CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2021/03/20

Search for the decay of a heavy Higgs boson H into two lighter Higgs bosons h and h_S in the $h(\tau\tau)h_S(bb)$ final state at 13 TeV

The CMS Collaboration

Abstract

A search for the decay of a heavy Higgs boson H into the observed Higgs boson h and another Higgs boson h_S with a mass of $m_{h_S} < m_H - m_h$ is presented. The h and h_S bosons are required to decay into a pair of tau leptons and a pair of b quarks, respectively. The search uses 137 fb^{-1} of proton-proton collisions collected with the CMS detector at a center-of-mass energy of 13 TeV. A mass range of $240 - 3000$ GeV for m_H and $60 - 2800$ GeV for m_{h_S} is covered. No signal has been observed. Therefore, model independent 95% confidence level upper limits on the product of the production cross section and the branching fractions of the signal process are set ranging from 125 fb (for $m_H = 240$ GeV) to 2.7 fb (for $m_H = 3000$ GeV). These limits are compared to predictions of the next-to-minimal supersymmetric extension of the standard model.

1 Introduction

The discovery of a Higgs boson h with a mass of 125 GeV at the CERN LHC [1–3] has turned the standard model of particle physics (SM) into a mathematically consistent theory, which could be realized up to the Planck scale. To date all properties of the observed particle are in agreement with the expectations of the SM within an experimental precision of between 5 to 20% [4–7]. Despite its success in describing a large wealth of phenomena in and beyond particle physics, the SM falls short of addressing a number of fundamental theoretical questions and striking observations in nature. In this respect it is considered still incomplete from the physics point of view.

A well motivated and therefore intensively studied extension of the SM is the addition of supersymmetry (SUSY) transforming bosons into fermions and vice versa [8, 9]. SUSY postulates a bosonic (fermionic) partner particle for each SM fermion (boson), with the same properties as the corresponding SM particle apart from its (half-)integer spin. The fact that to date no such SUSY particles have been observed implies that if SUSY were realized in nature it must be a broken symmetry. Apart from the prediction of a sizable number of new particles, SUSY requires the extension of the Brout-Englert-Higgs mechanism part [10–15] of the SM Lagrangian density. In the minimal supersymmetric extension of the SM (MSSM) [16, 17] one more doublet field is introduced with respect to the SM, leading to the prediction of two charged and three neutral Higgs bosons, one of which can be associated with h. A further extension of the MSSM by one additional singlet field S is theoretically well motivated by the so called "μ-problem" of the MSSM [18], which relates to the fact that in the MSSM superpotential a mass parameter μ has to be introduced for the Higgs doublet fields before spontaneous symmetry breaking. This dimensional μ parameter has to be chosen by hand close to the electroweak symmetry breaking scale $v = 246\,\text{GeV}$ to be consistent with observation. Adding S to the superpotential allows naturally to obtain v from the vacuum expectation value of S after electroweak symmetry breaking. The extension of the MSSM by S leads to the next-to-minimal supersymmetric SM (NMSSM), as reviewed, e.g., in Refs. [19, 20]. Since S is a complex singlet field this increases the number of predicted Higgs bosons by two, resulting in two charged and five neutral Higgs bosons, of which three are scalar and two are pseudoscalar in nature.

Many searches for additional Higgs bosons in the context of the MSSM have been performed by the LHC experiments. In the absence of signal these have led to the exclusion of large parts of the MSSM parameter space for masses of the additional neutral Higgs bosons up to $\approx$2 TeV [21–24]. The parameter space of the NMSSM on the other hand is still largely unconstrained [25].

This search focuses on the $\mathrm{H} \to \mathrm{h}\mathrm{h}_\mathrm{S}$ decay of a heavy Higgs boson H decaying into an h and another neutral Higgs boson h_S with a mass of $m_{\mathrm{h}_\mathrm{S}} < m_\mathrm{H} - m_\mathrm{h}$. The search is inspired by the NMSSM, where the h_S could have a dominant admixture of the additional singlet field S, leading to a significant suppression of its couplings to SM particles and thus of its direct production at the LHC [1]. In this case the production of an H boson and subsequent decay into $\mathrm{h}\mathrm{h}_\mathrm{S}$ would become the dominant process for h_S production. Despite the overall reduced coupling strengths to SM particles, the branching fractions of the h_S into SM particles are still expected to be similar to the h. A promising signature for the search is thus given by the decay of the h into a pair of tau leptons and the decay of the h_S into a pair of b quarks, $\mathrm{h}(\tau\tau)\mathrm{h}_\mathrm{S}(\mathrm{bb})$. For better readability we will not distinguish fermions by particle or anti-particle in this final state in subsequent notation throughout the text. The partonic process of H boson production

[1] Note that here we use the NMSSM as a concrete proxy motivating the decay channel that we are searching for. But also from a model independent perspective of an effective two Higgs doublet plus singlet model the h_S boson would have to have a significant admixture of a singlet field S for the arguments given in the text.

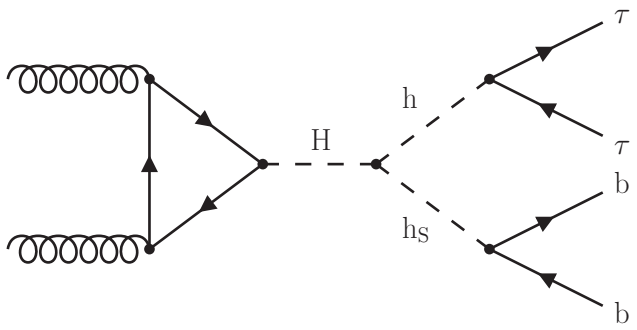

Figure 1: Feynman diagram of the $gg \to H \to h(\tau\tau)h_S(bb)$ process.

from gluon fusion up to the $\tau\tau bb$ final state is illustrated in Fig. 1. The search is performed in the mass ranges of 240 GeV $\leq m_H \leq$ 3000 GeV and 60 GeV $\leq m_{h_s} \leq$ 2800 GeV. It is the first search for this signature at the LHC.

After a brief introduction of the CMS detector and event reconstruction in Sections 2 and 3 the event selection and categorization are described in Section 4. In Section 5 the complete model used to describe the data and infer the potential signal including systematic uncertainties is introduced. The results of the search are described in Section 6 and the search is summarized in Section 7.

2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid.

The silicon tracker measures charged particles within the pseudorapidity range of $|\eta| < 2.5$. During the LHC data taking period up to 2017, the silicon tracker consisted of 1440 silicon pixel and 15 148 silicon strip detector modules. From 2017 on the silicon pixel detector was upgraded to 1856 modules. For nonisolated particles with a transverse momentum of $1 < p_T < 10$ GeV with respect to the beam axis and $|\eta| < 1.4$, the track resolutions are typically 1.5% in p_T and 25–90 (45–150) μm in the transverse (longitudinal) impact parameter [26]. From 2017 on the transverse impact parameter resolution improved to 20–60 μm when restricted to the same η range as before and 20–75 μm in the increased full η range [27].

The electron momentum is estimated by combining the energy measurement in the ECAL with the momentum measurement in the tracker. The momentum resolution for electrons with $p_T \approx$ 45 GeV from $Z \to ee$ decays ranges from 1.7 to 4.5%. It is generally better in the barrel region than in the endcaps, and also depends on the bremsstrahlung energy emitted by the electron traversing the material in front of the ECAL [28].

Muons are measured in the pseudorapidity range of $|\eta| < 2.4$, with detection planes made using three technologies: drift tubes, cathode strip chambers, and resistive plate chambers. Matching muons to tracks measured in the silicon tracker results in a relative p_T resolution, for

muons with $20 < p_T < 100\,\text{GeV}$, of 1.3–2.0% in the barrel and better than 6% in the endcaps. The p_T resolution in the barrel is better than 10% for muons with p_T up to 1 TeV [29].

In the barrel section of the ECAL, an energy resolution of about 1% is achieved for unconverted or late-converting photons that have energies in the range of tens of GeV. The remaining barrel photons have a resolution of about 1.3% up to $|\eta| = 1$, rising to about 2.5% at $|\eta| = 1.4$. In the endcaps, the resolution of unconverted or late-converting photons is about 2.5%, while the remaining endcap photons have a resolution between 3 and 4% [30]. When combining information from the entire detector, the jet energy resolution amounts typically to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [31].

Events of interest are selected using a two-tiered trigger system. The first level (L1), composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of about 4 μs [32]. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to around 1 kHz before data storage [33].

A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [34].

3 Event reconstruction

The reconstruction of the proton-proton (pp) collision products is based on the particle-flow (PF) algorithm as described in Ref. [35], combining the available information from all CMS subdetectors to reconstruct an unambiguous set of individual particle candidates, categorized into electrons, photons, muons, charged hadrons, and neutral hadrons. During the LHC Run 2 data taking period the CMS experiment was operating with, on average, 29 inelastic pp collisions per bunch crossing under varying beam conditions. The fully recorded data of a bunch crossing defines an *event* for further processing. The candidate vertex with the largest value of summed physics-object p_T^2 is taken to be the primary pp interaction vertex (PV). The physics objects for this purpose are the jets, clustered using the anti-k_T jet clustering algorithm as implemented in the FASTJET package [36] with the tracks assigned to the corresponding candidate vertex as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the p_T of those jets. Any other collision vertices in the event are associated with additional soft inelastic pp collisions called *pileup* (PU).

Electrons are reconstructed by combining clusters of energy deposits in the ECAL with hits in the tracker [28, 37]. To increase their purity, reconstructed electrons are required to pass a multivariate electron identification discriminant, which combines information on track quality, shower shape, and kinematic quantities. For this analysis a working point with an identification efficiency of 90% is used for a misidentification rate of $\approx$1%.

Muons in the event are reconstructed by performing a simultaneous track fit to hits in the tracker and in the muon chambers [38]. The presence of hits in the muon chambers already leads to a strong suppression of particles misidentified as muons. Additional identification requirements on the track fit quality and the compatibility of individual track segments with the fitted track can reduce the misidentification rate further. For this analysis muon identification requirements with an efficiency of $\approx$99% are chosen.

The contributions from backgrounds to the electron (muon) selection are further reduced by requiring the corresponding lepton to be isolated from any hadronic activity in the detector.

This property is quantified by an isolation variable

$$I_{\text{rel}}^{\text{e}(\mu)} = \frac{1}{p_{\text{T}}^{\text{e}(\mu)}} \left(\sum p_{\text{T}}^{\text{charged}} + \max \left(0, \sum E_{\text{T}}^{\text{neutral}} + \sum E_{\text{T}}^{\gamma} - p_{\text{T}}^{\text{PU}} \right) \right), \tag{1}$$

where $p_{\text{T}}^{\text{e}(\mu)}$ corresponds to the electron (muon) p_{T} and $\sum p_{\text{T}}^{\text{charged}}$, $\sum E_{\text{T}}^{\text{neutral}}$, and $\sum E_{\text{T}}^{\gamma}$ to the sum of the p_{T} (E_{T}) of all charged particles, neutral hadrons, and photons, in a predefined cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around the lepton direction at the PV, where $\Delta\eta$ and $\Delta\phi$ (measured in radians) correspond to the angular distances of the particle to the lepton in the η and ϕ directions. The chosen cone sizes are $\Delta R < 0.3$ for electrons and 0.4 for muons. The lepton itself is excluded from the calculation. To mitigate any distortions from PU, only those charged particles whose tracks are associated with the PV are taken into account. Since for neutral hadrons and photons an unambiguous association to the PV or PU is not possible an estimate of the contribution from PU (p_{T}^{PU}) is subtracted from the sum of $\sum E_{\text{T}}^{\text{neutral}}$ and $\sum E_{\text{T}}^{\gamma}$. In the case of negative values the result is set to zero.

For further characterization of the event all reconstructed PF objects are clustered into jets using the anti-k_{T} jet clustering algorithm with a distance parameter of 0.4. To identify jets resulting from the hadronization of b quarks (b jets) the DEEPJET algorithm is used as described in Refs. [39, 40]. In this analysis a working point of this algorithm is chosen that corresponds to a b jet identification efficiency of ≈80% for a misidentification rate for jets originating from light quarks and gluons of 1% [41]. Jets with $p_{\text{T}} > 30\,\text{GeV}$ and $|\eta| < 4.7$ and b jets with $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.4\,(2.5)$ are used, where the value in braces corresponds to the selection after the upgrade of the silicon pixel detector from 2017 on.

Jets are also used as seeds for the reconstruction of hadronic τ decays (τ_{h}). This is done by further exploiting the substructure of the jets, using the *hadrons-plus-strips* algorithm, as described in Ref. [42]. For the analysis the decays into one or three charged hadrons with up to two neutral pions with $p_{\text{T}} > 2.5\,\text{GeV}$ are used. The neutral pions are reconstructed as *strips* with dynamic size in η–ϕ from reconstructed electrons and photons contained in the seeding jet, where the strip size varies as a function of the p_{T} of the electron or photon candidate. The τ_{h} decay mode is then obtained by combining the charged hadrons with the strips. To distinguish τ_{h} decays from jets originating from the hadronization of quarks or gluons, electrons, and muons, the DEEPTAU algorithm is used as described in [43]. This algorithm exploits the information of $\gtrsim$100 000 features of the reconstructed event record, comprising the kinematic and object identification properties of the PF candidates in the vicinity of the τ_{h} candidate and the τ_{h} candidate itself, and several characterizing quantities of the whole event. It results in a multi-classification output $y_{\alpha}^{\text{DT}}\,(\alpha = \tau, \text{jet}, \text{e}, \mu)$ equivalent to a Bayesian probability of the τ_{h} candidate to originate from a genuine tau, the hadronization of a quark or gluon, an electron, or a muon. From this output three discriminators are built according to

$$D_{\alpha} = \frac{y_{\tau}^{\text{DT}}}{y_{\tau}^{\text{DT}} + y_{\alpha}^{\text{DT}}}, \quad \alpha = \text{jet, e, } \mu. \tag{2}$$

For this analysis a working point of D_{jet} with an efficiency of 70% for a misidentification rate of 10^{-2} is chosen. For D_{e} and D_{μ}, depending on the $\tau\tau$ final state, different working points with efficiencies of 80% and >99% with misidentifications rates between 10^{-4} and 10^{-2} are chosen.

The pileup per particle identification (PUPPI) algorithm [44] is applied to reduce the PU dependence of the $\vec{p}_{\text{T}}^{\text{miss}}$ observable. The $\vec{p}_{\text{T}}^{\text{miss}}$ is computed from the PF candidates weighted by their probability to originate from the PV [45]. Its magnitude is referred to as $p_{\text{T}}^{\text{miss}}$. It is used

Table 1: Offline selection requirements applied to electrons, muons, and τ_h candidates used for the selection of the τ pair. The p_T values in braces correspond to the selection criteria for events selected by a single electron or single muon trigger. These requirements depend on the year of data taking. For D_{jet} the efficiency and for $D_{e(\mu)}$ the rejection rates for the chosen working points are given in braces. A detailed discussion is given in the text.

Final state	Electron/Muon	τ_h
$e\tau_h$	$p_T > 25(26, 28, 33)$ GeV $\lvert\eta\rvert < 2.1$ $I^e_{rel} < 0.15$	$p_T > 35(30)$ GeV $\lvert\eta\rvert < 2.3$ $D_{jet}(70\%)$, $D_e(10^{-4})$, $D_\mu(10^{-3})$
$\mu\tau_h$	$p_T > 20(23, 25)$ GeV $\lvert\eta\rvert < 2.1$ $I^\mu_{rel} < 0.15$	$p_T > 35(30)$ GeV $\lvert\eta\rvert < 2.3$ $D_{jet}(70\%)$, $D_e(10^{-2})$, $D_\mu(10^{-4})$
$\tau_h\tau_h$	–	$p_T > 40$ GeV $\lvert\eta\rvert < 2.1$ $D_{jet}(70\%)$, $D_e(10^{-2})$, $D_\mu(10^{-3})$

for the estimation of the invariant mass of the two tau leptons before their decay, as discussed in Section 4.

4 Event selection and classification

4.1 Event selection

The analyzed data correspond to an integrated luminosity of 137 fb^{-1} of pp collisions collected at a center-of-mass energy of 13 TeV. The selection is targeted at the reconstruction of a pair of tau leptons originating from the h boson with a mass of $m_{\tau\tau} = 125$ GeV and a pair of b quarks originating from the h_S boson with a mass varying between 60 and 2800 GeV. For the τ pair the $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states are used. In the online HLT step the selection is based on the presence of a single electron or muon, an $e\tau_h(\mu\tau_h)$ pair, or a $\tau_h\tau_h$ pair in the event, depending on the final state. In the offline selection further requirements on the p_T, η, $I^{e(\mu)}_{rel}$, and the D_α discriminators are applied in addition to the object identification requirements described in Section 3, as summarized in Table 1.

In the $e\tau_h$ ($\mu\tau_h$) final state, an electron (muon) with at least 25(20) GeV is required, if the event was selected by a trigger based on the presence of the $e\tau_h$ ($\mu\tau_h$) pair in the event. If the event was selected by a single electron trigger, the p_T requirement on the electron is increased to 26–33 GeV depending on the data-taking period, to ensure a sufficiently high efficiency of the HLT selection. For muons, the p_T requirement is increased to 23 GeV for 2016, and 25 GeV for 2017 or 2018, if selected by a single muon trigger. The electron (muon) is required to be contained in the central detector with $\lvert\eta\rvert < 2.1$, and to be isolated from any hadronic activity in the detector according to $I^{e(\mu)}_{rel} < 0.15$. The τ_h candidate is required to have $\lvert\eta\rvert < 2.3$ and $p_T > 35$ GeV if selected by an $e\tau_h(\mu\tau_h)$ pair trigger, or $p_T > 30$ GeV if selected by a single electron (muon) trigger. In the $\tau_h\tau_h$ final state, both τ_h candidates are required to have $\lvert\eta\rvert < 2.1$ and $p_T > 40$ GeV. The working points of the DEEPTAU discriminants as described in Section 3 are chosen depending on the final state. Events with additional leptons fulfilling looser selection criteria are discarded to avoid the assignment of single events to more than one final state.

The selected τ decay candidates are required to be of opposite charge and to be separated by

more than $\Delta R = 0.5$ in the $\eta - \phi$ plane. The closest distance of their tracks to the PV is required to be $d_z < 0.2\,\text{cm}$ along the beam axis. For electrons and muons, an additional requirement of $d_{xy} < 0.045\,\text{cm}$ in the transverse plane is applied. In rare cases in which more than one τ_h candidate fulfilling all selection requirements is found, the candidate with the higher score of D_jet is chosen.

In addition to the di-τ pair at least one b jet fulfilling the selection criteria as described in Section 3 is required. Events without b jets are removed from the analysis. If at least two b jets exist, the pair is built from the two b jets leading in p_T. If only one b jet exists the b pair is built using the b jet and the jet with the highest b jet score of the DEEPJET classifier. Events that contain only one b jet and no further jet are also rejected from the event selection. The energies of the two jets used for the b pair are corrected using the multivariate energy-momentum regression described in Ref. [46].

4.2 Event classification

After this selection all remaining events are further categorized to enrich signal-like events and isolate different background processes in control regions. The latter is used to further constrain systematic uncertainties in the background estimates during the statistical inference of the signal. This categorization is based on neural network (NN) multi-classification exploiting fully connected feed-forward NNs with two hidden layers of 200 nodes each, and five output nodes. Each output node represents one of the following distinct event classes in consideration: (i) Events containing genuine τ pairs (labeled "$\tau\tau$"); (ii) events with quark or gluon induced jets misidentified as τ_h (labeled "jet $\rightarrow \tau_\text{h}$"); (iii) top quark pair events where the intermediate W boson decays into an electron or muon (not included in (i) or (ii); labeled as "tt"); (iv) remaining background processes that are of minor importance for the analysis and not included in any of the previous event classes, yet (labeled as "misc"); (v) $\text{H} \rightarrow \text{h}(\tau\tau)\text{h}_\text{S}(\text{bb})$ signal events (labeled as "signal"). Event class (iv) comprises diboson production, single t quark production, Z boson decays to electrons or muons, and single h boson production. For the latter, production rates and branching fractions as predicted by the SM are assumed. For each node in the hidden layers the hyperbolic tangent is chosen as the activation function. The activation function for the output layer is chosen to be the softmax function allowing for a Bayesian conditional probability interpretation $y_i^{(k)}$ of an event k to be associated to an event class i, given its input features $\vec{x}^{(k)}$ to the NN. The highest value of $y_i^{(k)}$, $\max(y_i^{(k)})$, defines which class the event is associated to and will define the discriminator for the statistical inference of the signal. All other outputs $y_j^{(k)}$, $j \neq i$ are discarded from any further consideration so that any event is used only once for the statistical inference of the signal. The choice of event classes, which is based on observable signatures rather than explicit processes, closely resembles the data model described in Section 5.

In the $\text{e}\tau_\text{h}$ and $\mu\tau_\text{h}$ final states the input space to the NNs is spanned by 20 features $\vec{x}$ of the event including the p_T of the τ candidates and the jets forming the b quark pair; mass and p_T estimates of the τ pair, b quark pair, and $\tau\tau\text{bb}$ system; the number of (b) jets; and further kinematic properties of the selected jets. For this purpose a likelihood based estimate of the $\tau\tau$ mass before decay [47] and a kinematic fit to the $\tau\tau\text{bb}$ system for each given m_H and m_{h_S} hypothesis, similar to the approach described in Ref. [48], are used. In the $\tau_\text{h}\tau_\text{h}$ final state these features are complemented by the masses of the two jets used for the b quark pair system and their associated output values of the DEEPJET algorithm, to allow for a better discrimination of genuine b jets from light quark or gluon induced jets. All input features have been selected from a superset of variables describing the properties of the event exploiting a ranking of indi-

vidual features and pairwise correlations of features as described in Ref. [49].

Since the kinematic properties of the signal strongly vary across the probed ranges of m_{H} and $m_{\mathrm{h_S}}$ a total of 68 NNs per final state are used for classification, which within each final state only differ by the kinematic properties of the signal that are used for training. For this purpose, adjacent sets of mass points are combined into single signal classes. Four mass points are combined for masses up to 1 000 GeV, and up to nine mass points beyond. Differences of the input features depending on the year of data taking are taken into account by a conditional training using a one-hot encoding of the data taking year in the NN training.

The parameters to be optimized during training are the weights ($\{w_a\}$) and biases ($\{b_b\}$) of the NN output functions y_i. Before training the weights are initialized with random numbers using the Glorot initialization technique [50] with values drawn from a uniform distribution. The biases are initialized with zero. The trainings are then performed using randomly sampled batches of $N = 30$ events per event class, drawn from the training datasets using a balanced batch approach [51] . This approach has shown improved convergence properties on training samples with highly imbalanced lengths. The classification task is encoded in the NN loss function, chosen as the cross entropy

$$L\left(\{y_i^{(k)}\},\{y_j'^{(k)}\}\right) = -\sum_{k=1}^{N} y_j'^{(k)} \log\left(y_i^{(k)}(\{w_a\},\{b_b\},\{\vec{x}^{(k)}\})\right) \; ; \qquad y_j'^{(k)} = \delta_{ij} \, , \tag{3}$$

which is to be minimized during the NN trainings. In Eq. (3) k runs over the events in the batch, on which L is evaluated. The NN output function for event k to belong to category i is given by $y_i^{(k)}$. The function $y_j'^{(k)}$ encodes the prior knowledge of the training. It is 1 if class i of event k coincides with the true event class j, and 0 otherwise. The $y_i^{(k)}$ depend on the weights, biases, and input features $\{\vec{x}^{(k)}\}$ of event k to the NN. The batch definition guarantees that each true event class enters the training with equal weight in the evaluation of L, i.e., without prevalence. Within the misc event class all contained processes are normalized according to their expected rates with respect to each other. On each batch a gradient step is applied defined by the partial derivatives of L in each weight, w_a, and bias, b_b, using the Adam minimization algorithm [52], with a constant multiplicative learning rate of 10^{-4}. To guarantee statistical independence those events that are used for training are not used for any other step of the analysis.

The performance of the NNs during training is monitored by evaluating L on a validation subset that contains a fraction of 25% of randomly chosen events from the training sample, which are excluded from the gradient computation. The training is stopped if the evaluation of L on the validation dataset does not indicate any further decrease for a sequence of 50 epochs, where an epoch is defined by 1 000 (100) batches in the $\mathrm{e}\tau_{\mathrm{h}}/\mu\tau_{\mathrm{h}}$ ($\tau_{\mathrm{h}}\tau_{\mathrm{h}}$) final state. The NNs used for the analysis are then defined by the weights and biases of the epoch with the minimal value of L on the validation sample. To improve the generalization property of the NNs, two regularization techniques are introduced. Firstly, after each hidden layer a layer with a dropout probability of 30% is added. Secondly, the weights of the NNs are subject to an L2 (Tikhonov) regularization [53] with a regularization factor of 10^{-5}.

After training, a very good separation between the non-peaking background events and the signal events is achieved, with a purity and classification sensitivity for the correct signal class of typically more than 80%.

5 Data model

The model used to infer the signal from the data is defined by an extended binned likelihood of the form

$$\mathcal{L}\left(\{k_i\}|\,\mu\,S(\{\theta_j\})+B(\{\theta_j\})\right)=\prod_i \mathcal{P}(k_i|\mu\,S_i(\{\theta_j\})+B_i(\{\theta_j\}))\prod_j \mathcal{C}(\hat{\theta}_j|\theta_j), \tag{4}$$

where i labels all bins of the distributions of the NN output functions $\max(y_i)$ of each signal and background class defined in Section 4, split by $\tau\tau$ final state and year of data taking, resulting in 45 individual input histograms for each pair of m_{H} and $m_{\mathrm{h_S}}$. The function $\mathcal{P}(k_i|\mu\,S_i(\{\theta_j\})+B_i(\{\theta_j\}))$ corresponds to the Poisson density to observe k_i events in bin i for a prediction of S_i signal and a total of B_i background events. The parameter μ is a single scaling parameter of the signal.

Systematic uncertainties are incorporated as penalty terms for additional nuisance parameters $\{\theta_j\}$ in the likelihood, appearing as a product with predefined probability density functions $\mathcal{C}(\hat{\theta}_j|\theta_j)$ to obtain a maximum likelihood estimate $\hat{\theta}_j$ for a given true value of θ_j, during the statistical inference of the signal.

All SM background sources of relevance after the event selection described in Section 4 are listed in Table 2. In the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states, the most abundant source of background is the production of top quark pairs ($\mathrm{t\bar{t}}$) that can easily result in a signature with genuine leptons and b quarks. After the event selection the expected fraction of $\mathrm{t\bar{t}}$ events in these final states is ≈70%. In the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state, events containing purely QCD induced gluon and light quark jets, referred to as QCD multijet production in the following, and the decay of Z bosons into tau leptons form the largest background sources with ≈35% each.

For the background modeling three different methods are used depending on the interpreted signature after reconstruction: $\tau\tau$ events are obtained from the τ-embedding method, discussed in Section 5.1; jet $\rightarrow \tau_{\mathrm{h}}$ events are obtained from the F_{F}-method, discussed in Section 5.2; all other background events are obtained from full event simulation, discussed in Section 5.3. The signal events are also obtained from simulation.

5.1 The τ-embedding method

For all events in which the decay of a Z or two W bosons results in two genuine tau leptons, the τ-embedding method as described in Ref. [54] is used. For this purpose $\mu\mu$ events are selected in data. All energy deposits of the muons are removed from the event record and replaced by simulated tau lepton decays with the same kinematic properties as the selected muons. In this way the method relies only on the simulation of the well-understood tau lepton decay and its energy deposits in the detector, while all other parts of the event, like the identification and reconstruction of (b) jets or the non-τ related parts of $p_{\mathrm{T}}^{\mathrm{miss}}$ are obtained from data. This results in an improved modeling of the data compared to the simulation of the full process. In turn several simulation-to-data corrections, as detailed in Section 5.4, are not needed. The selected muons predominantly originate from Z boson decays, however contributions from other processes resulting in two genuine tau leptons, like $\mathrm{t\bar{t}}$ or diboson production, are also covered by this model. For the selection described in Section 4, with the requirement of at least one b jet in the event, 84% of the $\mu\mu$ events selected for the τ-embedding method are expected to originate from Z boson decays, 14% from $\mathrm{t\bar{t}}$ production, and ≈2% from diboson production.

Table 2: Background processes contributing to the event selection, as given in Section 4. The symbol ℓ corresponds to an electron or muon. The second column refers to the experimental signature in the analysis, the last three columns indicate the estimation methods used to model each corresponding signature as described in Sections 5.1 – 5.3.

		Estimation method		
Background process	Final state signature	τ-emb.	F_{F}	Sim.
Z	$\tau\tau$	✓	–	–
	Jet $\to \tau_{\mathrm{h}}$	–	✓	–
	$\ell\ell$	–	–	✓
$\mathrm{t\bar{t}}$	$\tau\tau + X$	✓	–	–
	Jet $\to \tau_{\mathrm{h}}$	–	✓	–
	$\ell + X$	–	–	✓
Diboson+single t	$\tau\tau + X$	✓	–	–
	Jet $\to \tau_{\mathrm{h}}$	–	✓	–
	$\ell + X$	–	–	✓
W+jets	Jet $\to \tau_{\mathrm{h}}$	–	✓	–
QCD multijet	Jet $\to \tau_{\mathrm{h}}$	–	✓	–
Single h	$\tau\tau$	–	–	✓
	bb	–	–	✓

$\ell = \mathrm{e}, \mu$

5.2 The F_{F}-method

The main contributing processes to jet $\to \tau_{\mathrm{h}}$ events are QCD multijet production, the production of W bosons in association with jets (W+jets), and $\mathrm{t\bar{t}}$ production. These events are estimated using the F_{F}-method as described in Refs. [22, 55]. For this purpose the complete kinematic phase space is split into the disjoint signal region (SR), application region (AR), and determination regions (DR^i). The SR and the AR differ only in the working point chosen for the identification of the τ_{h} candidate, where for the AR a looser working point is chosen and the events from the SR are excluded. Three independent extrapolation factors F_{F}^i are then derived for QCD multijet, W+jets, and $\mathrm{t\bar{t}}$ production in three dedicated DR^i, defined to enrich each corresponding process. The F_{F}^i are then used to estimate the yields N_{SR}^i and kinematic properties of each corresponding background in the SR from the number of events N_{AR}^i in the AR according to

$$N_{\mathrm{SR}}^i = F_{\mathrm{F}}^i \, N_{\mathrm{AR}}^i \qquad i = \mathrm{QCD, W{+}jets, t\bar{t}}\,. \tag{5}$$

For the estimate of $F_{\mathrm{F}}^{\mathrm{QCD}}$ the charges of the two selected τ decay products are required to be of same sign. For the estimation of $F_{\mathrm{F}}^{\mathrm{W+jets}}$ the presence of no b jet and a high transverse mass of the lepton-$p_{\mathrm{T}}^{\mathrm{miss}}$ system are required. The estimation of $F_{\mathrm{F}}^{\mathrm{t\bar{t}}}$ is obtained from simulation. Each F_{F}^i is derived on an event-by-event basis, as a function of the p_{T} of the τ_{h} candidate, the p_{T} of the second τ decay in the event, and the mass of the visible $\tau\tau$ decay products. Each F_{F}^i is further subject to a number of non-closure corrections derived from control regions in data to take subleading dependencies of the F_{F}^i into account. Finally the F_{F}^i are combined into a weighted sum, using the simulation-based estimation of the fraction w_i of each process in the AR to derive the

final factor

$$F_{\mathrm{F}} = \sum_i w_i F_{\mathrm{F}}^i \tag{6}$$

to be used to model this background.

5.3 Simulation

In the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state the τ-embedding and F_{F}-methods cover ≈95% of all expected background events. In the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states the fraction of expected background events described by these two methods is ≈42%. All remaining events originate from processes like Z boson, $\mathrm{t\bar{t}}$, or diboson production, where at least one decay of a vector boson into an electron or muon is not covered by any of the two methods. These and the signal events are modeled using the simulation of the full processes.

The production of Z bosons in the ee and $\mu\mu$ final states is simulated at leading-order (LO) precision in the coupling strength α_{S}, using the MADGRAPH5_aMC@NLO 2.2.2 event generator [56, 57]. To increase the number of simulated events in regions of high signal purity supplementary samples are generated with up to four outgoing partons in the hard interaction. For diboson production MADGRAPH5_aMC@NLO is used at next-to-LO (NLO) precision in α_{S}. For $\mathrm{t\bar{t}}$ and single t quark production samples are generated at NLO precision using POWHEG 2.0 [58–63]. The kinematic properties of single h boson production are simulated using POWHEG separately for the production via gluon fusion, vector boson fusion, or in association with a Z boson, W boson, or a top quark pair. For this purpose the h is assumed to behave as expected from the SM.

When compared to data, Z boson, $\mathrm{t\bar{t}}$, and single t quark events in the tW-channel are normalized to their cross sections at next-to-NLO precision [64–66]. Single t quark production in the *t*-channel and diboson events are normalized to their cross sections at NLO precision or higher [66–68].

The signal process $\mathrm{H} \rightarrow \mathrm{h_S h}$ is generated using MADGRAPH5_aMC@NLO at LO precision. The analysis is restricted to the production of the H via gluon fusion, which is expected to be dominant, e.g., in the NMSSM. Due to the two unknown masses involved in the decay, a two dimensional grid of signal mass pairs is generated, resulting in 420 mass pairs spanning from 240 to 3 000 GeV in m_{H} and 60 to 2 800 GeV in $m_{\mathrm{h_S}}$ only taking pairs with $m_{\mathrm{h_S}} + 125\,\mathrm{GeV} \leq m_{\mathrm{H}}$ into account.

For the generation of all signal and background processes the NNPDF3.0 [69] (NNPDF3.1 [70]) parton distribution functions are used for the data taken in 2016 (from 2017 on). The description of the underlying event is parameterized according to the CUETP8M1 [71] and CP5 [72] tunes. Hadronic showering and hadronization, as well as the τ lepton decays, are modeled using the PYTHIA event generator [73, 74]. For all simulated events, additional inclusive inelastic pp collisions generated with PYTHIA are added according to the expected PU profile in data to take the effect of the observed PU into account. All events generated are passed through a GEANT4-based [75] simulation of the CMS detector and reconstructed using the same version of the CMS event reconstruction software as used for the data.

5.4 Corrections and control of the model

The capability of the model to describe the data is monitored in various control regions orthogonal to the selection used for the measurement and corrections and corresponding uncertainties are derived where necessary.

The following corrections equally apply to simulated and τ-embedded events, where the τ decay is also simulated. Since the simulation part for τ-embedded events happens under different detector conditions as for simulated events, corrections and related uncertainties may differ. Corrections are derived for residual differences in the efficiency of the selected triggers, differences in the electron and muon tracking efficiency, and in the efficiency of the identification and isolation requirements for electrons and muons. These corrections are obtained in bins of p_T and η of the corresponding lepton, using the tag-and-probe method as described in Ref. [76] with $Z \to ee$ and $Z \to \mu\mu$ events. They usually amount to not more than a few percent.

In a similar way, corrections are obtained for the efficiency of triggering the τ_h decays and for the τ_h-identification efficiency. The latter are derived as a function of the p_T of the τ_h in four bins below 40 GeV and one bin above. For $p_T(\tau_h) > 40$ GeV a correction is also derived for each τ_h decay mode individually, which is made use of only in the $\tau_h\tau_h$ final state. Corrections to the energy scale of the τ_h decays and to electrons misidentified as τ_h are derived for each year of data taking and each τ_h decay mode individually, from likelihood scans of discriminating observables. For the trigger efficiency the correction is obtained from parametric fits to the trigger turn-on curves derived for each corresponding sample and data.

The following corrections only apply to fully simulated events. During the 2016 and 2017 data taking, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the region at $|\eta| > 2.0$ caused a specific trigger inefficiency. For events containing an electron (a jet) with p_T larger than ≈ 50 GeV (≈ 100 GeV), in the region of $2.5 < |\eta| < 3.0$ the efficiency loss is 10–20%, depending on p_T, η, and time. Corresponding corrections have been derived from data and applied to the simulation.

The energies of jets are corrected to the expected response of the jet at the stable hadron level, using corrections measured in bins of the jet p_T and η. These corrections are usually not larger than 10–15%. Residual data-to-simulation corrections are applied to the simulated event samples. They usually range between sub-percent level at high jet p_T in the central part of the detector to a few percent in the forward region. A correction is applied to the direction and magnitude of $\vec{p}_T^{\,\text{miss}}$ based on differences between estimates of the hadronic recoil in $Z \to \mu\mu$ events in data and simulation. This correction is applied to the simulated Z boson, single h boson, and signal events, where a hadronic recoil against a single particle is well defined. The efficiencies for genuine and misidentified b jets to pass the working points of the b jet identification discriminator as given in Section 4 are determined from data, using $t\bar{t}$ events for genuine b jets and Z boson production in association with jets for jets originating from light quarks. Data-to-simulation corrections are obtained for these efficiencies and used to correct the number of b jets in the simulation.

Data-to-simulation corrections are further applied to $Z \to ee$ ($Z \to \mu\mu$) events in the $e\tau_h$ ($\mu\tau_h$) and $\tau_h\tau_h$ final states in which an electron (muon) is reconstructed as a τ_h candidate, to account for residual differences in the $e(\mu) \to \tau_h$-misidentification rate between data and simulation. Deficiencies in the modeling of Z boson events in the ee, $\mu\mu$ final states are corrected for by a weighting of the simulated $Z \to \mu\mu$ events to data in bins of $p_T^{\mu\mu}$ and $m_{\mu\mu}$. In addition all simulated $t\bar{t}$ events are weighted to better match the top quark p_T distribution, as observed in data [77].

The overall normalization of all backgrounds is constrained by the dedicated event categories described in Section 4. After the event selection and prior to the event classification, i.e., still at an inclusive state of the analysis, the marginal distributions and pairwise correlations, including self-correlations, of all input features to the NNs used for event classification are subject to

extensive scrutiny. This is done exploiting goodness-of-fit tests, based on a saturated likelihood model [78] including all systematic uncertainties of the model and their correlations. This guarantees a good understanding of the input space to the NNs and the input distributions used for the statistical inference of the signal.

5.5 Systematic uncertainties

The uncertainty model used for the analysis comprises theoretical uncertainties, experimental uncertainties, and uncertainties due to the limited population of template distributions available for the background model. The last group of uncertainties is incorporated for each bin of each corresponding template individually following the approach proposed in Ref. [79]. For this analysis, in which the signal significance is expected to be limited to a few bins with low background expectation, these uncertainties can often range among the leading sources of systematic uncertainty. All other uncertainties lead to correlated changes across bins either in form of normalization changes or as general non-trivial shape altering variations. Depending on the way they are derived correlations may also arise across years, samples, or individual uncertainties.

The following uncertainties related to the level of control of the reconstruction of electrons, muons, and τ_h decays after selection refer to simulated and τ-embedded events. Unless stated otherwise they are partially correlated across τ-embedded and simulated events.

- Uncertainties in the identification efficiency of electrons and muons amount to 2%, correlated across all years. Since no significant dependence on the p_T or η of each corresponding lepton is observed these uncertainties are introduced as normalization uncertainties.
- With a similar reasoning uncertainties in the electron and muon trigger efficiencies are also introduced as normalization uncertainties. They amount to 2%, each. Due to differences in the trigger leg definitions they are treated as uncorrelated for single lepton and pair triggers, which may result in shape altering effects in the overall model, since both triggers act on different regimes in lepton p_T.
- For fully simulated events an uncertainty in the electron energy scale is derived from the calibration of ECAL crystals, and applied on an event-by-event basis. For τ-embedded events uncertainties of 0.5–1.25%, split by the ECAL barrel and endcap regions, are derived for the corrections described in Section 5.4. Due to the different ways the uncertainties are determined and differences in detector conditions they are treated as uncorrelated across simulated and τ-embedded events. They lead to shape altering variations and are treated as correlated across years.
- Uncertainties in the τ_h-identification range between 3 to 5% in bins of τ_h p_T. Due to the nature of how they are derived these uncertainties are statistically dominated and therefore treated as uncorrelated across decay modes, p_T bins, and years. The same is true for the uncertainties in the τ_h-energy scale, which range from 0.2% to 1.1% depending on the p_T and the decay mode of the τ_h. For the energy scale of electrons misidentified as τ_h extra corrections are derived depending on the τ_h p_T and decay mode. Their uncertainties range from 1 to 2.5%. Concerning correlations the same statements apply as for the τ_h-energy scale. All uncertainties discussed in this paragraph lead to shape altering variations.
- Uncertainties in the τ_h-trigger efficiency are between 5 and 10% depending on the p_T of the τ_h. They are obtained from parametric fits to data and simulation, and lead to shape altering effects. They are treated as uncorrelated across triggers and years.

Two further sources of uncertainty are considered for τ-embedded events:

- A 4% normalization uncertainty accounts for the level of control in the efficiency of the $\mu\mu$ selection in data, which is unfolded during the τ-embedding procedure. The dominant part of this uncertainty originates from the trigger used for selection. Since these trigger setups differed across years this uncertainty is treated as uncorrelated across years.
- Another shape and normalization altering uncertainty in the normalization of $\mathrm{t\bar{t}} \to \mu\mu + \mathrm{X}$ decays, which are part of the τ-embedded event samples ranges between sub-percent level and 10% depending on the event composition of the model. For this uncertainty the number and shape of $\mathrm{t\bar{t}}$ events contained in the τ-embedded event samples are estimated from simulation, for which the corresponding decay has been selected at parton level. This estimate is then varied by $\pm 10\%$.

For fully simulated events the following additional uncertainties apply:

- Uncertainties in the $\mathrm{e}(\mu) \to \tau_\mathrm{h}$ misidentification rate amount to 40% for electrons and range from 10–70% for muons. These only apply to simulated $\mathrm{Z} \to \mathrm{ee}(\mu\mu)$ events, which are of marginal importance for the analysis. The same is true for an uncertainty in the reweighting in Z boson mass and p_T, discussed in Section 5.4, which ranges from 10–20%.
- Uncertainties in the energy calibration and resolution of jets are applied with different correlations depending on their sources, comprising statistical limitations of the measurements used for calibration, the time-dependence of the energy measurements in data due to detector aging, and non-closure corrections introduced to cover residual differences between simulation and data. They range between sub-percent level and $\mathcal{O}(10\%)$ depending on the kinematic properties of the jets in the event. Similar uncertainties are applied for the (mis)identification rates for (light quark or gluon induced jets) b jets, which are of a similar range.
- Depending on the process in consideration, two independent uncertainties on $p_\mathrm{T}^\mathrm{miss}$ are applied. For processes that are subject to recoil corrections, i.e., Z boson production, h boson production, or signal, uncertainties in the calibration and resolution of the hadronic recoil are applied, ranging from 1–5%. For all other processes an uncertainty in $p_\mathrm{T}^\mathrm{miss}$ is derived from the amount of unclustered energy in the event.
- A normalization uncertainty due to the timing shift of the inputs of the ECAL L1 trigger described in Section 5.4 amounts to 2–3%.
- A shape altering uncertainty is derived in the reweighting of the top quark p_T described in Section 5.4 by applying the correction twice or not applying it at all. This uncertainty is shape altering and has only a very small effect on the final discriminator.
- The luminosity is measured for each year of data taking individually [80–82]. These measurements are known to a precision of 2.5% for 2016 and 2018, and 2.3% for 2017. The corresponding normalization uncertainties comprise parts that are correlated and parts that are uncorrelated across the years.
- Uncertainties in the predictions of the normalizations of all simulated processes amount to 6% for $\mathrm{t\bar{t}}$ production, 5% for diboson and single t production, 2% for Z boson production, and 1.3–3.9% for the SM Higgs boson production rates used for h production, depending on the production mechanism. All these uncertainties are correlated across years.

For the F_F-method the following uncertainties apply:

- The F_F^i and their corrections are subject to statistical fluctuations in each corresponding DR^i. The corresponding uncertainties are split into a normalization and a shape altering part and propagated into the final discriminator. They are usually a few percent and are treated as uncorrelated across the kinematic and topological bins they are derived in.
- Additional uncertainties are applied to capture the needs and magnitudes of non-closure corrections and extrapolation factors, varying from a few percent to $\mathcal{O}(10\%)$, depending on the kinematic properties of the τ_h candidate and the topology of the event. These are both normalization and shape altering uncertainties.
- An additional source of uncertainty concerns the subtraction of processes other than the enriched process in each corresponding DR^i. These are subtracted from the data using simulated or τ-embedded events. The combined shape of the events to be removed is varied by 7%, and the measurements are repeated. The impact of these variations is then propagated to the final discriminator as shape altering uncertainties.
- An uncertainty in the estimation of the three main background fractions in the AR is estimated from a variation of each individual contribution by 7%, increasing or decreasing the remaining fractions such that the sum of all contributions remains unchanged. The effect of this variation is observed to be very small, since usually one of the contributions dominates the event composition in the AR.

Due to their mostly statistical nature and differences across years, all uncertainties related to the F_F-method are treated as uncorrelated across years.

6 Results

Sets of input distributions based on the NN classification for $m_H = 500\,\text{GeV}$ and $100 \leq m_{h_S} < 150\,\text{GeV}$ in the $\mu\tau_h$, $e\tau_h$, and $\tau_h\tau_h$ final states are shown in Figs. 2 – 4. For these Figures all years that have been analyzed have been combined. To retain the shape of the distributions of the y_i in each category the histogram bins have been divided by their widths in the upper panels of each figure. As a Bayesian probability estimate the values of y_i range from 0.2 to 1. The lower bound is given by the constraint that each event has to be associated to one of the five event categories. In each event category increasing purity of the associated process is expected. The points with error bars correspond to the data and the stacked filled histograms to the expectation from the background model. For the signal categories the expectation for a signal with $\sigma \times \mathcal{B}(H \to h(\tau\tau)h_S(bb)) = 200(50)\,\text{fb}$ depending on the $\tau\tau$ final state is also shown by a red line.

In the middle panels for all background categories the purity estimated for the background template of each corresponding event category is shown. For the signal categories the ratio of the indicated signal divided by the sum of all backgrounds is shown. In the lower panels of each figure the observed numbers of events divided by the numbers of events expected from the background model are shown for each bin.

No signal is observed in any of the investigated mass combinations and 95% confidence level (CL) upper limits on the $\sigma \times \mathcal{B}(H \to h(\tau\tau)h_S(bb))$ of a potential signal are set following the modified frequentist approach as described in Refs. [83, 84], using the same definition of the

profile likelihood test statistic as defined in Refs. [85, 86]:

$$q_\mu = -2\ln\left(\frac{\mathcal{L}(\{k_i\}|\,\mu\,S(\{\hat{\theta}_{j,\mu}\})+B(\{\hat{\theta}_{j,\mu}\}))}{\mathcal{L}(\{k_i\}|\,\hat{\mu}\,S(\{\hat{\theta}_{j,\hat{\mu}}\})+B(\{\hat{\theta}_{j,\hat{\mu}}\}))}\right),\quad 0\leq\hat{\mu}\leq\mu\,. \tag{7}$$

For this purpose all branching fractions in the signal sample have been set to 1. In Equation (7) $\hat{\mu}$, $\hat{\theta}_{j,\mu}$, and $\hat{\theta}_{j,\hat{\mu}}$ indicate the maximum likelihood estimates of the corresponding parameters from the fit to the data and the index of q_μ indicates that the fit to the data has been performed for a fixed value of μ. In the large number limit the distribution of q_μ can be approximated by analytic functions, from which the median and the uncertainty contours can be obtained as described in Ref. [87].

The observed and expected limits as a function of the tested values of $m_{\mathrm{h_S}}$ in a mass range from $240 \leq m_{\mathrm{H}} \leq 3000\,\mathrm{GeV}$ and for the combination of all $\tau\tau$ final states and the analyzed data from all years are shown in Fig. 5. The observed limits are given by the black points. The expected median values in the absence of signal are indicated by the dashed black line with the central 68 and 95% expected quantiles for the upper limit given by the green and yellow bands. The limits range from 125 fb for $m_{\mathrm{H}} = 240\,\mathrm{GeV}$ and $m_{\mathrm{h_S}} = 85\,\mathrm{GeV}$ to 2.7 fb for $m_{\mathrm{H}} = 1000\,\mathrm{GeV}$ and $m_{\mathrm{h_S}} = 350\,\mathrm{GeV}$ with a roughly flattening progression beyond. These limits are model independent. Since the analysis is not able to distinguish between scalar and pseudoscalar Higgs bosons the limits are equally applicable to both cases.

It should be noted that neighboring mass points differ only slightly in the kinematic properties of the tested signal hypotheses. Groups of hypothesis tests based on the same NN trainings for classification are indicated by discontinuities in the limits, which are linearly connected otherwise to improve the visibility of common trends.

Maximally allowed values for $\sigma \times \mathcal{B}(\mathrm{H} \to \mathrm{h}(\tau\tau)\mathrm{h_S}(\mathrm{bb}))$ in the context of the NMSSM are obtained from the NMSSMTOOLS 5.5.0 [88] and NMSSMCALC [89] programs and have been provided by the LHC Higgs Working Group. The mass region of the NMSSM accessible for constaint by this search is summarized in Fig. 6. The maximally allowed ranges for $\sigma \times \mathcal{B}(\mathrm{H} \to \mathrm{h}(\tau\tau)\mathrm{h_S}(\mathrm{bb}))$ within the NMSSM can be constrained by this search for masses between 400 and 600 GeV. The strongest constraint is achieved for $m(\mathrm{H}) = 450\,\mathrm{GeV}$ and $60 \leq m_{\mathrm{h_S}} \leq 80\,\mathrm{GeV}$, where the maximally allowed value for $\sigma \times \mathcal{B}(\mathrm{H} \to \mathrm{h}(\tau\tau)\mathrm{h_S}(\mathrm{bb}))$ is reduced by a factor of five by this search.

7 Summary

A search for the decay of a heavy Higgs boson H into the observed Higgs boson h and another Higgs boson $\mathrm{h_S}$ with a mass of $m_{\mathrm{h_S}} < m_{\mathrm{H}} - m_{\mathrm{h}}$ has been presented. The h and the $\mathrm{h_S}$ bosons are required to decay into a pair of tau leptons and a pair of b quarks, respectively. The search is based on $137\,\mathrm{fb}^{-1}$ of proton-proton collisions collected with the CMS detector during the LHC Run 2 data taking period at a center-of-mass energy of 13 TeV. A mass range of 240 – 3000 GeV for m_{H} and 60 – 2800 GeV for $m_{\mathrm{h_S}}$ is covered. No signal has been observed. Therefore model independent 95% confidence level upper limits on the product of the production cross section and the branching fractions of the searched process are set ranging from 125 fb (for $m_{\mathrm{H}} = 240\,\mathrm{GeV}$) to 2.7 fb (for $m_{\mathrm{H}} = 3000\,\mathrm{GeV}$). These limits have been compared to predictions of the next-to-minimal supersymmetric extension of the standard model.

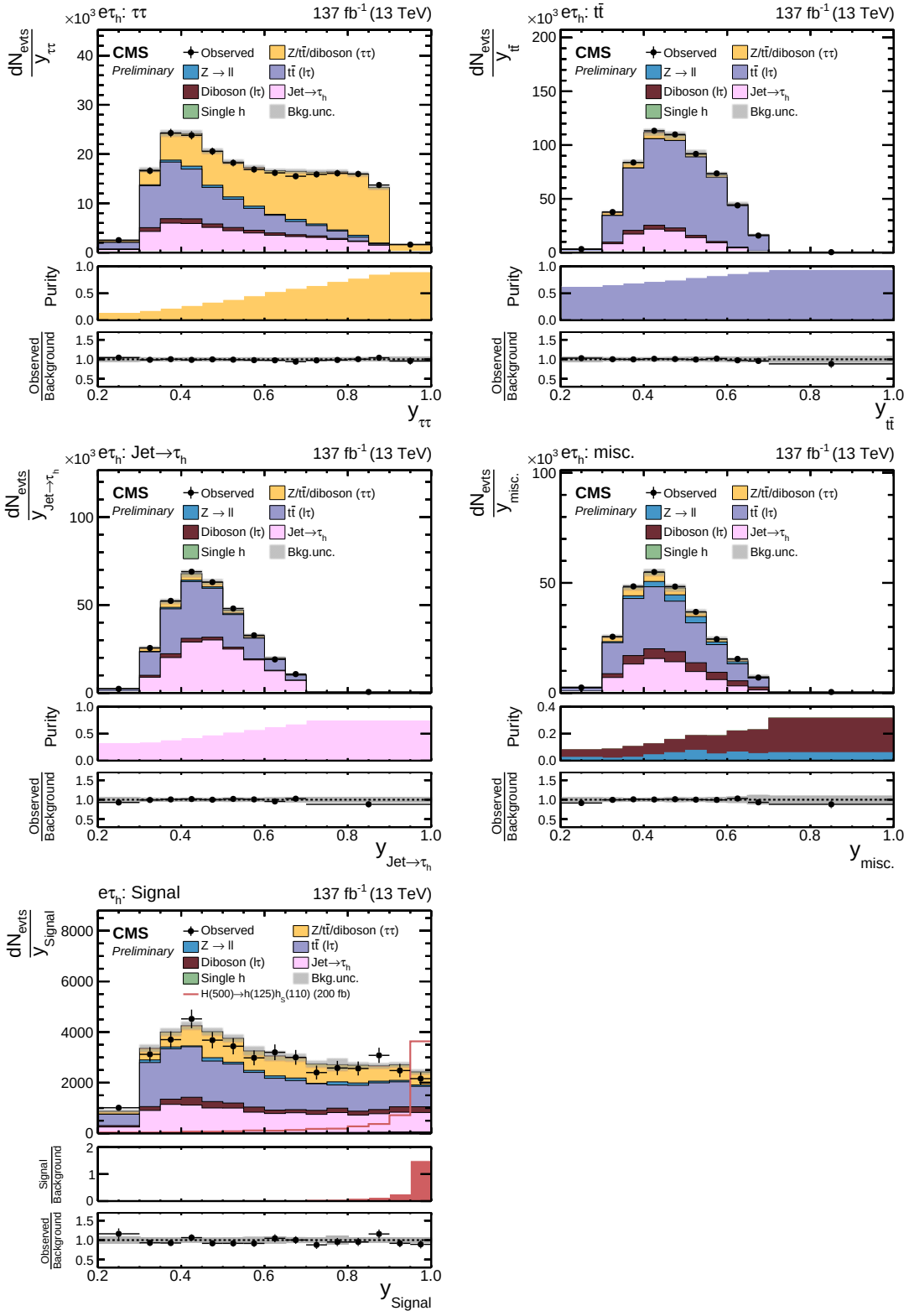

Figure 2: Event categories after NN classification based on a training for $m_{\mathrm{H}} = 500\,\mathrm{GeV}$ and $100 \leq m_{\mathrm{h_S}} < 150\,\mathrm{GeV}$ in the $\mathrm{e}\tau_{\mathrm{h}}$ final state. Shown are the (upper left) $\tau\tau$, (upper right) tt, (middle left) jet $\rightarrow \tau_{\mathrm{h}}$, (middle right) misc, and (lower left) signal categories. For these figures the datasets of all years have been combined. The uncertainty bands correspond to the combination of statistical and systematic uncertainties after the fit to the signal plus background hypothesis for $m_{\mathrm{H}} = 500\,\mathrm{GeV}$ and $m_{\mathrm{h_S}} = 110\,\mathrm{GeV}$.

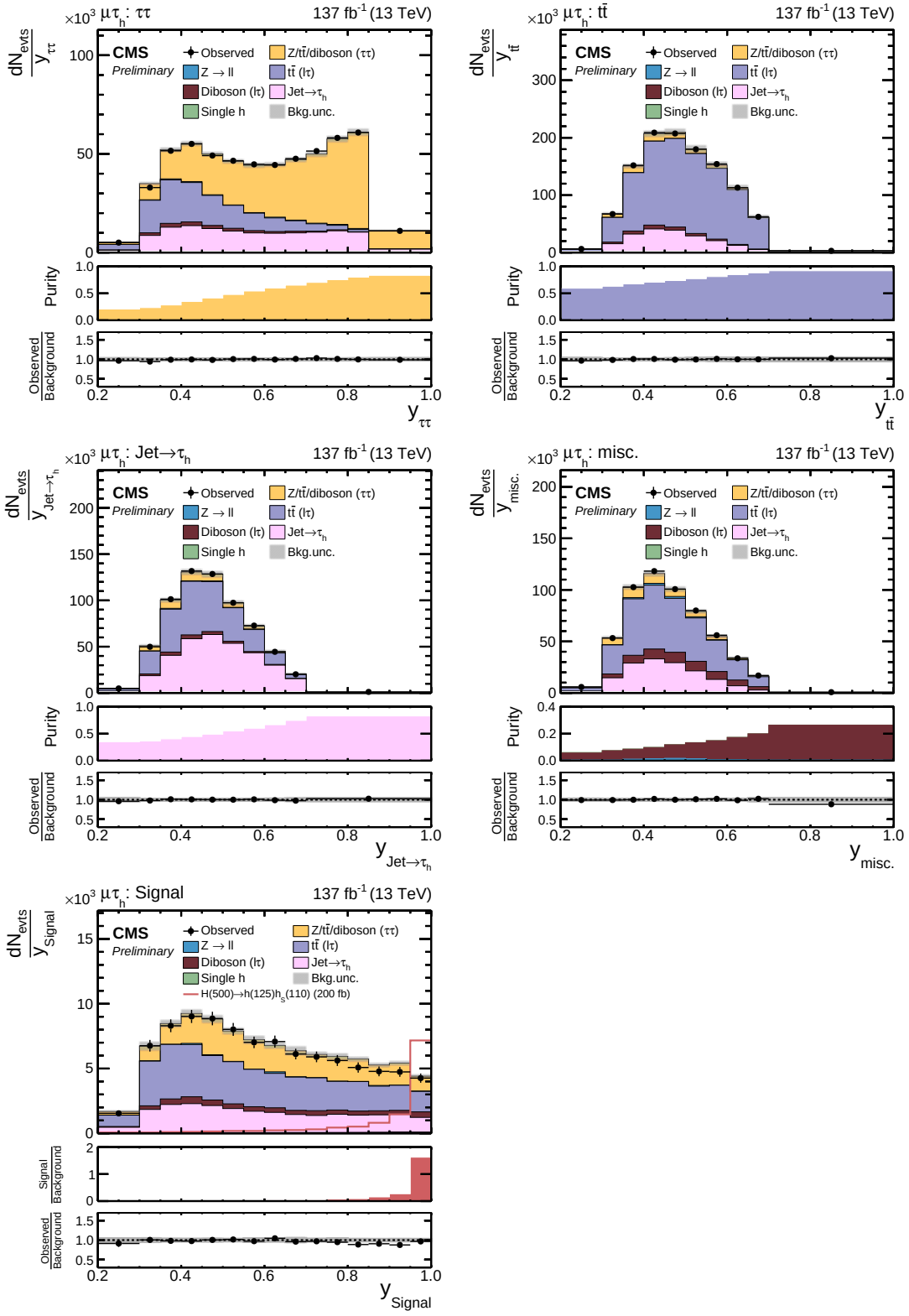

Figure 3: Event categories after NN classification based on a training for $m_{\mathrm{H}} = 500\,\mathrm{GeV}$ and $100 \leq m_{\mathrm{h_S}} < 150\,\mathrm{GeV}$ in the $\mu\tau_{\mathrm{h}}$ final state. Shown are the (upper left) $\tau\tau$, (upper right) tt, (middle left) jet $\rightarrow \tau_{\mathrm{h}}$, (middle right) misc, and (lower left) signal categories. For these figures the datasets of all years have been combined. The uncertainty bands correspond to the combination of statistical and systematic uncertainties after the fit to the signal plus background hypothesis for $m_{\mathrm{H}} = 500\,\mathrm{GeV}$ and $m_{\mathrm{h_S}} = 110\,\mathrm{GeV}$.

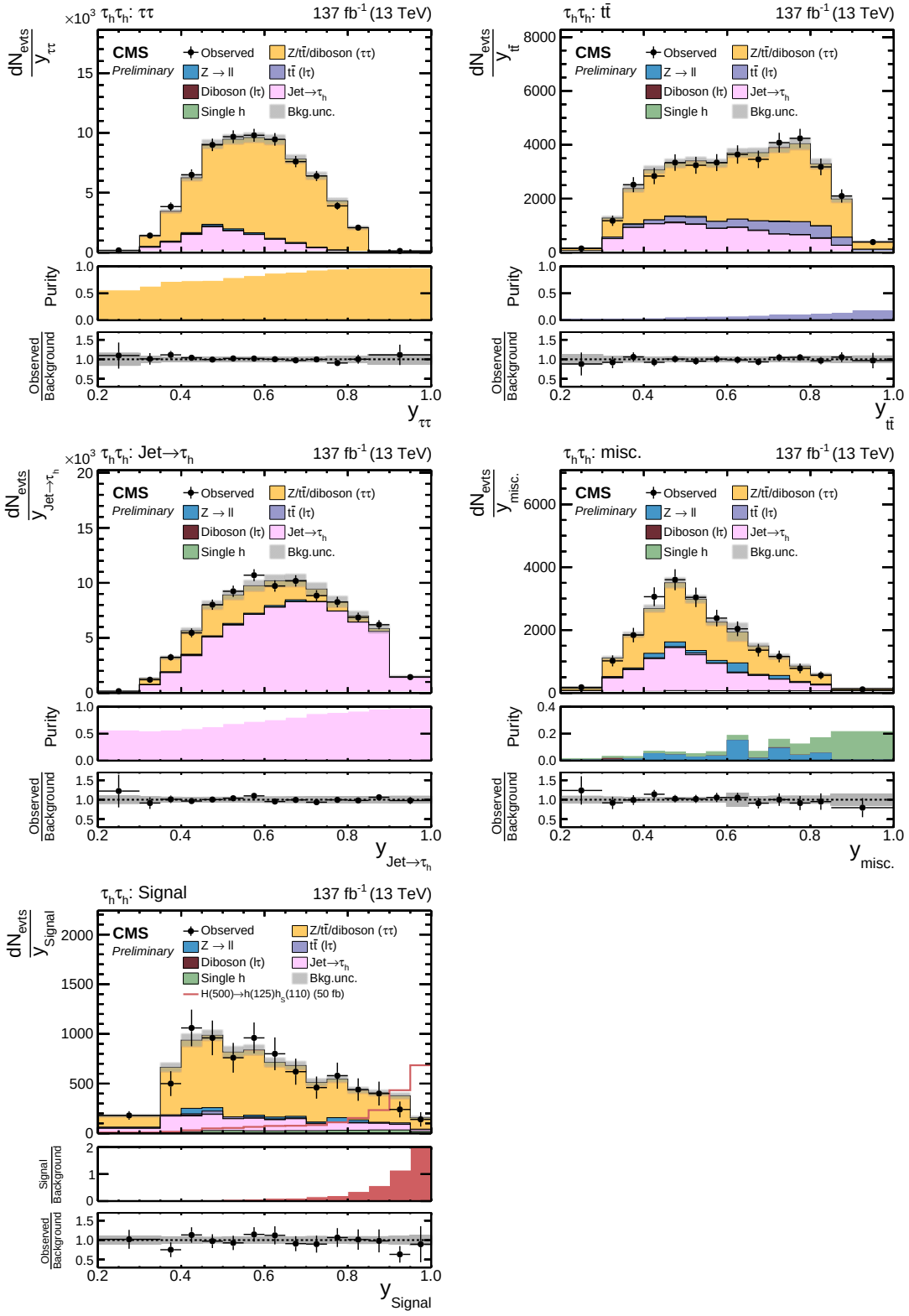

Figure 4: Event categories after NN classification based on a training for $m_{\mathrm{H}} = 500\,\mathrm{GeV}$ and $100 \leq m_{\mathrm{h_S}} < 150\,\mathrm{GeV}$ in the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state. Shown are the (upper left) $\tau\tau$, (upper right) tt, (middle left) jet $\to \tau_{\mathrm{h}}$, (middle right) misc, and (lower left) signal categories. For these figures the datasets of all years have been combined. The uncertainty bands correspond to the combination of statistical and systematic uncertainties after the fit to the signal plus background hypothesis for $m_{\mathrm{H}} = 500\,\mathrm{GeV}$ and $m_{\mathrm{h_S}} = 110\,\mathrm{GeV}$.

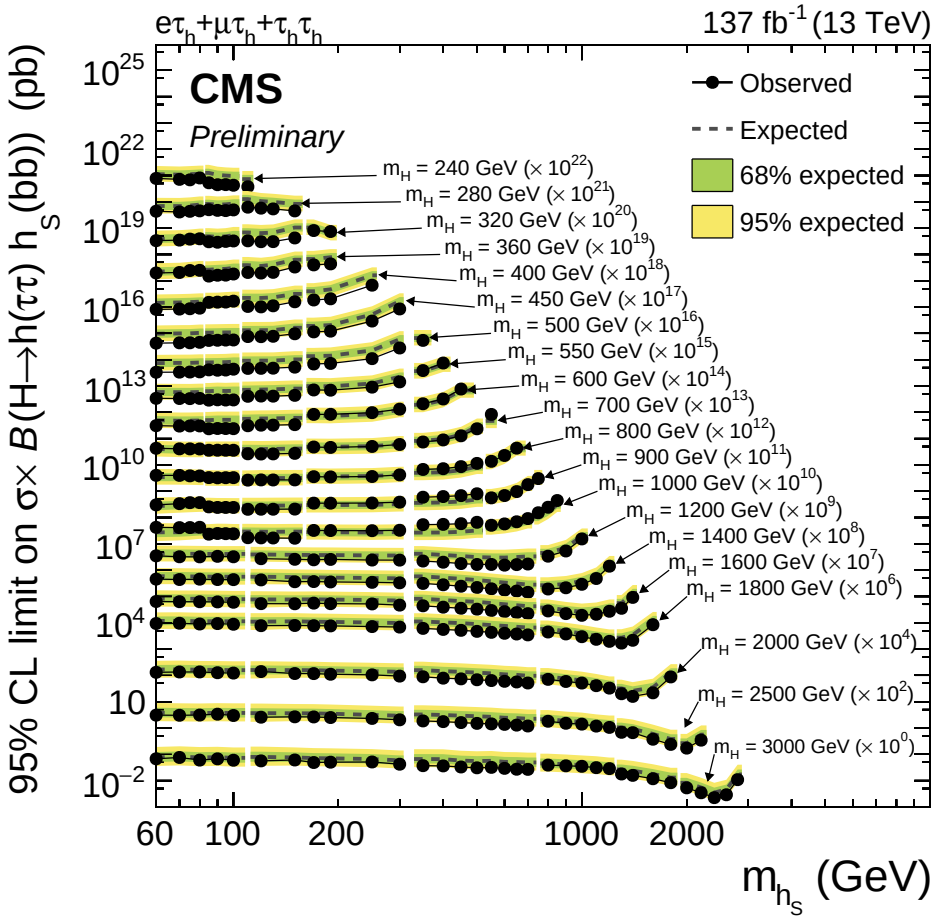

Figure 5: Expected and observed 95% CL upper limits on $\sigma \times \mathcal{B}(\mathrm{H} \rightarrow \mathrm{h}(\tau\tau)\mathrm{h_S}(\mathrm{bb}))$ for all tested m_{H} values. The limits for each corresponding mass value have been scaled by orders of ten as indicated in the annotations.

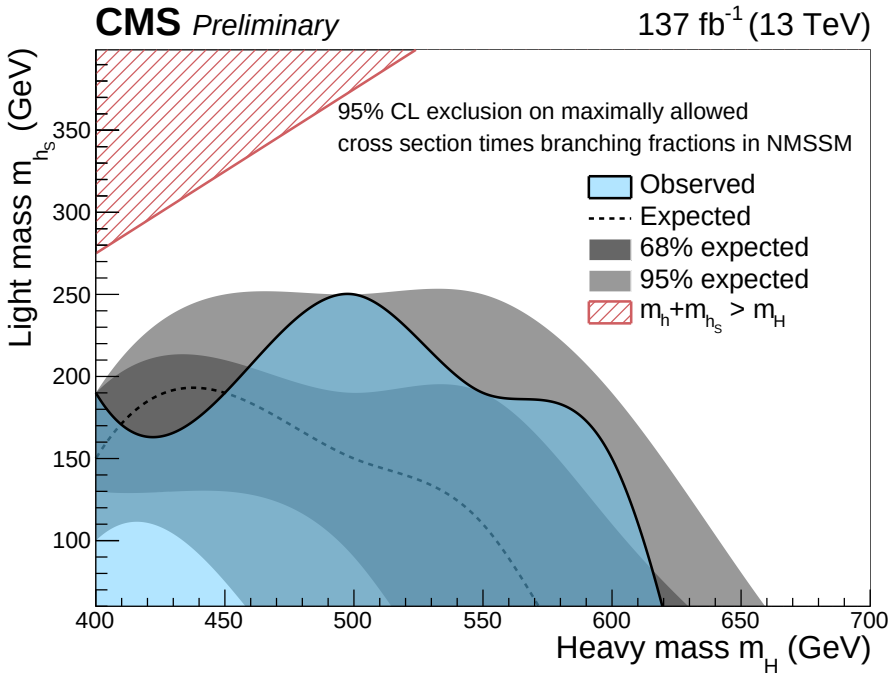

Figure 6: Mass range in m_{H} and $m_{\mathrm{h_S}}$ for which the maximally allowed $\sigma \times \mathcal{B}(\mathrm{H} \to \mathrm{h}(\tau\tau)\mathrm{h_S}(\mathrm{bb}))$ within the NMSSM can be excluded at 95% CL by this search.

References

[1] ATLAS Collaboration, "Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[3] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **06** (2013) 081, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.

[4] 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", *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.

[5] CMS Collaboration, "Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **79** (2019), no. 5, 421, doi:10.1140/epjc/s10052-019-6909-y, arXiv:1809.10733.

[6] ATLAS Collaboration, "Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment", *Phys. Rev. D* **101** (2020), no. 1, 012002, doi:10.1103/PhysRevD.101.012002, arXiv:1909.02845.

[7] CMS Collaboration, "Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state", *Phys. Rev. D* **99** (2019), no. 11, 112003, doi:10.1103/PhysRevD.99.112003, arXiv:1901.00174.

[8] Yu. A. Golfand and E. P. Likhtman, "Extension of the algebra of Poincaré group generators and violation of p invariance", *JETP Lett.* **13** (1971) 323.

[9] J. Wess and B. Zumino, "Supergauge transformations in four-dimensions", *Nucl. Phys. B* **70** (1974) 39, doi:10.1016/0550-3213(74)90355-1.

[10] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132, doi:10.1016/0031-9163(64)91136-9.

[11] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles", *Phys. Rev. Lett.* **13** (1964) 585, doi:10.1103/PhysRevLett.13.585.

[13] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.

[14] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons", *Phys. Rev.* **145** (1966) 1156, doi:10.1103/PhysRev.145.1156.

[15] T. W. B. Kibble, "Symmetry breaking in non-abelian gauge theories", *Phys. Rev.* **155** (1967) 1554, doi:10.1103/PhysRev.155.1554.

[16] P. Fayet, "Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino", *Nucl. Phys. B* **90** (1975) 104, doi:10.1016/0550-3213(75)90636-7.

[17] P. Fayet, "Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions", *Phys. Lett. B* **69** (1977) 489, doi:10.1016/0370-2693(77)90852-8.

[18] J. E. Kim and H. P. Nilles, "The mu problem and the strong CP problem", *Phys. Lett. B* **138** (1984) 150, doi:10.1016/0370-2693(84)91890-2.

[19] U. Ellwanger, C. Hugonie, and A. M. Teixeira, "The next-to-minimal supersymmetric standard model", *Phys. Rept.* **496** (2010) 1, doi:10.1016/j.physrep.2010.07.001, arXiv:0910.1785.

[20] M. Maniatis, "The next-to-minimal supersymmetric extension of the standard model reviewed", *Int. J. Mod. Phys. A* **25** (2010) 3505, doi:10.1142/S0217751X10049827, arXiv:0906.0777.

[21] ATLAS Collaboration, "Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.* **125** (2020), no. 5, 051801, doi:10.1103/PhysRevLett.125.051801, arXiv:2002.12223.

[22] CMS Collaboration, "Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **09** (2018) 007, doi:10.1007/JHEP09(2018)007, arXiv:1803.06553.

[23] ATLAS Collaboration, "Search for charged Higgs bosons decaying via $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ in the τ+jets and τ+lepton final states with 36 fb^{-1} of pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS experiment", *JHEP* **09** (2018) 139, doi:10.1007/JHEP09(2018)139, arXiv:1807.07915.

[24] CMS Collaboration, "Search for charged Higgs bosons in the $\mathrm{H}^{\pm} \to \tau^{\pm}\nu_{\tau}$ decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **07** (2019) 142, doi:10.1007/JHEP07(2019)142, arXiv:1903.04560.

[25] S. King, M. Muehlleitner, R. Nevzorov, and K. Walz, "Discovery prospects for NMSSM Higgs bosons at the high-energy large hadron collider", *Phys. Rev. D* **90** (2014), no. 9, 095014, doi:10.1103/PhysRevD.90.095014, arXiv:1408.1120.

[26] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[27] CMS Collaboration, "Track impact parameter resolution for the full pseudorapidity coverage in the 2017 dataset with the CMS Phase-1 pixel detector", CMS Detector Performance Note CMS-DP-2020-049, CERN, 2020.

[28] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[29] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[30] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV", *JINST* **10** (2015), no. 08, P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[31] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[32] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.

[33] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

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

[35] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[36] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[37] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", arXiv:2012.06888.

[38] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018), no. 06, P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[39] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[40] E. Bols et al., "Jet Flavour Classification Using DeepJet", *Submitted to: JINST* (2020) arXiv:2008.10519.

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

[42] CMS Collaboration, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018), no. 10, P10005, doi:10.1088/1748-0221/13/10/P10005, arXiv:1809.02816.

[43] CMS Collaboration, "Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons", CMS Detector Performance Note CMS-DP-2019-033, CERN, 2019.

[44] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup per particle identification", *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

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

[46] CMS Collaboration, "Evidence for the Higgs boson decay to a bottom quark–antiquark pair", *Phys. Lett. B* **780** (2018) 501, doi:10.1016/j.physletb.2018.02.050, arXiv:1709.07497.

[47] L. Bianchini, J. Conway, E. K. Friis, and C. Veelken, "Reconstruction of the Higgs mass in $H \to \tau\tau$ events by dynamical likelihood techniques", *J. Phys. Conf. Ser.* **513** (2014) 022035, doi:10.1088/1742-6596/513/2/022035.

[48] CMS Collaboration, "Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with $h \to \tau\tau$", *Phys. Lett. B* **755** (2016) 217, doi:10.1016/j.physletb.2016.01.056, arXiv:1510.01181.

[49] S. Wunsch, R. Friese, R. Wolf, and G. Quast, "Identifying the relevant dependencies of the neural network response on characteristics of the input space", *Comput. Softw. Big Sci.* **2** (2018), no. 1, 5, doi:10.1007/s41781-018-0012-1, arXiv:1803.08782.

[50] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, p. 249. 2010.

[51] R. Shimizu et al., "Balanced mini-batch training for imbalanced image data classification with neural network", in *2018 First International Conference on Artificial Intelligence for Industries (AI4I)*, p. 27. 2018.

[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", arXiv:1412.6980.

[53] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method", *Soviet Math. Dokl.* **4** (1963) 1035.

[54] CMS Collaboration, "An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data", *JINST* **14** (2019), no. 06, P06032, doi:10.1088/1748-0221/14/06/P06032, arXiv:1903.01216.

[55] 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", *Eur. Phys. J. C* **78** (2018), no. 9, 708, doi:10.1140/epjc/s10052-018-6146-9, arXiv:1801.03535.

[56] J. Alwall et al., "MadGraph 5: Going beyond", *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[57] 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", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[58] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[59] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[60] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO Higgs boson production via gluon fusion matched with shower in POWHEG", *JHEP* **04** (2009) 002, doi:10.1088/1126-6708/2009/04/002, arXiv:0812.0578.

[61] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[62] S. Alioli et al., "Jet pair production in POWHEG", *JHEP* **04** (2011) 081, doi:10.1007/JHEP04(2011)081, arXiv:1012.3380.

[63] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini, "Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM", *JHEP* **02** (2012) 088, doi:10.1007/JHEP02(2012)088, arXiv:1111.2854.

[64] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$", *Phys. Rev. D* **74** (2006) 114017, doi:10.1103/PhysRevD.74.114017, arXiv:hep-ph/0609070.

[65] M. Czakon and A. Mitov, "Top++: A program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[66] N. Kidonakis, "Top quark production", in *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013): JINR, Dubna, Russia, July 15-28, 2013*, p. 139. 2014. arXiv:1311.0283. doi:10.3204/DESY-PROC-2013-03/Kidonakis.

[67] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.

[68] T. Gehrmann et al., "W^+W^- production at hadron colliders in next to next to leading order QCD", *Phys. Rev. Lett.* **113** (2014), no. 21, 212001, doi:10.1103/PhysRevLett.113.212001, arXiv:1408.5243.

[69] NNPDF Collaboration, "Parton distributions for the LHC run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[70] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017), no. 10, 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

[71] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[72] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020), no. 1, 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[73] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A brief introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[74] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[75] S. Agostinelli et al., "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[76] CMS Collaboration, "Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **01** (2011) 080, doi:10.1007/JHEP01(2011)080, arXiv:1012.2466.

[77] CMS Collaboration, "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J. C* **75** (2015), no. 11, 542, doi:10.1140/epjc/s10052-015-3709-x, arXiv:1505.04480.

[78] S. Baker and R. D. Cousins, "Clarification of the use of chi square and likelihood functions in fits to histograms", *Nucl. Instrum. Meth.* **221** (1984) 437, doi:10.1016/0167-5087(84)90016-4.

[79] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219, doi:10.1016/0010-4655(93)90005-W.

[80] CMS Collaboration, "CMS luminosity measurements for the 2016 data taking period", Technical Report CMS-PAS-LUM-17-001, CERN, 2017.

[81] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-LUM-17-004, CERN, 2018.

[82] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-LUM-18-002, CERN, 2019.

[83] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[84] A. L. Read, "Presentation of search results: The CL_s technique", *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

[85] ATLAS and CMS Collaborations, "Procedure for the LHC Higgs boson search combination in summer 2011", Technical Report ATL-PHYS-PUB 2011-11, CMS NOTE 2011/005, CERN, 2011.

[86] CMS Collaboration, "Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett. B* **710** (2012) 26, doi:10.1016/j.physletb.2012.02.064, arXiv:1202.1488.

[87] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: doi:doi:10.1140/epjc/s10052-013-2501-z].

[88] U. Ellwanger and C. Hugonie, "NMHDECAY 2.0: An updated program for sparticle masses, Higgs masses, couplings and decay widths in the NMSSM", *Comput. Phys. Commun.* **175** (2006) 290, doi:10.1016/j.cpc.2006.04.004, arXiv:hep-ph/0508022.

[89] J. Baglio et al., "NMSSMCALC: A program package for the calculation of loop-corrected Higgs boson masses and decay widths in the (complex) NMSSM", *Comput. Phys. Commun.* **185** (2014), no. 12, 3372, doi:10.1016/j.cpc.2014.08.005, arXiv:1312.4788.