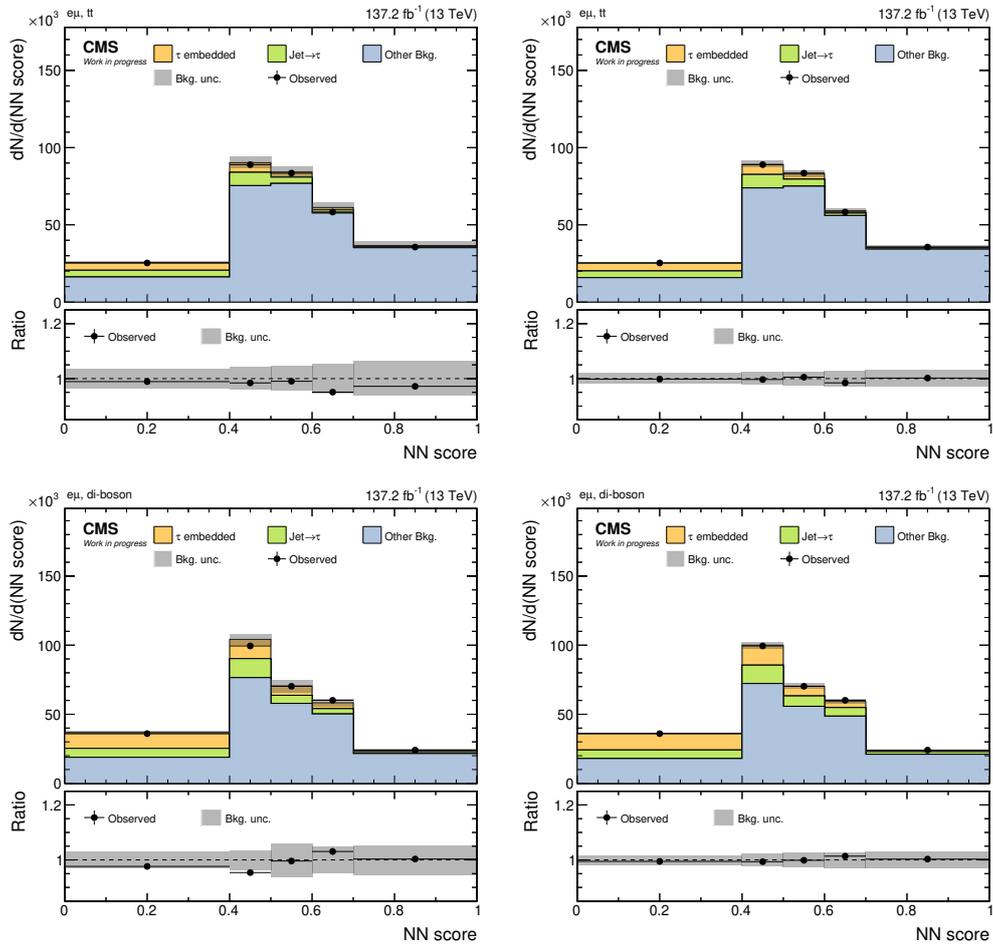
Figure B.18: STXS stage-1 categories in the $e\mu$ channel (VII). Prefit distributions are shown on the left, postfit distributions on the right.

Figure B.19: STXS stage-1 categories in the $e\tau_h$ channel (I). Prefit distributions are shown on the left, postfit distributions on the right.
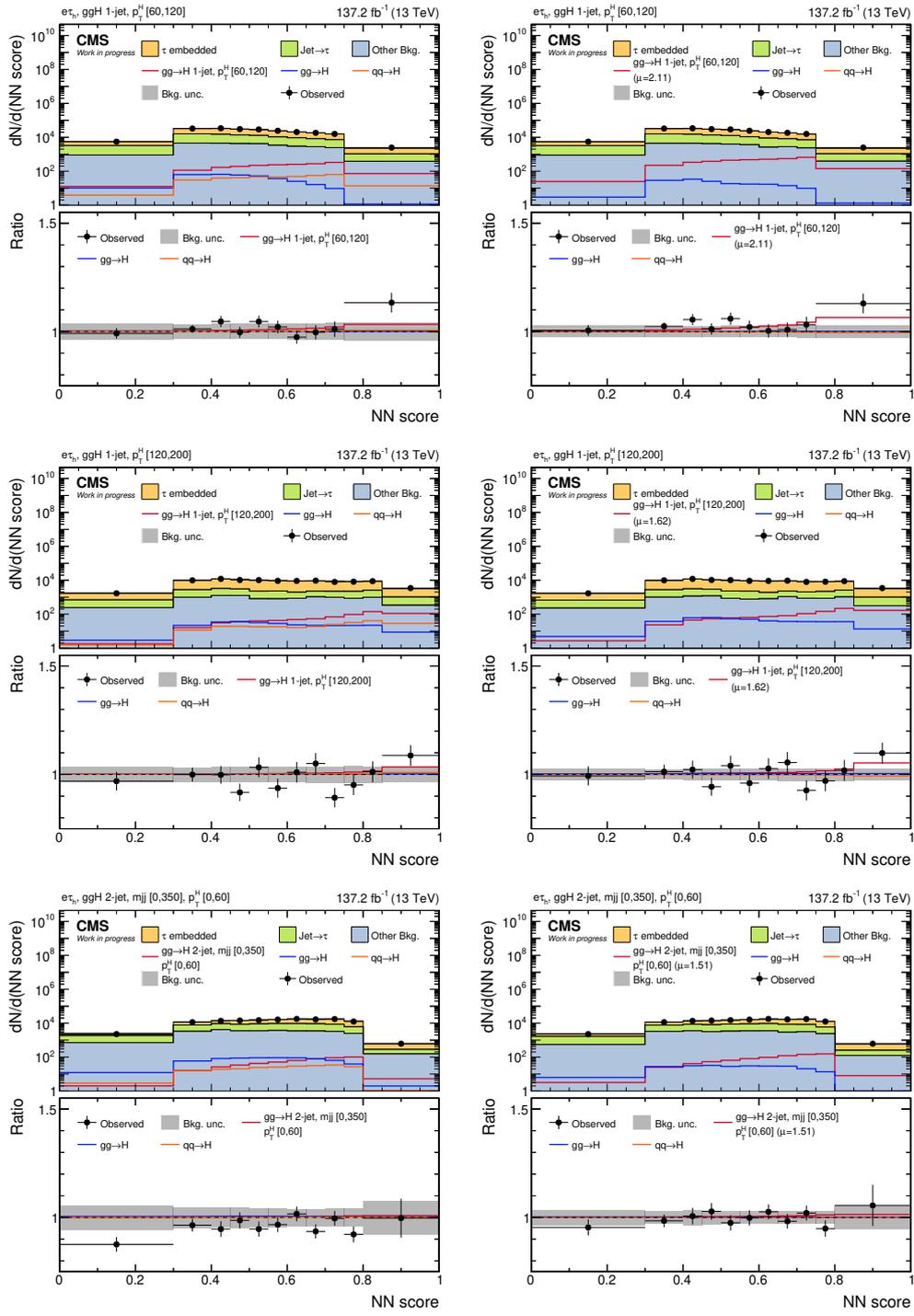
Figure B.20: STXS stage-1 categories in the $e\tau_\mathrm{h}$ channel (II). Prefit distributions are shown on the left, postfit distributions on the right.
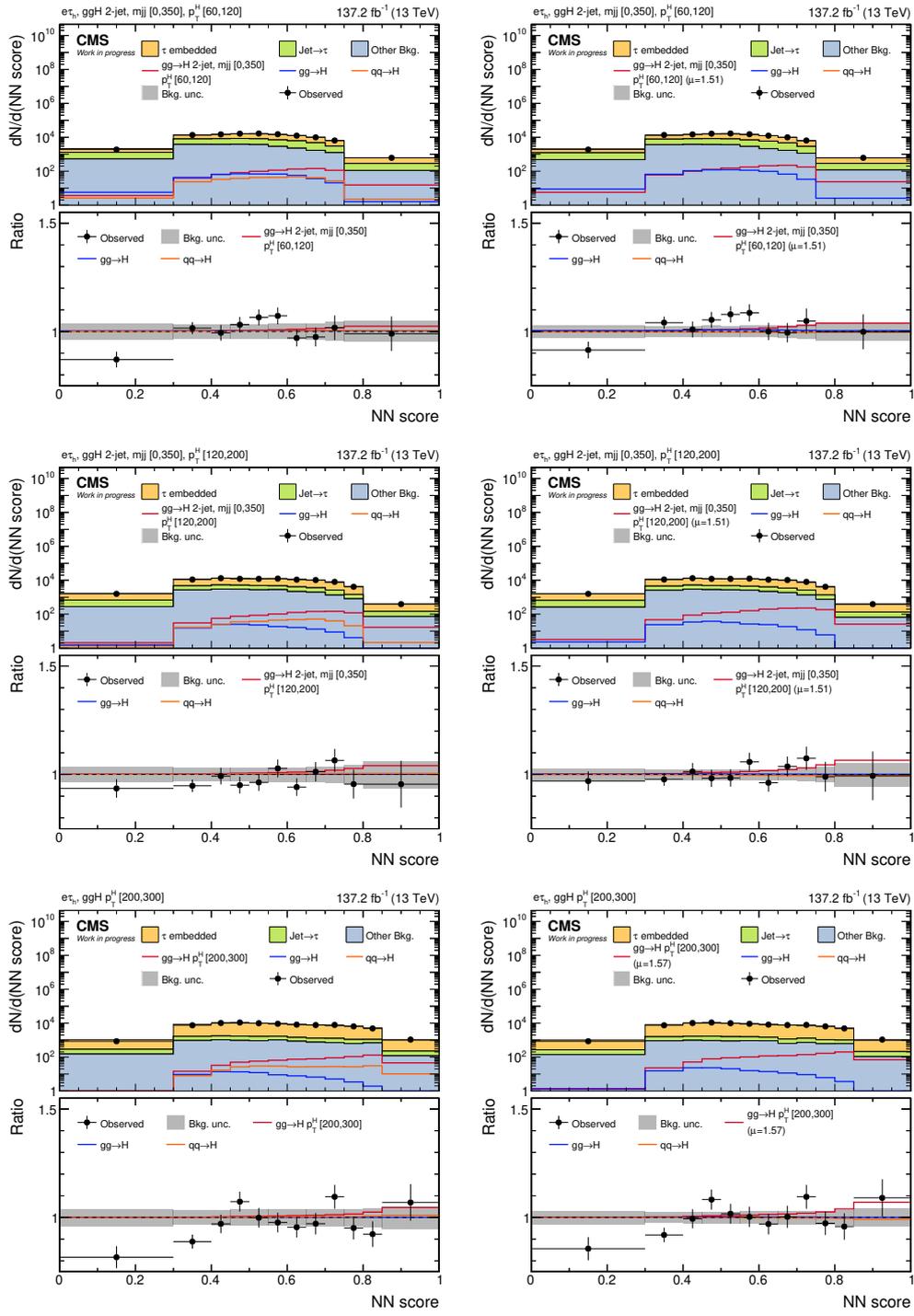
Figure B.21: STXS stage-1 categories in the $e\tau_h$ channel (III). Prefit distributions are shown on the left, postfit distributions on the right.
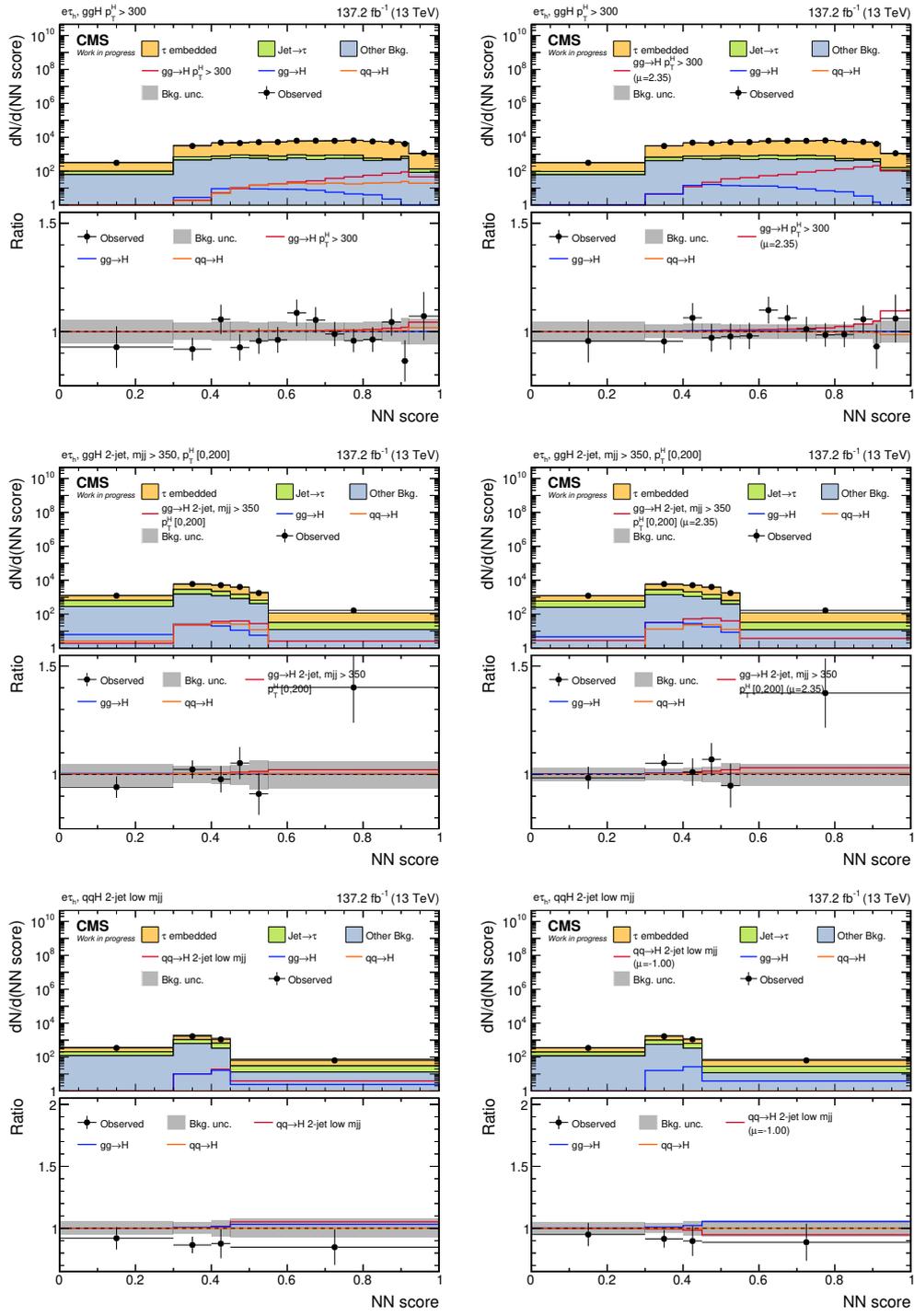
Figure B.22: STXS stage-1 categories in the $e\tau_{\mathrm{h}}$ channel (IV). Prefit distributions are shown on the left, postfit distributions on the right.
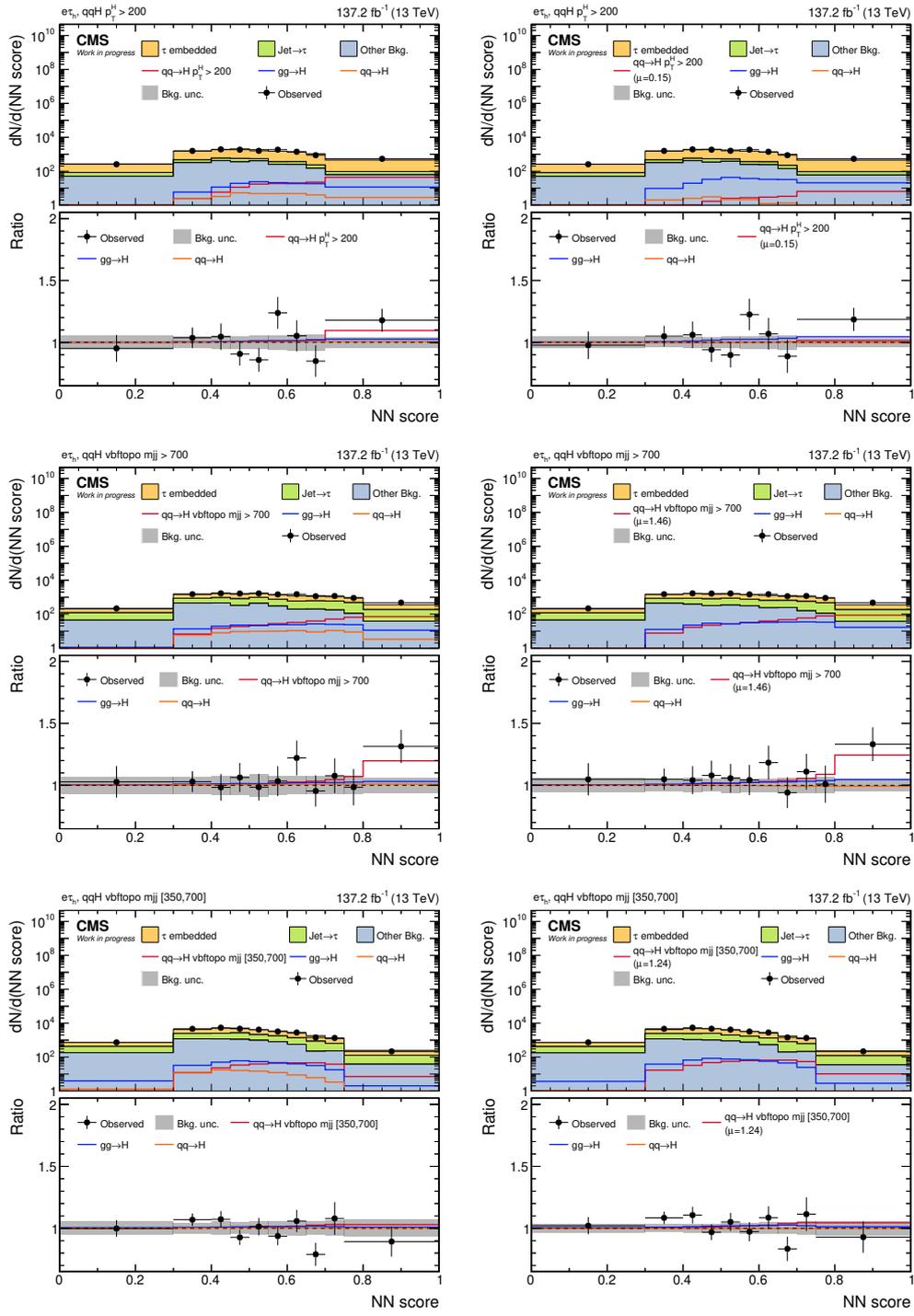
Figure B.23: STXS stage-1 categories in the $e\tau_\mathrm{h}$ channel (V). Prefit distributions are shown on the left, postfit distributions on the right.
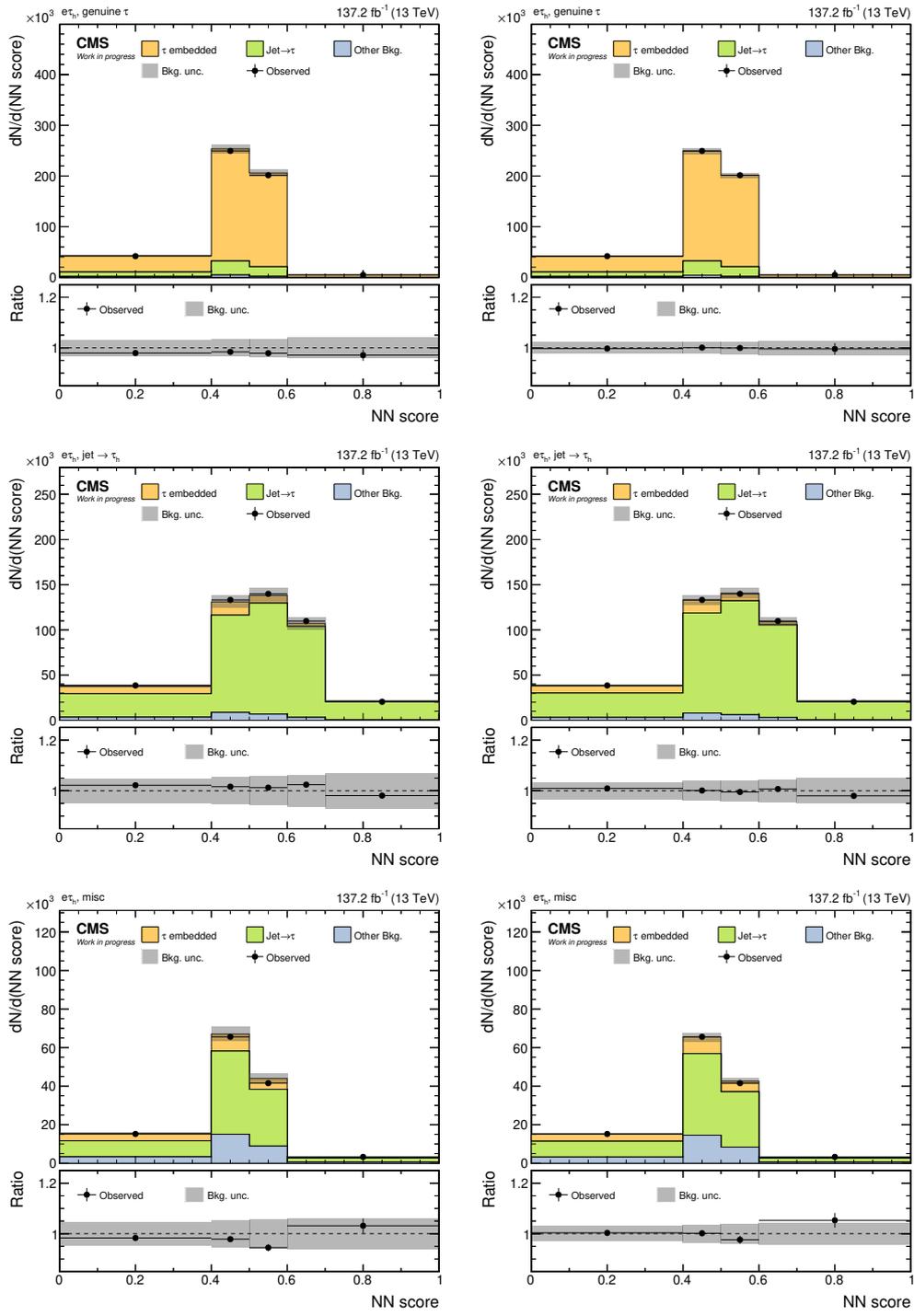
Figure B.24: STXS stage-1 categories in the $e\tau_h$ channel (VI). Prefit distributions are shown on the left, postfit distributions on the right.

Figure B.25: STXS stage-1 categories in the $e\tau_h$ channel (VII). Prefit distributions are shown on the left, postfit distributions on the right.
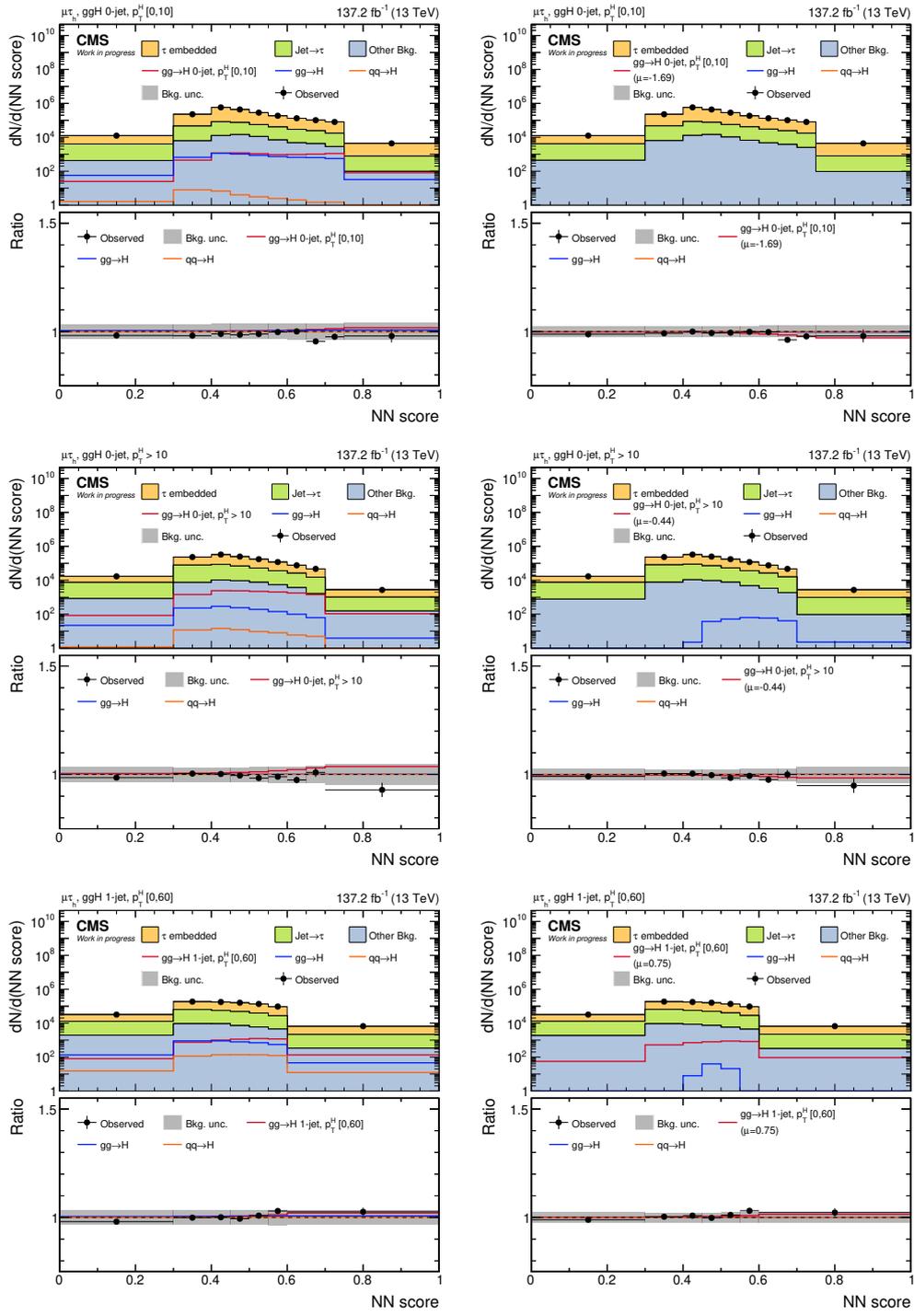
Figure B.26: STXS stage-1 categories in the $\mu\tau_h$ channel (I). Prefit distributions are shown on the left, postfit distributions on the right.
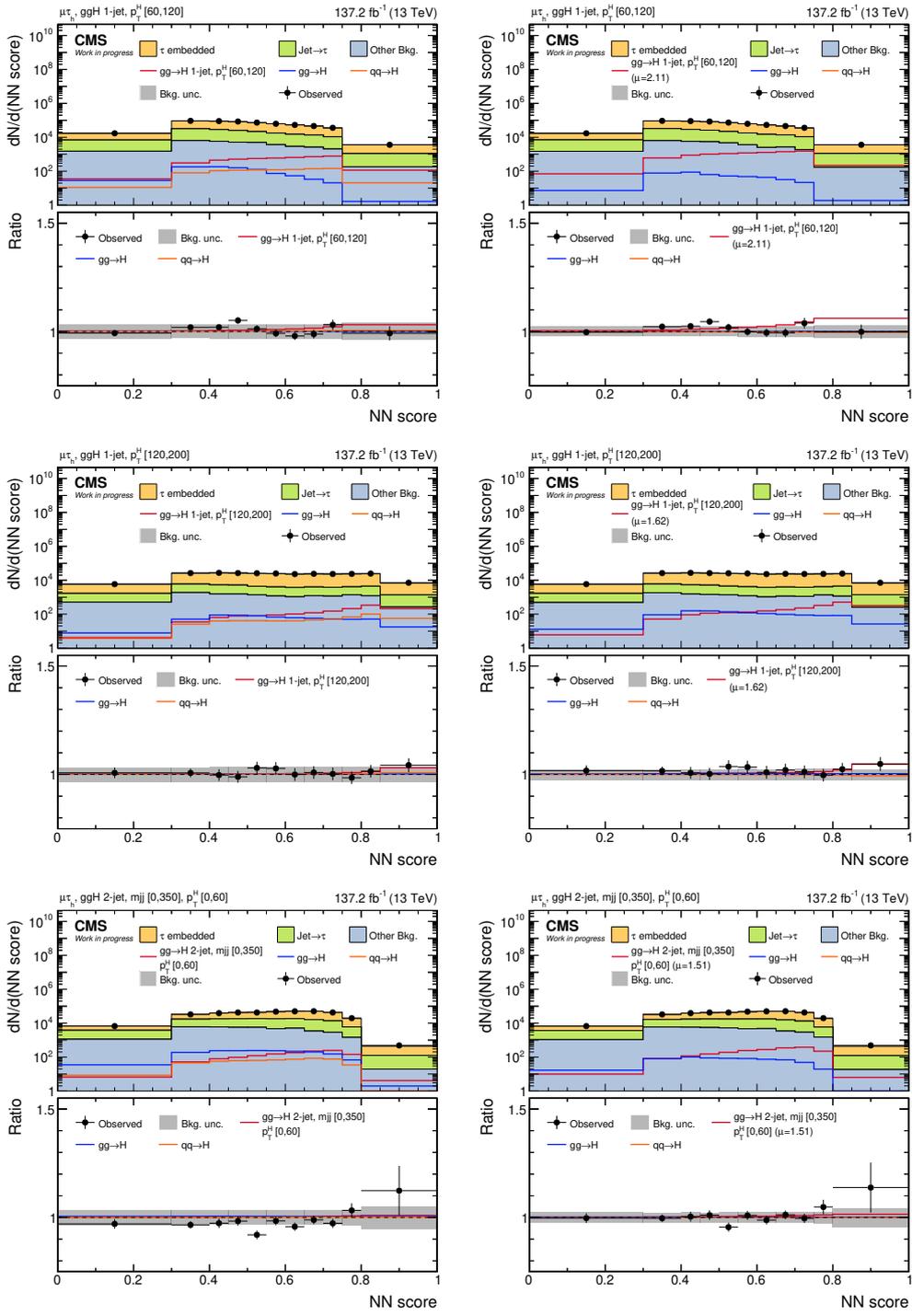
Figure B.27: STXS stage-1 categories in the $\mu\tau_\mathrm{h}$ channel (II). Prefit distributions are shown on the left, postfit distributions on the right.
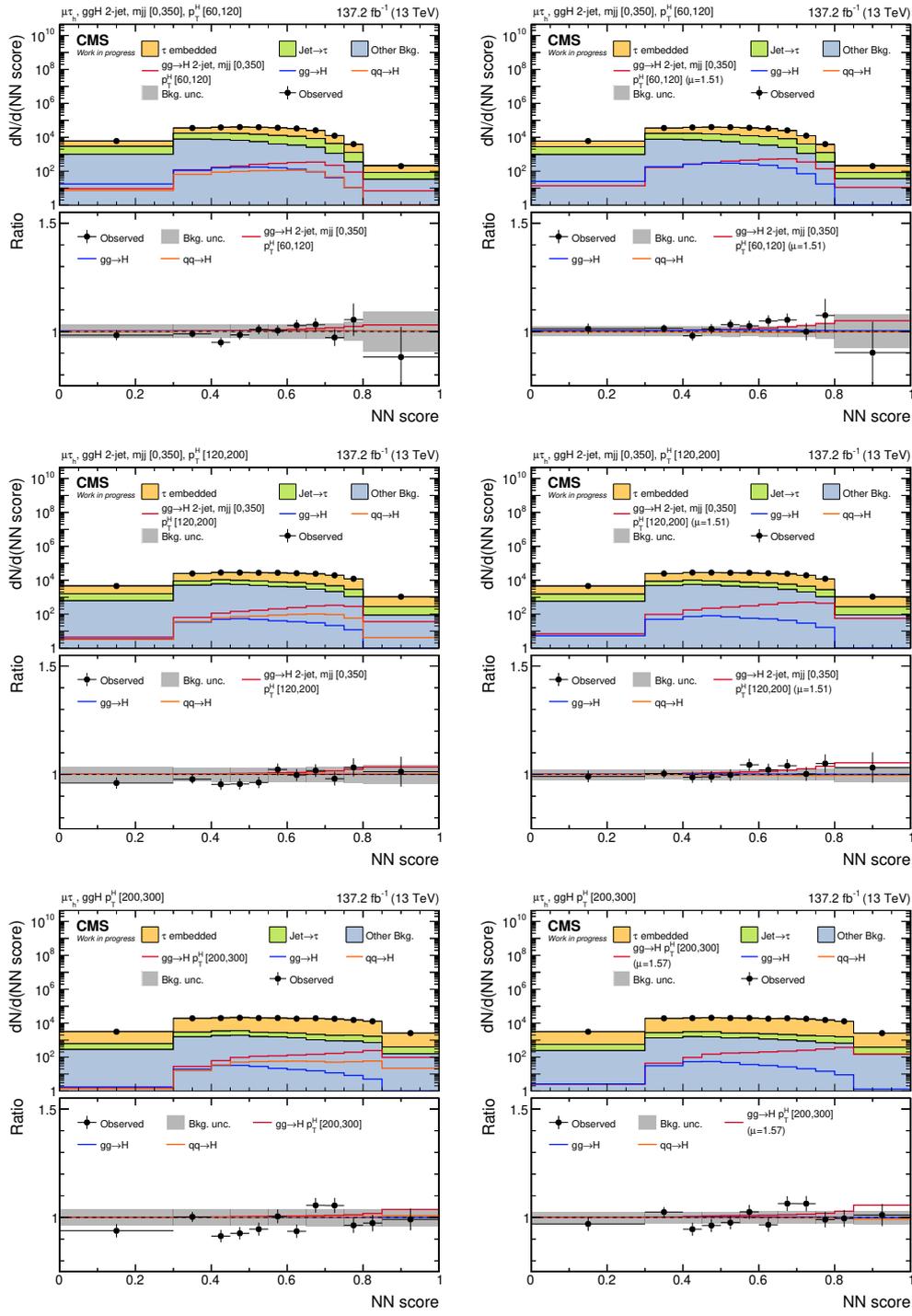
Figure B.28: STXS stage-1 categories in the $\mu\tau_h$ channel (III). Prefit distributions are shown on the left, postfit distributions on the right.
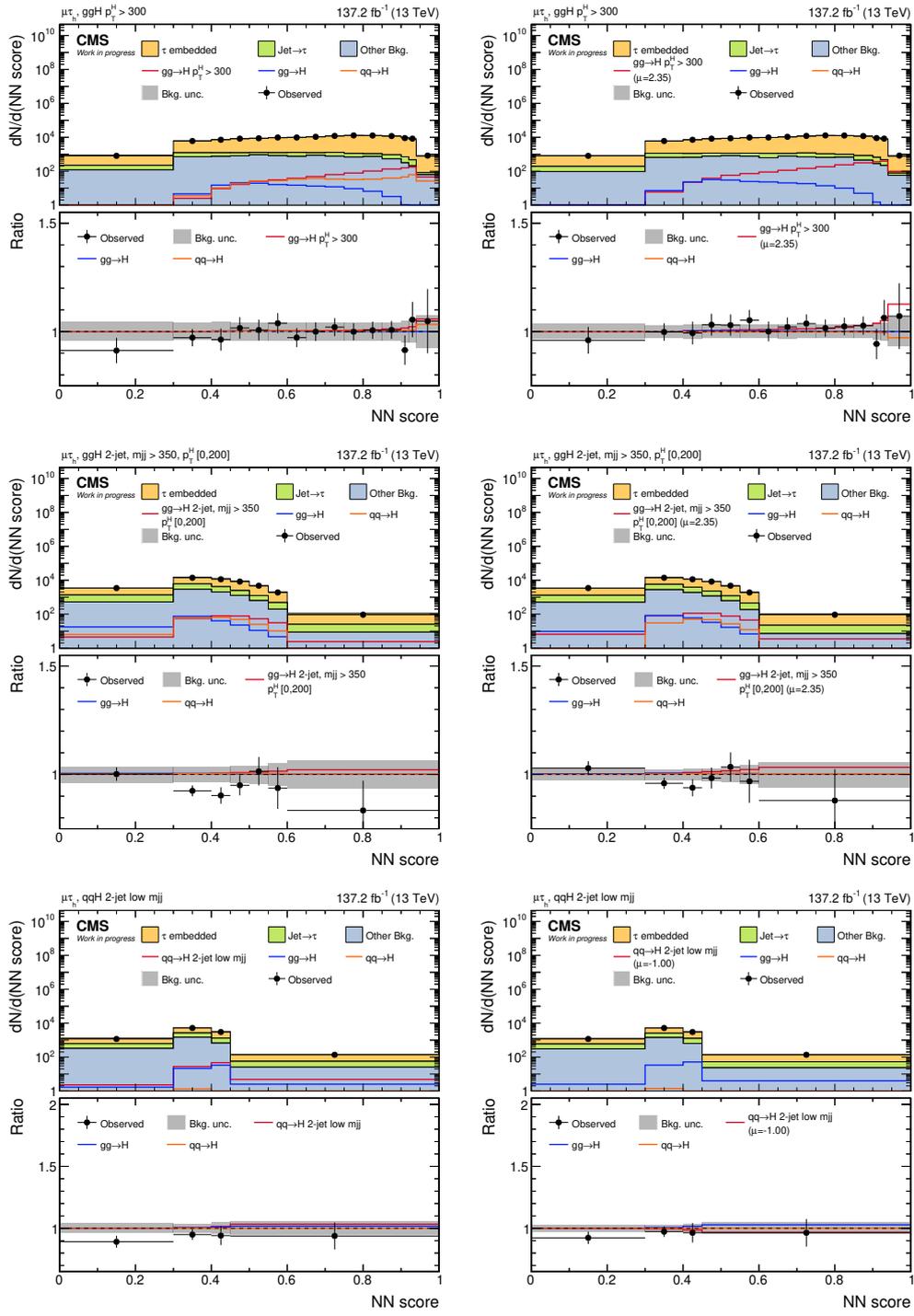
Figure B.29: STXS stage-1 categories in the $\mu\tau_\text{h}$ channel (IV). Prefit distributions are shown on the left, postfit distributions on the right.
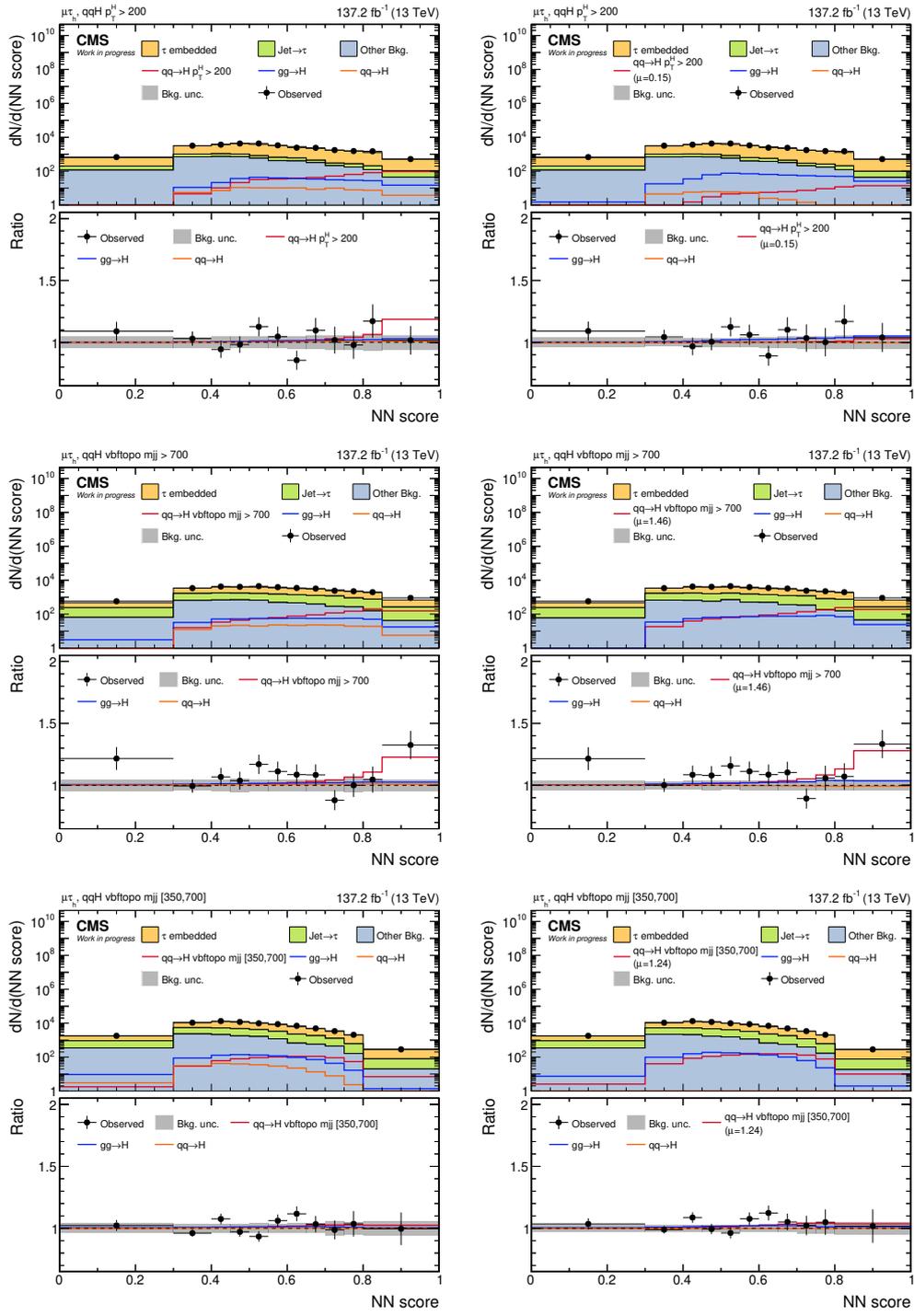
Figure B.30: STXS stage-1 categories in the $\mu\tau_h$ channel (V). Prefit distributions are shown on the left, postfit distributions on the right.
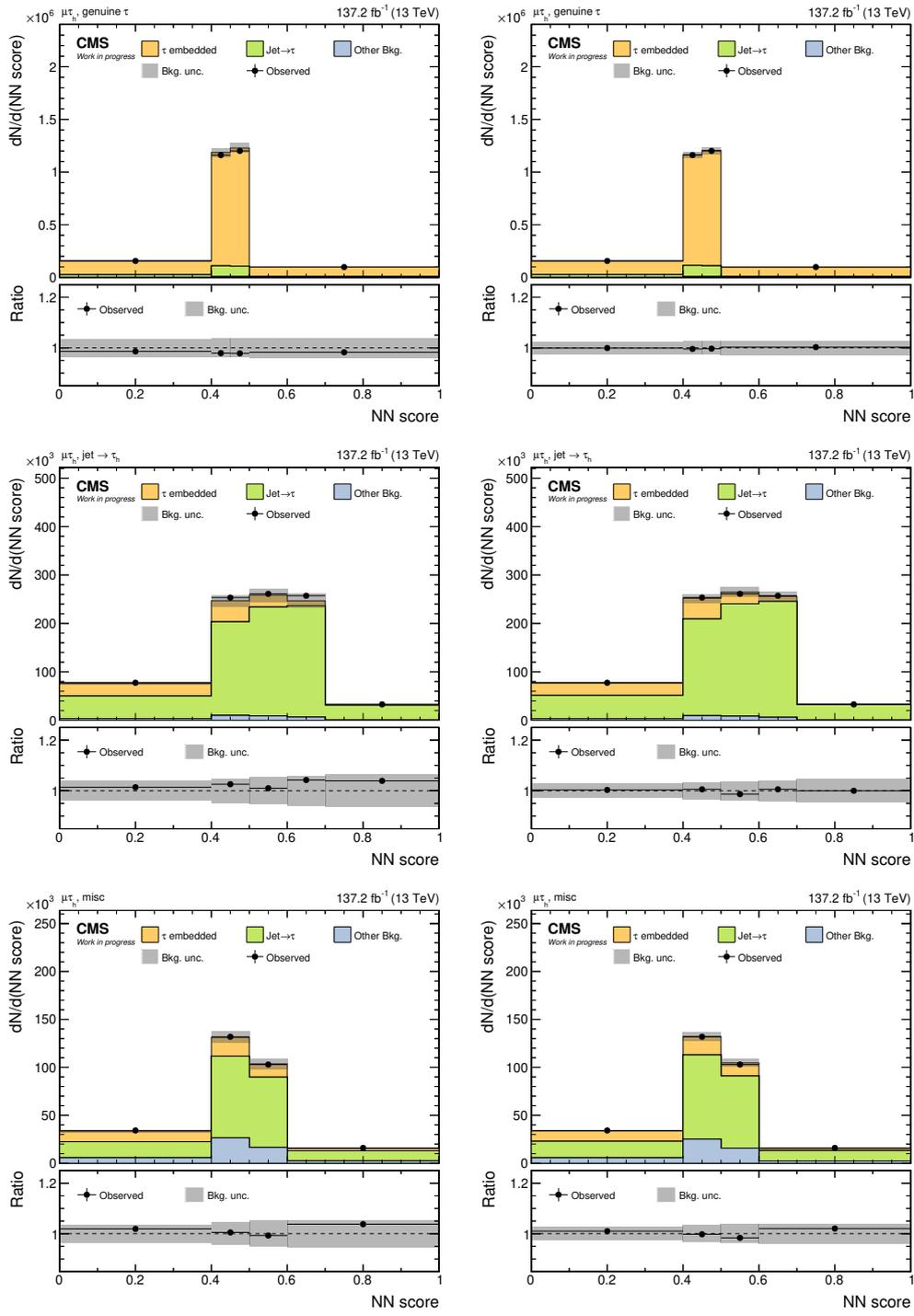
Figure B.31: STXS stage-1 categories in the $\mu\tau_\mathrm{h}$ channel (VI). Prefit distributions are shown on the left, postfit distributions on the right.
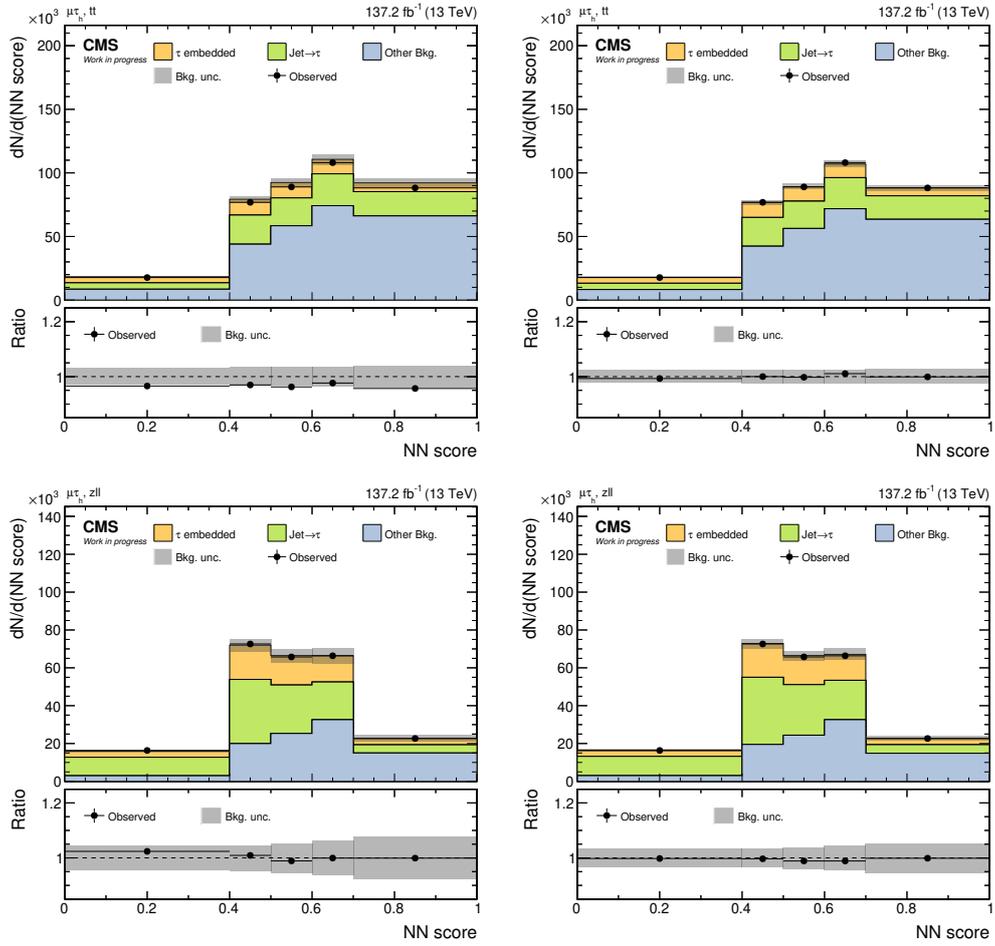
Figure B.32: STXS stage-1 categories in the $\mu\tau_\mathrm{h}$ channel (VII). Prefit distributions are shown on the left, postfit distributions on the right.
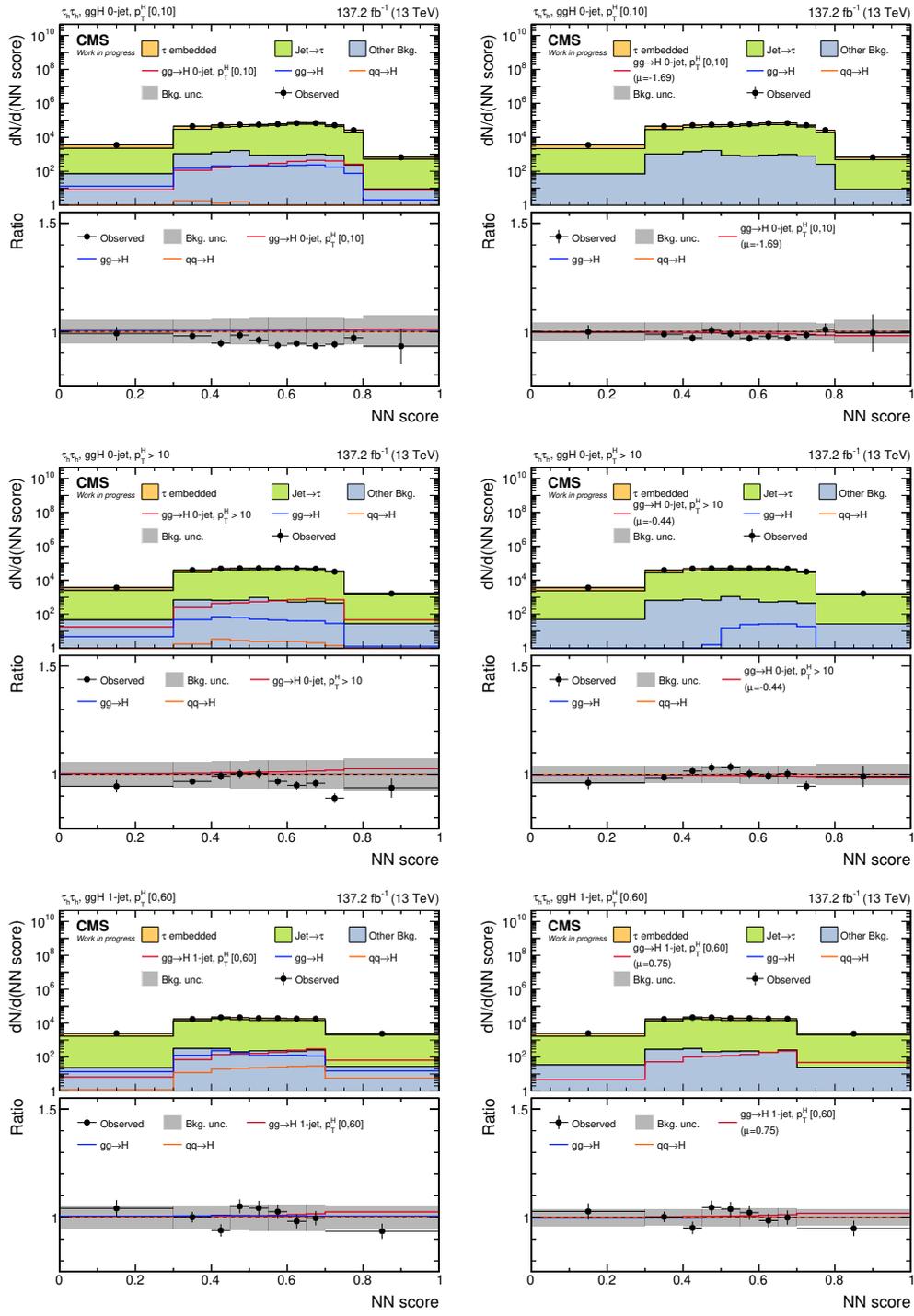
Figure B.33: STXS stage-1 categories in the $\tau_h\tau_h$ channel (I). Prefit distributions are shown on the left, postfit distributions on the right.
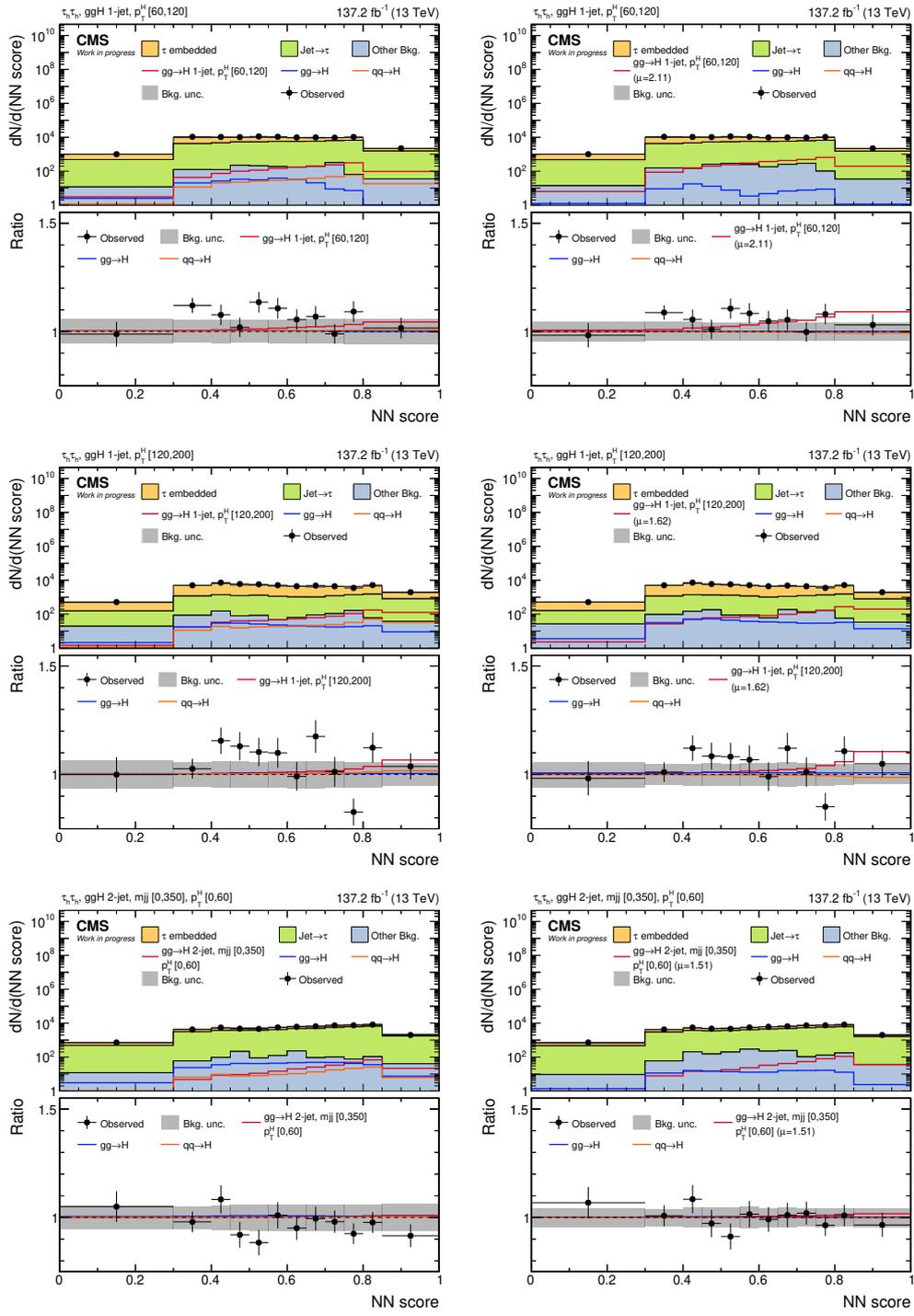
Figure B.34: STXS stage-1 categories in the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel (II). Prefit distributions are shown on the left, postfit distributions on the right.
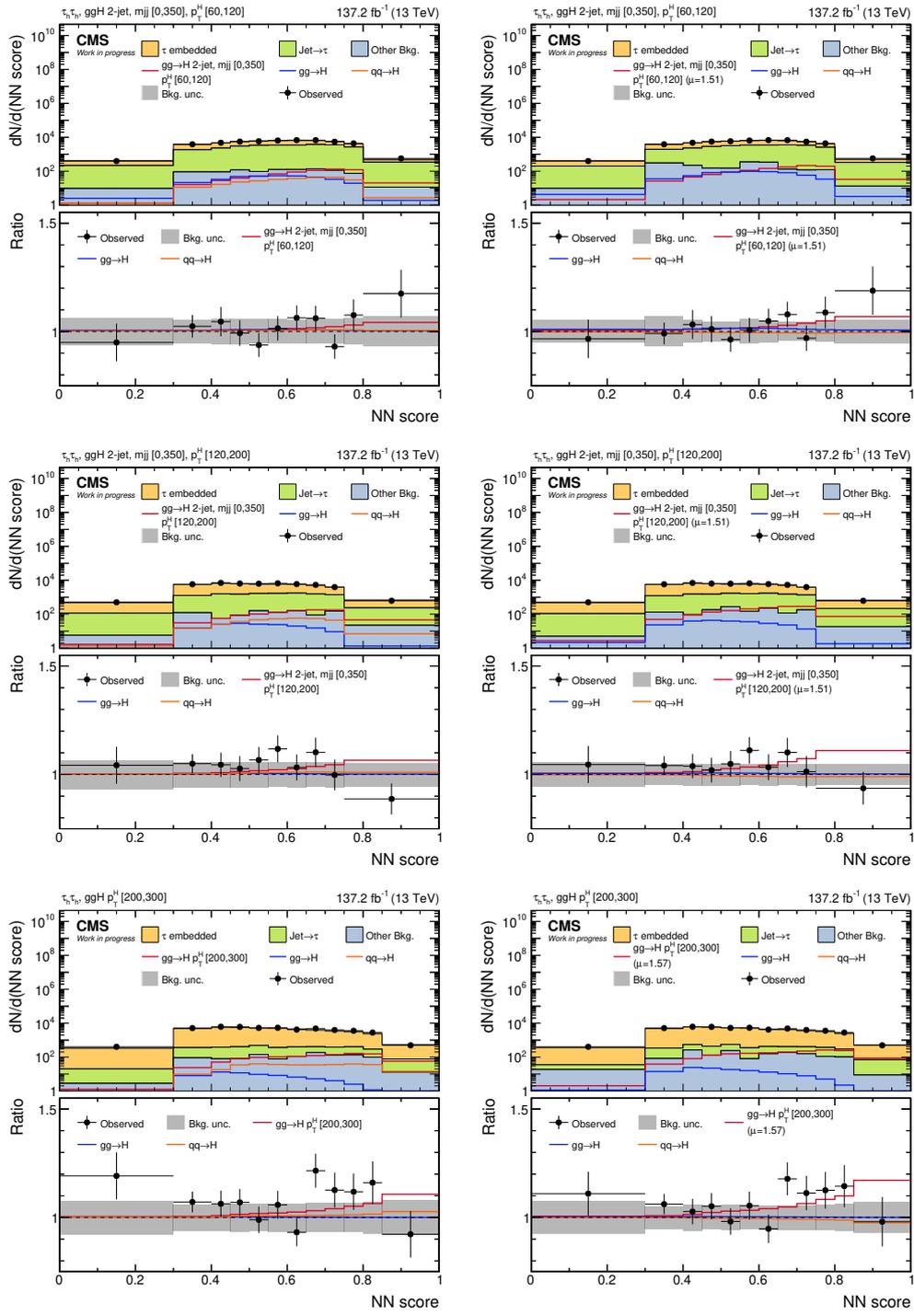
Figure B.35: STXS stage-1 categories in the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel (III). Prefit distributions are shown on the left, postfit distributions on the right.
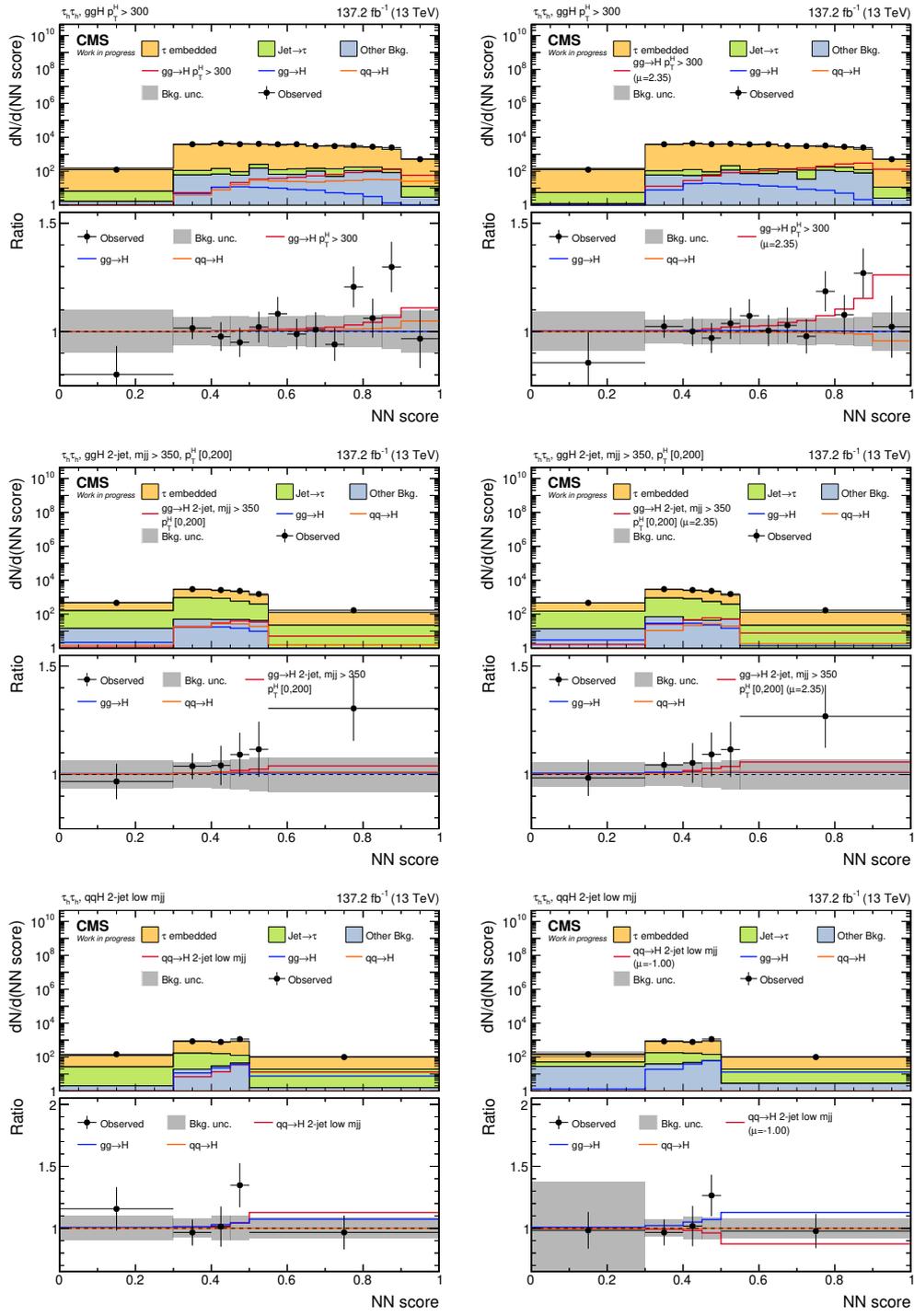
Figure B.36: STXS stage-1 categories in the $\tau_h\tau_h$ channel (IV). Prefit distributions are shown on the left, postfit distributions on the right.
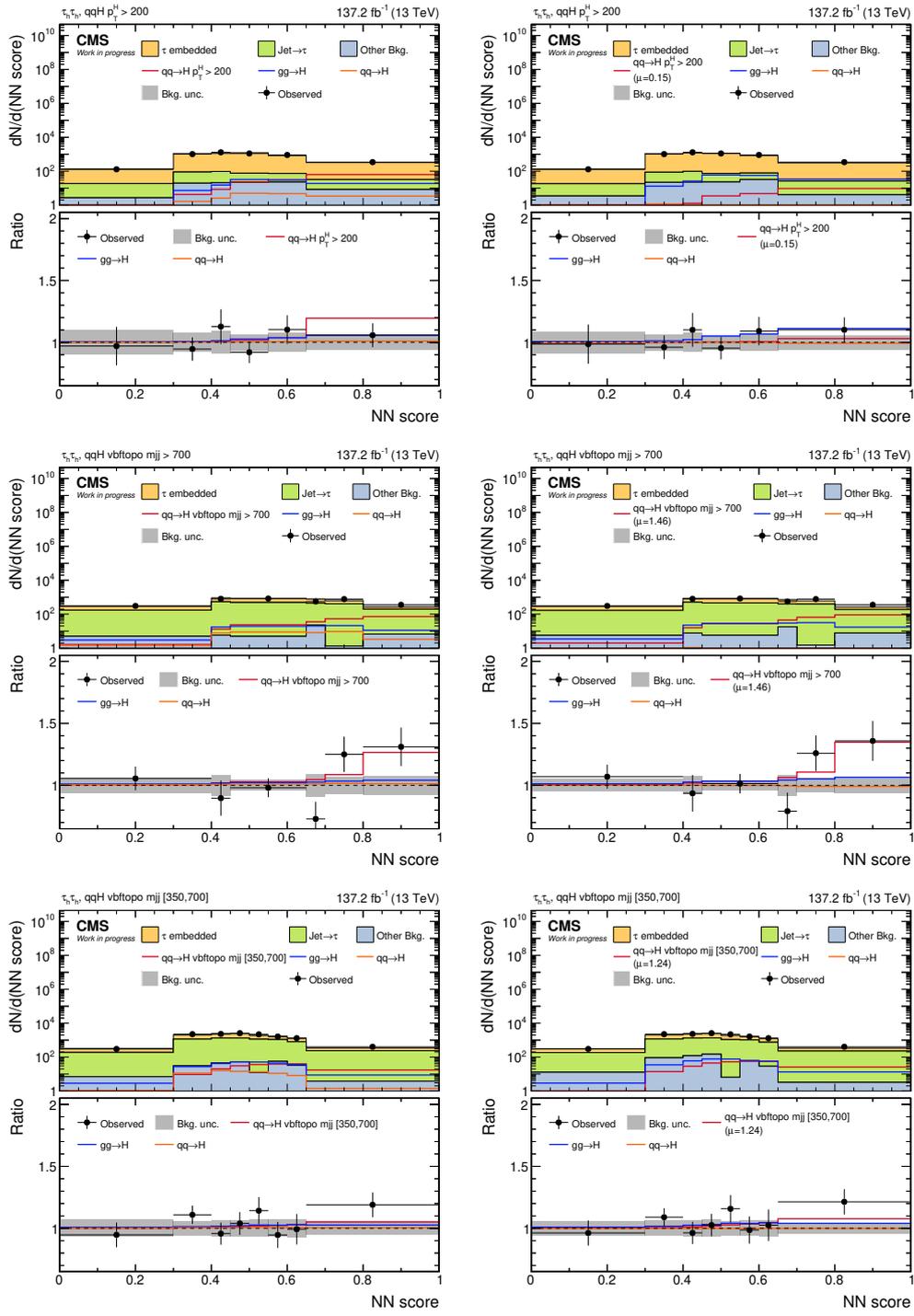
Figure B.37: STXS stage-1 categories in the $\tau_h\tau_h$ channel (V). Prefit distributions are shown on the left, postfit distributions on the right.
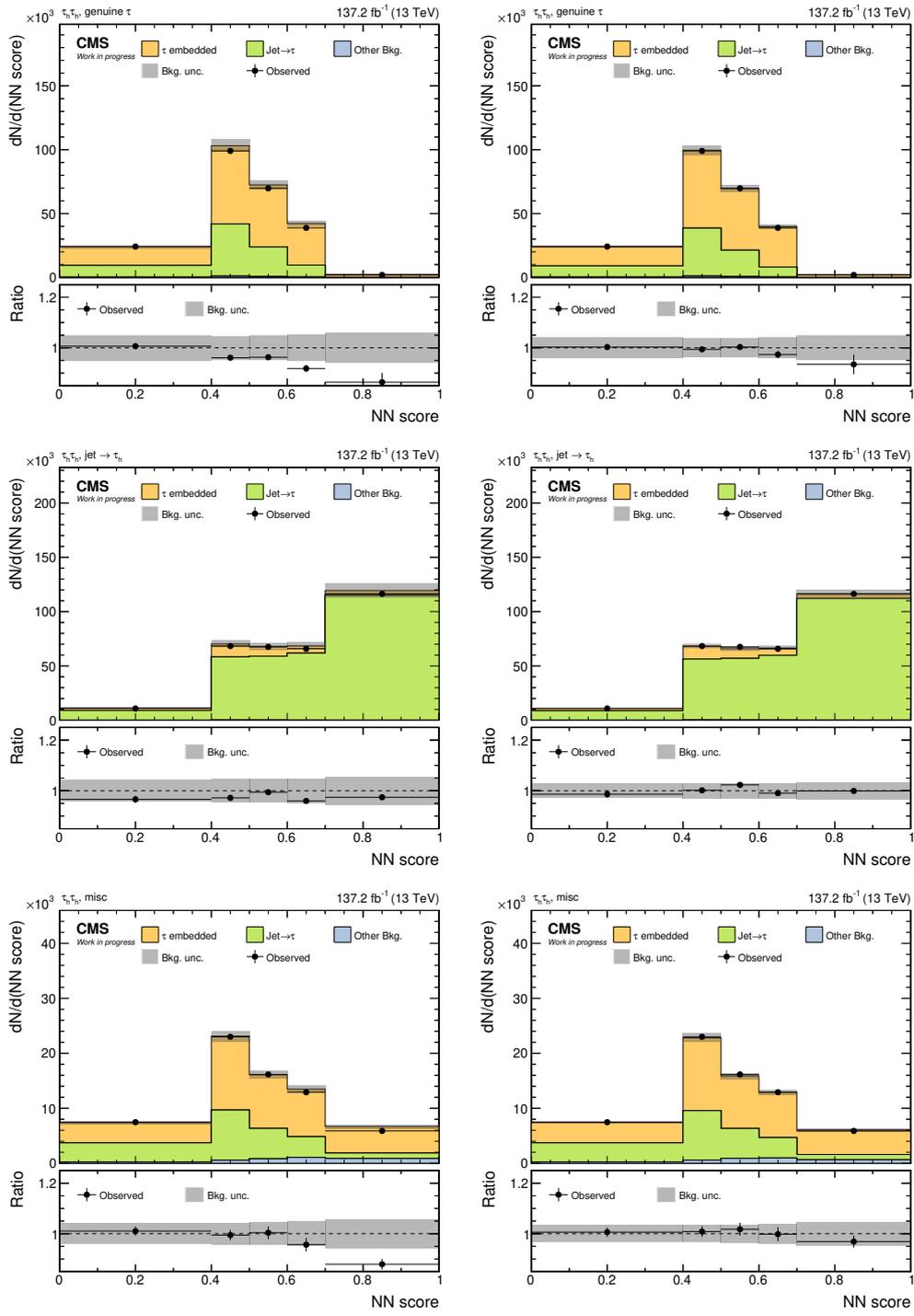
Figure B.38: STXS stage-1 categories in the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel (VI). Prefit distributions are shown on the left, postfit distributions on the right.

Confusion matrices of the event classifiers

Figure C.1: Confusion matrix of the stage-0 $e\mu$ classifier. Given numbers indicate the fraction of a true class directed to a predicted class.

Figure C.2: Confusion matrix of the stage-0 $e\tau_{\mathrm{h}}$ classifier. Given numbers indicate the fraction of a true class directed to a predicted class.

Figure C.3: Confusion matrix of the stage-0 $\tau_h \tau_h$ classifier. Given numbers indicate the fraction of a true class directed to a predicted class.

Figure C.4: Confusion matrix of the stage-1 $e\mu$ classifier. Given numbers indicate the fraction of a true class directed to a predicted class.

Figure C.5: Confusion matrix of the stage-1 $e\tau_\mathrm{h}$ classifier. Given numbers indicate the fraction of a true class directed to a predicted class.

165

$\tau_h\tau_h$ CMS *Work in progress*

Predicted event class (rows) vs True event class (columns):

Predicted \ True	ggH 0 Jet p_T^H[0,10]	ggH 0 Jet p_T^H[10,200]	ggH 1 Jet p_T^H[0,60]	ggH 1 Jet p_T^H[60,120]	ggH 1 Jet p_T^H[120,200]	ggH ≥2 Jet m_{jj}[0,350] p_T^H[0,60]	ggH ≥2 Jet m_{jj}[0,350] p_T^H[60,120]	ggH ≥2 Jet m_{jj}[0,350] p_T^H[120,200]	ggH ≥2 Jet m_{jj}≥350 p_T^H[0,200]	ggH p_T^H[200,300]	ggH p_T^H≥300	qqH ≥2 Jet m_{jj}[0,350]	qqH ≥2 Jet m_{jj}≥350 p_T^H≥200	qqH ≥2 Jet m_{jj}≥700 p_T^H[0,200]	qqH ≥2 Jet m_{jj}[350,700] p_T^H[0,200]	genuine τ	jet→τ_h	misc
ggH 0 Jet p_T^H[0,10]	0.89	0.21	0.04	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.17	0.17
ggH 0 Jet p_T^H[10,200]	0.04	0.58	0.13	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.14	0.10
ggH 1 Jet p_T^H[0,60]	0.01	0.13	0.67	0.05	0.00	0.18	0.01	0.00	0.02	0.00	0.00	0.01	0.00	0.01	0.02	0.03	0.05	0.04
ggH 1 Jet p_T^H[60,120]	0.00	0.01	0.03	0.74	0.03	0.01	0.15	0.00	0.04	0.00	0.00	0.03	0.00	0.04	0.04	0.04	0.03	0.03
ggH 1 Jet p_T^H[120,200]	0.00	0.00	0.00	0.02	0.81	0.00	0.01	0.09	0.04	0.00	0.03	0.00	0.00	0.05	0.04	0.04	0.01	0.02
ggH ≥2 Jet m_{jj}[0,350] p_T^H[0,60]	0.01	0.02	0.10	0.01	0.00	0.73	0.03	0.00	0.02	0.00	0.00	0.08	0.00	0.01	0.02	0.02	0.02	0.03
ggH ≥2 Jet m_{jj}[0,350] p_T^H[60,120]	0.00	0.00	0.00	0.12	0.01	0.01	0.73	0.04	0.03	0.00	0.00	0.17	0.00	0.01	0.02	0.02	0.01	0.02
ggH ≥2 Jet m_{jj}[0,350] p_T^H[120,200]	0.00	0.00	0.00	0.00	0.08	0.00	0.02	0.76	0.05	0.03	0.00	0.25	0.00	0.01	0.03	0.03	0.01	0.03
ggH ≥2 Jet m_{jj}≥350 p_T^H[0,200]	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.03	0.25	0.01	0.00	0.01	0.01	0.04	0.14	0.01	0.00	0.01
ggH p_T^H[200,300]	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.68	0.04	0.13	0.10	0.00	0.00	0.04	0.00	0.03
ggH p_T^H≥300	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.84	0.09	0.13	0.00	0.00	0.03	0.00	0.03
qqH ≥2 Jet m_{jj}[0,350]	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.05	0.02	0.14	0.00	0.00	0.01	0.00	0.00
qqH ≥2 Jet m_{jj}≥350 p_T^H≥200	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.12	0.08	0.01	0.71	0.02	0.01	0.01	0.00	0.02
qqH ≥2 Jet m_{jj}≥700 p_T^H[0,200]	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.19	0.00	0.00	0.01	0.03	0.76	0.04	0.00	0.00	0.01
qqH ≥2 Jet m_{jj}[350,700] p_T^H[0,200]	0.00	0.00	0.01	0.02	0.02	0.02	0.02	0.01	0.31	0.00	0.00	0.02	0.01	0.04	0.63	0.01	0.00	0.01
genuine τ	0.02	0.01	0.00	0.01	0.02	0.00	0.00	0.02	0.01	0.03	0.01	0.01	0.01	0.00	0.01	0.34	0.10	0.16
jet→τ_h	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.08	0.43	0.05
misc	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.10	0.03	0.24

True event class

Figure C.6: Confusion matrix of the stage-1 $\tau_h\tau_h$ classifier. Given numbers indicate the fraction of a true class directed to a predicted class.

Profile likelihood scans of the STXS stage-1 measurement

Figure D.1: Profile likelihood scans of the STXS stage-1 measurement (I)

Figure D.2: Profile likelihood scans of the STXS stage-1 measurement (II)

List of Figures

List of Tables

Acronyms

ATLAS A Toroidal LHC AparatuS. 1, 2, 9, 11, 12, 14, 20, 74, 81, 82

bbH bottom quark associated production. 9, 42

CERN European Organization for Nuclear Research. 1, 19

CMS Compact Muon Solenoid. 1, 2, 9, 11, 13, 14, 20–22, 29, 31, 49, 68, 74, 81, 82

$\not{E}_{\mathrm{T}}$ Missing Transverse Energy. 11, 31–33, 35, 40, 44, 46, 48, 49, 55, 71, 83

F_F **method** fake rate method. 44, 47, 71

ggH gluon fusion. 8, 9, 12–15, 17, 41, 42, 51, 57–59, 63–65, 71, 73, 74, 76, 82

GoF goodness of fit. 45, 55, 56

HLT High-Level Trigger. 34, 35, 37, 38, 71

HPS hadron-plus-strip. 29

LHC Large Hadron Collider. 1–3, 8, 11, 14, 19, 20, 23, 29, 33, 81, 82

MELA Matrix Element Likelihood Analysis. 55

MSSM Minimal Supersymmetric Extension to the Standard Model. 8, 14, 17

NN score neural network output value. 52–55, 57–60, 63, 65, 121, 169

PDF parton distribution function. 9

PF Particle-Flow. 25–29, 31, 32, 34

POI parameter of interest. 64, 68, 74

PUPPI Pileup-Per-Particle-Identification. 31, 49

qqH Higgs boson production via vector boson fusion or Higgs strahlung with a hadronically decaying vector boson. 9, 14–16, 42, 55, 57–59, 63–65, 71, 73, 76, 82

SM Standard Model of Particle Physics. 3, 7, 11, 12, 14, 37, 73, 74, 77, 82

STXS Simplified Template Cross Sections. 2, 14–17, 41, 46, 51, 57, 58, 63, 64, 70, 72–74, 76, 77, 81, 82, 122–132, 134–160, 167, 168

ttH top quark associated production. 9–11, 42, 71

VBF vector boson fusion. 8, 9, 12, 13, 15, 16, 27, 42, 49, 51, 55

VH Higgs strahlung. 9, 16, 42, 63

VH$_{lep}$ Higgs strahlung with a leptonically decaying vector boson. 9, 11, 12, 42, 71

WH Higgs strahlung with a W boson. 9

ZH Higgs strahlung with a Z boson. 9

Bibliography

[1] CMS Collaboration. "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector". *Physics Letters B* 779 (Apr. 2018), pp. 283–316. ISSN: 0370-2693.
DOI: 10.1016/j.physletb.2018.02.004.

[2] Sheldon Lee Glashow. "Partial-Symmetries of Weak Interactions". *Nucl. Phys.* 22 (4 1961), pp. 579–588.
DOI: 10.1016/0029-5582(61)90469-2.

[3] Abdus Salam. "Weak and Electromagnetic Interactions". *Elementary particle theory. Relativistic groups and analyticity.* Proceedings of the Eighth Nobel Symposium. (Aspenäsgarden, Lerum, May 19–25, 1968). Ed. by Nils Svartholm. Stockholm: Almquist & Wiksell, 1968, pp. 367–377.

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

[5] Peter W. Higgs. "Broken symmetries, massless particles and gauge fields". *Phys. Lett.* 12.2 (1964), pp. 132–133.
DOI: 10.1016/0031-9163(64)91136-9.

[6] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Phys. Rev. Lett.* 13 (1964), pp. 508–509.
DOI: 10.1103/PhysRevLett.13.508.

[7] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323.
DOI: 10.1103/PhysRevLett.13.321.

[8] "Particle Data Group: Gauge and Higgs Bosons". URL: http://pdg.lbl.gov/2016/tables/rpp2016-sum-gauge-higgs-bosons.pdf (visited on 08/17/2020).

[9] Franz Mandl and Graham Shaw. "Quantum field theory". 2. ed. Chichester: Wiley, 2010. ISBN: 978-0-471-49683-0; 978-0-471-49684-7.

[10] Ta-Pei Cheng and Ling-Fong Li. "Gauge theory of elementary particle physics". Repr. Oxford science publications. Oxford [u.a.]: Clarendon Press, 2008. ISBN: 978-0-19-851961-4.

[11] Michael Edward Peskin. "An introduction to quantum field theory". Ed. by Daniel V. Schroeder. Student economy edition. The advanced book program. Boulder: Westview Press, a member of the Perseus Books Group, 2016. ISBN: 0813350190; 9780813350196.

[12] Sebastian Wozniewski. "Development of multivariate analysis methods for enhancing the sensitivity to supersymmetric Higgs bosons in the $H \to \tau\tau$ decay channel". IEKP-KA/2017-18. MA thesis. Karlsruhe Institute of Technology (KIT), 2017.

[13] Roger Wolf. "The Higgs Boson Discovery at the Large Hadron Collider". Springer Tracts in Modern Physics. Cham: Springer, 2015. ISBN: 978-331-91851-2-5.

[14] "CERN, LHC Higgs Cross Section WG Picture Gallery". URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGCrossSectionsFigures (visited on 08/10/2020).

[15] CMS Collaboration. "Higgs boson production in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = $ 13 TeV". Tech. rep. CMS-PAS-HIG-19-008. Geneva: CERN, 2020.

[16] ATLAS Collaboration. "Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector". *Phys. Rev. D* 97 (7 Apr. 2018), p. 072003.
DOI: 10.1103/PhysRevD.97.072003.

[17] CMS Collaboration. "Measurements of Higgs boson properties in the diphoton decay channel at $\sqrt{s} = 13$ TeV". Tech. rep. CMS-PAS-HIG-19-015. Geneva: CERN, 2020.

[18] ATLAS Collaboration. "Measurement of the properties of Higgs boson production at $\sqrt{s}$=13 TeV in the $H \to \gamma\gamma$ channel using 139 fb^{-1} of pp collision data with the ATLAS experiment". Tech. rep. ATLAS-CONF-2020-026. Geneva: CERN, Aug. 2020.

[19] CMS Collaboration. "Measurements of differential Higgs boson production cross sections in the leptonic WW decay mode at $\sqrt{s} = 13$ TeV". Tech. rep. CMS-PAS-HIG-19-002. Geneva: CERN, 2019.

[20] ATLAS Collaboration. "Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Physics Letters B* 789 (2019), pp. 508–529. ISSN: 0370-2693.
DOI: https://doi.org/10.1016/j.physletb.2018.11.064.

[21] ATLAS Collaboration. "Observation of vector-boson-fusion production of Higgs bosons in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". Tech. rep. ATLAS-CONF-2020-045. Geneva: CERN, Aug. 2020.

[22] "Particle Data Group: Leptons". URL: http://pdg.lbl.gov/2016/tables/rpp2016-sum-leptons.pdf (visited on 08/17/2020).

[23] CMS Collaboration. "Measurements of properties of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". Tech. rep. CMS-PAS-HIG-19-001. Geneva: CERN, 2019.

[24] ATLAS Collaboration. "Higgs boson production cross-section measurements and their EFT interpretation in the 4ℓ decay channel at $\sqrt{s} =$13 TeV with the ATLAS detector". *Eur. Phys. J. C* 80.10 (2020), p. 957.
DOI: 10.1140/epjc/s10052-020-8227-9. arXiv: 2004.03447 [hep-ex].

[25] ATLAS Collaboration. "A search for the $Z\gamma$ decay mode of the Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Physics Letters B* 809 (2020), p. 135754. ISSN: 0370-2693.
DOI: https://doi.org/10.1016/j.physletb.2020.135754.

[26] ATLAS Collaboration. "Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in *pp* collisions at 13 TeV with the ATLAS detector" (July 2020).
DOI: 10.3204/PUBDB-2020-02666. arXiv: 2007.02873 [hep-ex].

[27] ATLAS Collaboration. "Measurement of VH, $H \to b\bar{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector". *Journal of High Energy Physics* 05 (2019), p. 141.
DOI: 10.1007/JHEP05(2019)141. arXiv: 1903.04618 [hep-ex].

[28] CMS Collaboration. "Observation of Higgs Boson Decay to Bottom Quarks". *Phys. Rev. Lett.* 121 (12 Sept. 2018), p. 121801.
DOI: 10.1103/PhysRevLett.121.121801.

[29] CMS Collaboration. "A search for the standard model Higgs boson decaying to charm quarks". *Journal of High Energy Physics* 03 (2020), p. 131.
DOI: 10.1007/JHEP03(2020)131. arXiv: 1912.01662 [hep-ex].

[30] ATLAS Collaboration. "Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment". *Phys. Rev. Lett.* 120 (21 May 2018), p. 211802.
DOI: 10.1103/PhysRevLett.120.211802.

[31] CMS Collaboration. "Measurement of Higgs boson production in the decay channel with a pair of τ leptons". Tech. rep. CMS-PAS-HIG-19-010. Geneva: CERN, 2020.

[32] CMS Collaboration. "Measurement of Higgs boson production and decay to the $\tau\tau$ final state". Tech. rep. CMS-PAS-HIG-18-032. Geneva: CERN, 2019.

[33] ATLAS Collaboration. "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Phys. Rev. D* 99 (7 Apr. 2019), p. 072001.
DOI: 10.1103/PhysRevD.99.072001.

[34] CMS Collaboration. "Measurement of Higgs boson decay to a pair of muons in proton-proton collisions at $\sqrt{s} = 13$ TeV". Tech. rep. CMS-PAS-HIG-19-006. Geneva: CERN, 2020.

[35] ATLAS Collaboration. "A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector" (July 2020). arXiv: 2007.07830 [hep-ex].

[36] S. Badger et al. "Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report" (2016). arXiv: 1605.04692 [hep-ph].

[37] S. Amoroso et al. "Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report" (2020). arXiv: 2003.01700 [hep-ph].

[38] "CERN, LHC Higgs Cross Section WG about STXS". URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGFiducialAndSTXS (visited on 08/20/2020).

[39] R.K. Ellis et al. "Higgs decay to $\tau^+\tau^-$: A possible signature of intermediate mass Higgs bosons at high energy hadron colliders". *Nuclear Physics B* 297.2 (1988), pp. 221–243. ISSN: 0550-3213.
DOI: https://doi.org/10.1016/0550-3213(88)90019-3.

[40] U. Baur and E.W.N. Glover. "Higgs boson production at large transverse momentum in hadronic collisions". *Nuclear Physics B* 339.1 (1990), pp. 38–66. ISSN: 0550-3213.
DOI: https://doi.org/10.1016/0550-3213(90)90532-I.

[41] Esma Mobs. "The CERN accelerator complex. Complexe des accélérateurs du CERN". General Photo. July 2016. URL: https://cds.cern.ch/record/2197559 (visited on 10/15/2020).

[42] Lyndon Evans and Philip Bryant. "LHC Machine". *JINST* 3, S08001 (Aug. 2008).
DOI: 10.1088/1748-0221/3/08/S08001.

[43] "CMS Luminosity - Public Results". URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (visited on 10/21/2020).

[44] CMS Collaboration. "The CMS Experiment at the CERN LHC". *JINST* 3, S08004 (Aug. 2008).
DOI: 10.1088/1748-0221/3/08/S08004.

[45] "Interactive Slice of the CMS detector". URL: https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=4172 (visited on 03/01/2017).

[46] Joram Berger. "Search for the Higgs Boson Produced via Vector-Boson Fusion in the Decay Channel H to Tau Tau". Karlsruhe, KIT. PhD thesis. 2014.

[47] M. Lipinski. "The Phase-1 upgrade of the CMS pixel detector". *Journal of Instrumentation* 12.07 (July 2017), pp. C07009–C07009.
DOI: 10.1088/1748-0221/12/07/c07009.

[48] CMS HCAL/ECAL Collaborations. "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c". *The European Physical Journal C* 60.03 (Apr. 2009), P0359–P0373.
DOI: 10.1140/epjc/s10052-009-0959-5.

[49] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–P10003. DOI: 10.1088/1748-0221/12/10/p10003.

[50] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. DOI: 10.1088/1126-6708/2008/04/063.

[51] CMS Collaboration. "Pileup Removal Algorithms". Tech. rep. CMS-PAS-JME-14-001. Geneva: CERN, 2014.

[52] CMS Collaboration. "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector". Tech. rep. CMS-DP-2018-058. Geneva: CERN, 2018.

[53] CMS Collaboration. "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS". *Journal of Instrumentation* 11.01 (Jan. 2016), P01019–P01019. DOI: 10.1088/1748-0221/11/01/p01019.

[54] CMS Collaboration. "Performance of reconstruction and identification of τ leptons decaying to hadrons and $\nu\tau$ in pp collisions at $\sqrt{s} = 13$ TeV". *Journal of Instrumentation* 13.10 (Oct. 2018), P10005–P10005. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/10/p10005.

[55] CMS Collaboration. "Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons". Tech. rep. CMS-DP-2019-033. Geneva: CERN, 2019.

[56] Daniele Bertolini et al. "Pileup per particle identification". *Journal of High Energy Physics* 2014.10 (Oct. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep10(2014)059.

[57] CMS Collaboration. "Pileup mitigation at CMS in 13 TeV data". *Journal of Instrumentation* 15.09 (Sept. 2020), P09018–P09018. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/09/p09018.

[58] CMS Collaboration. "CMS Technical Design Report for the Level-1 Trigger Upgrade". Tech. rep. CERN-LHCC-2013-011. CMS-TDR-12. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch. June 2013.

[59] V. Khachatryan et al. "The CMS trigger system". *Journal of Instrumentation* 12.01 (Jan. 2017), P01020–P01020. DOI: 10.1088/1748-0221/12/01/p01020.

[60] CMS Collaboration. "CMS High Level Trigger performance at 13 TeV". *PoS* ICHEP2018 (2019), 226. 4 p. DOI: 10.22323/1.340.0226.

[61] CMS Collaboration. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". *JINST* 14.07 (2019), P07004.
DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[62] Artur Gottmann. "Global Interpretation of $\tau\tau$ Events in the Context of the Standard Model and Beyond". Karlsruhe, KIT. PhD thesis. 2020.

[63] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655.
DOI: https://doi.org/10.1016/j.cpc.2015.01.024.

[64] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". *Journal of High Energy Physics* 2014.7 (2014), p. 79. ISSN: 1029-8479.
DOI: 10.1007/JHEP07(2014)079.

[65] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". *Journal of High Energy Physics* 11 (2007), p. 070.
DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[66] S. Agostinelli et al. "Geant4—a simulation toolkit". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002.
DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.

[67] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator.* 2015. arXiv: 1501.04637 [hep-ph].

[68] CMS Collaboration. "An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data". *Journal of Instrumentation* 14.06 (June 2019), P06032–P06032.
DOI: 10.1088/1748-0221/14/06/p06032.

[69] CMS Collaboration. "Measurement of the $Z\gamma^* \to \tau\tau$ cross section in pp collisions at $\sqrt{s} = 13$ TeV and validation of τ lepton analysis techniques". *The European Physical Journal C* 78.9 (Sept. 2018). ISSN: 1434-6052.
DOI: 10.1140/epjc/s10052-018-6146-9.

[70] CMS Collaboration. "Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Journal of High Energy Physics* 2018.9 (Sept. 2018). ISSN: 1029-8479.
DOI: 10.1007/jhep09(2018)007.

[71] Sebastian Brommer. "Production of Hybrid Data Samples for Data Driven Background Determination in the $H \to \tau\tau$ Channel". ETP-KA/2019-05. MA thesis. Karlsruhe Institute of Technology (KIT), 2019.

[72] Stefan Wunsch. "A Novel Strategy for the Standard Model $H \to \tau\tau$ Analysis with Emphasis on Minimizing Systematic Uncertainties in Presence of Modern Multi-Variate Methods". ETP-KA/2017-26. MA thesis. Karlsruhe Institute of Technology (KIT), 2017.

[73] Moritz Scham. "Standard Model H $\to \tau\tau$ Analysis with a Neural Network Trained on a Mix of Simulation and Data Samples". ETP-KA/2020-20. MA thesis. Karlsruhe Institute of Technology (KIT), 2020.

[74] Simon Jörger. "Studies of the usage of neural networks in particle physics analyses". ETP-KA/2020-10. MA thesis. Karlsruhe Institute of Technology (KIT), 2020.

[75] Stefan Wunsch et al. "Identifying the Relevant Dependencies of the Neural Network Response on Characteristics of the Input Space". *Computing and Software for Big Science* 2.5 (2018).
      DOI: 10.1007/s41781-018-0012-1.

[76] Lorenzo Bianchini et al. "Reconstruction of the Higgs mass in H $\to \tau\tau$ Events by Dynamical Likelihood techniques". *Journal of Physics: Conference Series* 513.2 (June 2014), p. 022035.
      DOI: 10.1088/1742-6596/513/2/022035.

[77] Andrei V. Gritsan et al. "Constraining anomalous Higgs boson couplings to the heavy-flavor fermions using matrix element techniques". *Physical Review D* 94.5 (Sept. 2016). ISSN: 2470-0029.
      DOI: 10.1103/physrevd.94.055023.

[78] Robert D. Cousins. "Generalization of the chisquare goodness-of-fit test for binned data using saturated models, with application to histograms". Apr. 2010. URL: http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf (visited on 10/10/2020).

[79] J. S. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra" (2011). arXiv: 1103.0354 [physics.data-an].

[80] Sebastian Wozniewski. "Relative Kalibrierung der IR Absorptionsspektren gegen die H2-, HD- und D2-Konzentrationen". IEKP-BACHELOR-KA/2014-12. BA thesis. Karlsruhe Institute of Technology (KIT), 2014.

[81] Roger Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". *Computer Physics Communications* 77.2 (1993), pp. 219–228. ISSN: 0010-4655.
      DOI: https://doi.org/10.1016/0010-4655(93)90005-W.

[82] ATLAS and CMS Collaborations. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV". *Journal of High Energy Physics* 2016.8 (Aug. 2016). ISSN: 1029-8479.
      DOI: 10.1007/jhep08(2016)045.

[83]   Glen Cowan. "Statistics for Searches at the LHC". *Proceedings, 69th Scottish Universities Summer School in Physics : LHC Phenomenology (SUSSP69): St.Andrews, Scotland, August 19-September 1, 2012*. 2013, pp. 321–355.
       DOI: 10.1007/978-3-319-05362-2_9. arXiv: 1307.2487 [hep-ex].

[84]   Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". *Eur.Phys.J.* C71 (2011), p. 1554.
       DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[85]   Glen Cowan et al. "Erratum to: Asymptotic formulae for likelihood-based tests of new physics". *The European Physical Journal C* 73.7 (2013), p. 2501. ISSN: 1434-6052.
       DOI: 10.1140/epjc/s10052-013-2501-z.

Danksagung