CMS Physics Analysis Summary

Contact: cms-pag-conveners-b2g@cern.ch 2021/11/17

Search for a massive scalar resonance decaying to a light scalar and a Higgs boson in the four b quark final state with boosted topology

The CMS Collaboration

Abstract

We search for the resonant production of a new massive scalar X decaying into a new light scalar Y and the standard model Higgs boson H through the process $X \to YH \to b\bar{b}b\bar{b}$. The search uses proton-proton collision data from the CERN LHC, collected at a centre-of-mass energy of 13 TeV in 2016-2018 and corresponding to an integrated luminosity of 138 fb^{-1}. The search is dedicated to mass ranges of X (0.9-4 TeV) and Y (60-600 GeV) where both the Y and the H are highly Lorentz-boosted and therefore their b quark-antiquark daughter particles are sufficiently collimated to be reconstructed using single large-area jets each. The mass of the one of the jets is required to be compatible with that of the Higgs boson, which is 125 GeV. A scan is performed in a two dimensional phase space spanned by the mass of the other jet, associated with Y, and the invariant mass of the two jets used to reconstruct X. The results are interpreted in the context of scalar resonances predicted in the next-to-minimal supersymmetric standard model and upper limits are placed on the production cross section as a function of the masses of X and Y. This is the first search for this process using Lorentz-boosted event topologies and it significantly extends the constraints on the model under consideration.

1 Introduction

The discovery of a Higgs boson of mass 125 GeV [1–3] at the CERN LHC has, on the one hand vindicated the predictions of the Brout-Englert-Higgs mechanism [4–9] of the standard model (SM), while on the other, raised the question of its validity at higher energy scales [10–13]. Notwithstanding the possibility of the SM being valid all the way up to the Planck scale, it is considered incomplete according to many theoretical and observational considerations. Beyond-standard models (BSMs) like supersymmetry (SUSY) [14] and extradimensions [15] seek to alleviate in part or in full the shortcomings of the SM.

Among the many BSMs, SUSY has been a favourite as an elegant solution to many of SM's shortcomings. Although no BSMs, including SUSY, have been observed at the LHC, the masses of their predicted particles may lie above the currently explored ranges. The unexplored regions of the SUSY parameter space include its scalar sector. The minimal supersymmetric extension of the SM (MSSM) [16, 17] postulates two complex scalar field doublets of the SU(2) gauge symmetry, giving rise to five Higgs bosons. The next-to-minimal model (NMSSM) [18, 19] has been proposed to solve the "μ-problem" of the MSSM [20]. It contains an extra complex scalar field S, an SU(2) singlet, totalling to seven Higgs bosons: three neutral scalars $\mathrm{H}^{'}$, H, $\mathrm{H_s}$; two neutral pseudoscalars A and $\mathrm{A}^{'}$, and two charged scalars H^{+} and H^{-}. If the NMSSM were true, one of the neutral scalars H would be associated with the 125 GeV Higgs boson. Searches for a heavier scalar $\mathrm{H}^{'}$ decaying to SM particles have set a limit on its mass at $M_{\mathrm{H}^{'}} > 2$ TeV [21–24].

An interesting prospect in the NMSSM is that the scalar $\mathrm{H_s}$, arising from the mixing of the doublet and singlet scalar fields, may predominantly contain S, which holds in most of the NMSSM parameter space. This will lead to a suppression of $\mathrm{H_s}$ coupling to the SM fermions leading to suppressed production cross section. It can then be produced mostly through the production and decay of $\mathrm{H}^{'}$ via $\mathrm{H}^{'} \to \mathrm{H_s H_s}$ or $\mathrm{H}^{'} \to \mathrm{H H_s}$. The latter can be the most dominant process. The maximum branching fractions of both H and $\mathrm{H_s}$ are to a pair of b quark-antiquarks giving the final state $\mathrm{H}^{'} \to \mathrm{H H_s} \to \mathrm{b\bar{b}b\bar{b}}$ (for Y with mass less than twice that of the top quark t) [25]. At one-tenth of this rate, the second dominant process is $\mathrm{H}^{'} \to \mathrm{H H_s} \to \tau\tau\mathrm{b\bar{b}}$, and has been searched for by the CMS Collaboration [26]. Models besides the NMSSM have also postulated such a process involving the production and decay of scalar particles [27]. We generically label the new scalars as X and Y, which in the case of NMSSM would correspond to $\mathrm{H}^{'}$ and $\mathrm{H_s}$, respectively.

This note describes the search for two new scalars, X and Y, with X being more massive and decaying through $\mathrm{X} \to \mathrm{YH}$, where H is the Higgs boson discovered at the LHC. The search uses LHC proton-proton (pp) collision data collected by the CMS experiment in 2016–2018 and amounting to an integrated luminosity of 138 fb^{-1} [28]. The masses of the three particles satisfy $M_{\mathrm{X}} > M_{\mathrm{Y}} + M_{\mathrm{H}}$; Y may be lighter or heavier than H and both decay through $\mathrm{b\bar{b}}$. This search focuses on the kinematic regions where $M_{\mathrm{X}} \gg M_{\mathrm{Y}}$ or M_{H} such that both Y and H are imparted considerable momentum and therefore their decay products, i.e. the $\mathrm{b\bar{b}}$ pairs are highly collimated. Correspondingly, we explore the mass ranges $0.9 < M_{\mathrm{X}} < 4$ TeV and $60 < M_{\mathrm{Y}} < 600$ GeV in this analysis. In this interesting kinematic regime, special techniques have been used to reconstruct the final states containing the collimated $\mathrm{b\bar{b}}$, in order to increase the signal sensitivity well beyond those covered by previous searches.

2 The CMS detector and event reconstruction

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 η coverage provided by the barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. 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. [29]. 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 [30]. 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 [31].

The candidate vertex with the largest value of summed physics-object p_{T}^2 is taken to be the primary pp interaction vertex. The physics objects are the jets and the associated missing transverse momentum, taken as the negative vector sum of the p_{T} of those jets. The jets are clustered using the infrared and collinear safe anti-k_{T} algorithm [32, 33] with the tracks assigned to candidate vertices as inputs, with a distance parameter of 0.4.

A particle-flow algorithm (PF) [34] aims to reconstruct and identify each individual particle in an event, with an optimized combination of information from the various elements of the CMS detector. The photon energy is obtained from the ECAL measurements. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies.

Jets are clustered from PF candidates using the anti-k_{T} algorithm with a distance parameter of either 0.4 (AK4 jets) or 0.8 (AK8 jets). Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5 to 10% of the true momentum over the whole p_{T} spectrum and detector acceptance. Additional pp interactions within the same or nearby bunch crossings (pileup) can contribute additional tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect, charged particles identified to be originating from pileup vertices are discarded and an offset correction is applied to correct for remaining contributions [35]. Jet energy corrections are derived from simulation to bring the measured response of jets to that of particle level jets on average.

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event, and its magnitude is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$ [36]. The $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is modified to account for corrections to the energy scale of the reconstructed jets in the event.

3 Signal and background processes

Monte Carlo simulations of the signal process $X \to YH \to b\bar{b}b\bar{b}$, with $0.9 < M_X < 4\,\text{TeV}$ and $60 < M_Y < 600\,\text{GeV}$ and a width of 1 MeV, are generated using the MADGRAPH5_aMC@NLO2.3.3 [37] event generator. The NNPDF3.1 next-to-leading order parton distribution functions (PDFs) [38], part of the LHAPDF6 PDF set [39], are used. The showering and hadronization of partons are simulated with PYTHIA8 [40].

The two main backgrounds are $t\bar{t}$+jets, where the top quarks decay hadronically, and events with jets arising purely from the SM quantum chromodynamics (QCD) interactions (multijet events). The former is modelled using POWHEG 2.0 [41–43] and PYTHIA8, with the CP5 tune, with corrections derived from collision data using events requiring the presence of either an electron or a muon. The multijet events background is estimated using data events containing only jets. A sample of $t\bar{t}$ with one of the top quarks decaying via $t \to Wb \to \ell\nu b$, ℓ being a lepton (electron or a muon) is also simulated using the same event generators and tune. Both the samples are scaled using a cross section of 832^{+46}_{-52} pb, calculated at next-to-next-to-leading order in QCD [44].

The simulation of the events in the CMS detector is done using GEANT4 [45]. The effects of additional pp interactions in the same or adjacent bunch crossings as the main collision event (pileup) is modelled, assuming a total inelastic pp cross section of 69.2 mb [46]. All simulated event samples are weighted to match the distribution of the number of pileup events observed in the data.

4 Event selection

The events were selected in two mutually-exclusive categories: "hadronic", containing only jets and "semileptonic", containing one lepton and which serves as a control sample to calibrate the simulated $t\bar{t}$+jets yields using data. Collision events in the hadronic category were collected using the logical OR of several HLT algorithms.

One trigger criterion required a single AK8 jet with a p_T >450 and >500 GeV in 2016 and in 2017–2018, respectively. To complement this, a second trigger required that the scalar sum p_T of all AK4 jets, H_T be >800 or 900 GeV in 2016, depending on the LHC beam luminosity. In 2017 and 2018, H_T >1050 GeV was required.

The AK8 jet trimmed mass is obtained after removing remnants of soft radiation with the jet trimming technique [47], using a subjet size parameter of 0.3 and a subjet-to-AK8 jet p_T fraction of 0.1. The third set of trigger algorithms requiring an AK8 jet with p_T >360 GeV along with a trimmed mass >30 GeV was used in 2016. In 2017–2018, the AK8 jet p_T threshold was raised to 400 and 420 GeV, keeping the trimmed mass criterion the same, depending on the LHC beam luminosity.

In 2016, a fourth trigger requiring two AK8 jets, one with p_T >280 and 200 GeV with one having a trimmed mass of 30 GeV was used. The same requirements, along with a jet passing a loose b-tagging criterion using the "combined secondary vertex" algorithm [48] constituted the fifth trigger. In 2016, a sixth trigger required $H_T \geq 650$ GeV with a pair of AK4 jets having invariant mass above 900 GeV and a pseudorapidity separation $|\Delta\eta| < 1.5$.

The final set of triggers combined the H_T and the trimmed mass requirements. In 2016, such triggers required H_T >650 or 700 GeV with an AK8 jet having trimmed mass >50 GeV. In 2017 and 2018, the H_T thresholds were increased to >800 GeV, with the trimmed mass threshold

unchanged.

The triggers in the semileptonic category require events to have either an isolated muon of p_T >24 or 27 GeV; or an isolated electron having p_T >27, 32, or 35 GeV; or a photon with energy greater than 175 or 200 GeV.

The hadronic event category trigger efficiency was measured in the data requiring a single AK4 jet with $p_T > 260$ GeV and was found to be between 92–100%. Simulated events were weighted by this efficiency as a function of the invariant mass of the two leading-p_T AK8 jets in the event, M_{JJ}. The semileptonic event trigger efficiencies were measured in a sample of $Z\to\ell\ell$ events and was found to be close to 100%.

Events are required to have two AK8 jets with p_T >350 (450) GeV and $|\eta| < 2.4(2.5)$ for 2016 (2017–2018) data or simulations. A requirement on their separation in pseudorapidity is imposed to suppress the multijet background, $|\Delta\eta| < 1.3$. The soft-drop algorithm [49, 50], is used to remove wide-angled soft radiation from a jet in order to mitigate the effects of contamination from initial state radiation, underlying event, and multiple hadron scattering. The algorithm is used with radiation fraction parameter z=0.1, which determines the minimum p_T fraction that the jet constituents need to have in order not to be removed, and the angular exponent parameter β=0, which scales the p_T fraction threshold as a function of the distance between the jet constituents. With β=0, there is no dependence of the threshold on the distance between the constituents. The effect of pileup on the soft-drop algorithm is mitigated using the pileup per particle identification (PUPPI) algorithm [51].

An $H\to b\bar{b}$ candidate or an "H jet" is that whose soft-drop mass is $110 < M_J^H < 140$ GeV. The second jet is designated as the $Y\to b\bar{b}$ candidate or the "Y jet" if its soft-drop mass satisfies $M_J^Y > 60$ GeV. When both AK8 jets satisfy the H jet soft-drop mass requirement, the Y jet is chosen at random. Events without an H or a Y jet are not selected for further analysis.

A graph convolutional neural network based algorithm, ParticleNet [52] is used to identify the boosted $H\to b\bar{b}$ or $Y\to b\bar{b}$ decays against a background of other jets. It estimates the probability of a jet to arise from a certain decay, such as from $H\to b\bar{b}$ or $Y\to b\bar{b}$ ($P(Y/H\to b\bar{b})$) or from a light-flavoured quark or a gluon ($P(QCD)$). The ParticleNet algorithm can also identify jets from the decay of a boosted top quark. The probability of all possible origins of a jet sum to one. In this analysis the ParticleNet score of a jet is defined as $P(Y/H\to b\bar{b})/(Y/H\to b\bar{b} + P(QCD)$. The algorithm is trained on AK8 jets using simulated Lorentz-boosted spin-0 particles (X) with a flat mass spectrum (15-250 GeV), decaying to a pair of quarks as the signal. The QCD multijet samples are used for the background [53].

The background jet rejection power of the ParticleNet tagger is increased by at least a factor of two when compared to the DeepAK8-MD algorithm, used in a similar search for the resonant di-Higgs production [54], in the p_T range from 500 to 1000 GeV for the same signal efficiency [55] [53]. The improvement in the performance comes from the novel approach in considering a jet as an unordered set of its constituent particles and using a graph neural network to better exploit the correlation between the particles. It also benefits from a new mass-decorrelation technique. Since this search targets events with two AK8 jets, this translates to approximately a factor of four improvement in background rejection.

The ParticleNet scores used for selecting the $H\to b\bar{b}$ and the $Y\to b\bar{b}$ candidates ("signal jets") are either >0.98 (tight requirement) or >0.94 (loose requirement). Depending on the jet p_T, the former has an efficiency of 62–72% and a misidentification rate of 0.45%, while the latter has an efficiency of 80–85% and a misidentification rate of 1%. The efficiency of the ParticleNet

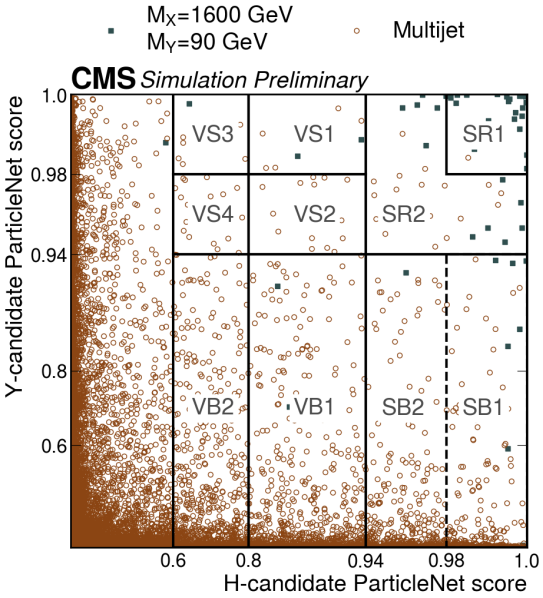

Figure 1: The distributions of the H and the Y candidate jets' ParticleNet scores for the signal with M_X = 1600 GeV and M_Y = 90 GeV (filled squares) and multijets background (open circles). The grid lines show the different event categories defined using the ParticleNet scores of the two jets. A description of the regions is given in Table 1 and in the text.

classifier is calibrated in data using a sample of jets originating from gluon splitting to $b\bar{b}$. A boosted decision tree is used to select such jets with properties close to jets from boosted $Y/H\to b\bar{b}$.

The ParticleNet scores of the two leading-p_T jets are used to classify events into either signal, sideband or validation categories. Signal regions require a high ($>$0.94 or $>$0.98) ParticleNet score for both the jets, indicating a higher probability of originating from the decay of a Higgs or a Y boson, than from a background process. The sideband event categories have jets with lower ParticleNet score and are hence enriched in light flavoured jets. They are deployed in the estimation of the multijets background using a method based on the rate of light flavoured jets being misidentified as a signal jet. The validation event categories serve to cross-check the background estimation method without using the signal region events and to provide an initial measurement of the misidentification rate of light flavoured jets for initiating the background estimation method.

A layout of the different signal, sideband, and validation regions using the two leading-p_T jet ParticleNet scores is shown in Fig. 1, with their descriptions given in Table 1. Referring to the numbered regions in Fig. 1, the two signal regions are: the "signal region 1" (SR1) and the "signal region 2" (SR2), which are mutually exclusive for the purpose of statistical combination. They are defined using the tight and loose ParticleNet requirements on the jets. The SR1 has a higher signal-to-background ratio and is thus more sensitive to the presence of signal. However, the SR2 improves the sensitivity for signal mass points with low background by increasing the signal efficiency.

Corresponding to the two signal regions, two "sideband regions" are defined, for the purpose of estimating the multijet background. The sidebands are formed of events whose Y jet candidates fail the loose ParticleNet tagger threshold and are almost entirely composed of multijet events. They are labelled as "Sideband 1" (SB1) and "Sideband 2" (SB2) in Fig. 1. The SB1 over-

laps with SB2 region in order to provide better sideband region characteristics for estimating the multijets background in their respective signal regions.

Table 1: Definition of the signal, sideband, and validation regions used for background estimation. The regions are defined in terms of the ParticleNet discriminators of the H and Y candidate jets, as shown in Fig. 1.

Region name and label in Fig. 1	ParticleNet discriminator		Purpose
	H Jet	Y jet	
Signal region 1 (SR1)	>0.98	>0.98	Signal
Signal region 2 (SR2) (excludes SR1)	>0.94	>0.94	Signal
Sideband 1 (SB1)	>0.98	<0.94	Sideband
Sideband 2 (SB2)	>0.94	<0.94	Sideband
Validation signal-like 1 (VS1)		>0.98	Validation
Validation signal-like 2 (VS2)	0.8–0.94	>0.94	Validation
Validation sideband 1 (VB1)		<0.94	Validation
Validation signal-like 3 (VS3)		>0.98	Validation
Validation signal-like 4 (VS4)	0.6–0.8	>0.94	Validation
Validation sideband 2 (VB2)		<0.94	Validation

Furthermore, six " validation regions", which are signal-free, are used to verify the efficacy of the background estimation method before the actual background estimation in the signal regions. They are defined and are grouped into threes based on their H and Y jet ParticleNet scores. The first set of regions (labelled VS1, VS2, VB1 in Fig. 1) require the H jet ParticleNet between 0.8–0.94. The H jet ParticleNet score for the second set of regions, labelled VS3, VS4, VB2 in Fig. 1, lies between 0.6 and 0.8. All these regions are enriched in QCD multijet events with almost no signals.

The other main background, $\mathrm{t\bar{t}}$+jets, is estimated from simulations. A sample of semileptonic $\mathrm{t\bar{t}}$+jets events (one of the top quarks decaying to $\mathrm{Wb}\to\ell\nu\mathrm{b}$) is used to derive data-to-simulation correction factors. Such events are required to have a lepton with p_T >40 GeV and $|\eta| < 2.4$. An AK4 jet, tagged as originating from a b quark (b-tagged) using the DeepJet algorithm [56], is required to be close to it, $\Delta R(\text{lepton, jet}) < 1.5$, where $\Delta R_{1,2} \equiv \sqrt{(\eta_1-\eta_2)^2+(\varphi_1-\varphi_2)^2}$ is the distance between two objects in the plane of pseudorapidity and azimuthal angle ϕ. The DeepJet score distributions of the AK4 jets are corrected using a weight extracted from measurements in the data [48]. The lepton and the b-tagged jets are a signature of the leptonic decay of a top quark. The requirements of $E_\mathrm{T}^\mathrm{miss} > 60\,\mathrm{GeV}$ and $H_\mathrm{T} > 500\,\mathrm{GeV}$ are also imposed. A hadronically decaying top quark candidate is reconstructed from an AK8 jet with $p_\mathrm{T} > 350(450)$ GeV and $|\eta| < 2.4(2.5)$ for 2016 (2017, 2018), a soft-drop mass >60 GeV, and satisfying $\Delta R(\text{lepton, AK8 jet}) > 2$. Events in the semileptonic category are split into two regions based on whether the AK8 jet passes the tight or loose ParticleNet scores and two separate correction factors are derived, one each for SR1 and SR2.

The total signal selection efficiencies, for M_X in 0.9–4 TeV and M_Y in 60–600 GeV, range from 1.7% to 12.6% in the SR1 category and from 1.3% to 5.6% in the SR2 category. The analysis uses the hadronic event category to search for a narrow signal in the 2-dimensional plane

spanned by M_{JJ} and $M_{\mathrm{J}}^{\mathrm{Y}}$. Most of the event distributions on this plane come from multijet and $\mathrm{t\bar{t}}$ events. The multijet event M_{JJ}–$M_{\mathrm{J}}^{\mathrm{Y}}$ distribution and yields are estimated using a pass-to-fail ratio method using the sideband regions. The nominal $\mathrm{t\bar{t}}$ event yields and distributions, taken from simulations, are corrected by fitting the top quark jet mass $M_{\mathrm{J}}^{\mathrm{t}}$ in the semileptonic region to the data. A p_{T}-dependent jet mass scale was found to have negligible effect based on validation region studies and was not considered.

5 Background estimation

The two-dimensional background event distributions as a function of M_{JJ} and the $M_{\mathrm{J}}^{\mathrm{Y}}$ are estimated for the signal regions SR1 and SR2. The multijet background, combined for the three data-taking years, is obtained starting from the the sideband region (SB1 or SB2) distributions and multiplying them by appropriate transfer functions. The transfer functions depend on both M_{JJ} and $M_{\mathrm{J}}^{\mathrm{Y}}$ and are defined as the ratios of events passing the ParticleNet classifier to the events failing the classifier ($R_{P/F}$). It is a priori unknown, without examining the signal region data.

An initial estimate ($R_{P/F}^{init}$) of the $R_{P/F}$ is made using the validation regions. It is multiplied with a correction function, R_{ratio} (described below), to account for the differences between the true $R_{P/F}$ and the initial estimate, giving the relation $R_{P/F} = R_{P/F}^{init} \times R_{ratio}$.

The ratios of the number of events in the binned $M_{\mathrm{J}}^{\mathrm{Y}}$ distribution in the regions VS1 to VB1 (VS2 to VB1), are fit using quadratic functions to give $R_{P/F}^{init}$ for the SR1 (SR2) signal regions background estimates. The $R_{P/F}^{init}$ is thus only a function of $M_{\mathrm{J}}^{\mathrm{Y}}$ (and is constant along M_{JJ}), unlike the final $R_{P/F}$ or the R_{ratio}. Using a 1-dimensional modelling reduces the statistical uncertainties in the estimated initial transfer function.

The R_{ratio} is modelled as a product of two polynomials in $M_{\mathrm{J}}^{\mathrm{Y}}$ and M_{JJ}. The parameters of R_{ratio} are determined in situ from the simultaneous fit of the background and the signal models to the data in the signal and the sideband regions. The order of the polynomial is chosen using a Fisher's F-test [57]. The F-test was performed starting from the base R_{ratio} of order one as a function of both M_{JJ} and $M_{\mathrm{J}}^{\mathrm{Y}}$ and adding terms in one of the two variables until there is no significant fit improvement with the additional terms. At 95% confidence level, the most favoured form for R_{ratio} was found to be the product of linear functions along both M_{JJ} and in $M_{\mathrm{J}}^{\mathrm{Y}}$, compared to higher order polynomials. The value of the correction in the two signal regions ranges from 0.4 to 2.9.

The $\mathrm{t\bar{t}}$+jets background is estimated from simulations with corrections measured in a region separate from the ones used to estimate multijet background. The data in the semileptonic region is used to constrain the event yields of the $\mathrm{t\bar{t}}$+jets background as well as the exact mass scale of the AK8 jets from the top quark decay. These jets fall into three categories, depending on the top quark boost. A high enough boost may result in a fully merged $\mathrm{t} \to \mathrm{Wb} \to \mathrm{q\bar{q}'b}$ decay, labelled as a bqq jet. At moderate boosts, the $\mathrm{W} \to \mathrm{q\bar{q}'}$ may be merged to form a W jet with the b quark forming its own jet. However, such events are nearly all eliminated in the event selection. Finally, one of the quarks from the W boson decay can merge with the b quark, which are labelled as bq jets. Unmerged and other combinatorial backgrounds constitute a small fraction outside these two categories. In the fit, the bqq jet and the bq components are scaled independently with the scales of each of these components tied to the corresponding backgrounds in the hadronic signal regions. They are also independent across the years, thus

giving six scales in total per signal region. The jet mass scales of bqq and bq are treated as decoupled nuisance parameters in the fit for each individual year.

The background estimation method was cross-checked using only the validation regions. Here, the goal was to predict the background in regions VS1 and VS2 using the region VB1 as sideband. The initial transfer functions $R^{init}_{P/F}$ were estimated from the ratios of events in the regions VS3 to VB2 and VS4 to VB2. The semileptonic $\mathrm{t\bar{t}}$+jets region was treated as it would be for the true background estimation in the signal regions. Similar to the actual background estimation process, a Fisher's test was used to decide the polynomial order of the R_{ratio} function. A goodness-of-fit test confirmed a good agreement between data and estimated background.

6 Systematic uncertainties

We consider several sources of systematic uncertainty that cover experimental effects, uncertainties due to the extraction of the multijet background, and uncertainties in the predicted $\mathrm{t\bar{t}}$ background.

The main uncertainties on the signal are:

- *ParticleNet scale factor*: The uncertainty in the correction applied in the simulation to match the efficiency of the ParticleNet discriminator in data. The uncertainty is 7–37%, depending on the AK8 jet p_{T}, and affects the signal by 15%.
- *Jet mass scale*: The jet mass scale uncertainty is modelled as a 5% shift in the AK8 jet soft-drop mass. It is decorrelated between the bqq jets, the bq jets, and the signal jets. Besides changing the shape of distributions, it also affects the yields because of the Higgs mass condition applied to one of the jets. It impacts the signal by 13%. The JMS uncertainty on the $\mathrm{t\bar{t}}$+jets background is reduced by including the semileptonic control region.
- *Jet energy scale and resolution*: The uncertainties are applied to both AK4 and AK8 jets, and are fully correlated between the two sets of jets. The signal is impacted by 5%.
- *Jet mass resolution*: The nominal jet mass resolution from simulation is taken as the downward uncertainty and a 20% resolution smear is applied as the upward variation to the AK8 jets, resulting in 4% impact on the signal.

The following uncertainties mostly affect the background determination:

- *$t\bar{t}$ normalization*: The uncertainties in the parameters used to describe the data-based correction for the $\mathrm{t\bar{t}}$+jets background range from 6% to 16%.
- *Top quark p_{T} modelling*: An uncertainty is associated with the top quark p_{T} modelling in Monte Carlo simulations of the $\mathrm{t\bar{t}}$ process [58], resulting in a 2% uncertainty coming from the $\mathrm{t\bar{t}}$+jets background.
- *Pass-to-fail ratio uncertainty*: The main source of uncertainty for the multijet background comes from the uncertainty in the $R^{init}_{P/F}(M^{\mathrm{Y}}_{\mathrm{J}})$ fit and is proportional to the statistical uncertainty because of the sample size in the VS1, VS2 and VB1 regions. Its impact of 7–11% is evaluated by calculating the yield change when applying the $R^{init}_{P/F}(M^{\mathrm{Y}}_{\mathrm{J}})$ to the failing regions with uncertainties.

Other considered sources of systematic uncertainties with minor impact are as follows:

- *Trigger*: The uncertainty in the trigger efficiencies, ϵ, are relevant for only $M_{\mathrm{JJ}} < 1100\,\mathrm{GeV}$ where it drops below 100% and is taken as $(1-\epsilon)/2$. The uncertainties go

up to 3%.

- *Trigger timing correction*: During the 2016 and 2017 data taking, a gradual shift in the timing of the inputs of the ECAL hardware level trigger in the region of $|\eta| > 2.0$ caused a specific trigger inefficiency. To take this effect into account, a 2% normalization uncertainty is applied to $\mathrm{t\bar{t}}$ and signal.
- *Integrated luminosity*: The uncertainty on the total Run 2 (2016–2018) integrated luminosity is 1.6%.
- *Pileup*: The value of the pp total inelastic cross section that is used in the simulation of pileup events is varied upwards and downwards from its assumed value of 69.2 mb by its uncertainty of 4.6% [46].
- *PDF and scale uncertainties*: The impact of the PDF and the QCD factorization and renormalization scale uncertainties on the signal acceptance and selection is estimated to be 1.0%. The former is derived using the PDF4LHC procedure [59] and the NNPDF3.1 PDF sets. The later is evaluated by changing the renormalization (μ_R) and factorization (μ_F) scales in simulation by factors of 0.5 and 2.
- *Sample size of sideband regions*: The statistical uncertainties associated with the number of events in the sideband regions SB1 and SB2 impact the estimated multijets background in the SR1 and SR2 signal regions. These uncertainties are small compared with the uncertainty in the $R^{init}_{P/F}(M_{\mathrm{J}}^{\mathrm{Y}})$, are uncorrelated between the bins of the histograms used to model these distributions, and are included in the Likelihood using the Barlow–Beeston Lite prescription [60, 61].
- *Lepton ID and isolation*: The lepton identification and isolation efficiency data-to-simulation correction factors have an uncertainty which affect the event yields by 1–2% in the semileptonic selection.
- *b-tagging scale factor uncertainty*: The uncertainty on the correction applied in the simulation to match the shape of the DeepJet discriminator in data of 2% is applied and affects the $\mathrm{t\bar{t}}$ in the semileptonic region.

All uncertainties are decorrelated between years, except the PDF, pileup, luminosity and top quark p_{T} modelling uncertainties.

7 Results

The joint likelihood of the M_{JJ}–$M_{\mathrm{J}}^{\mathrm{Y}}$ distributions in the hadronic (signal and sideband) regions in both the SR1 and the SR2 categories along with the $M_{\mathrm{J}}^{\mathrm{t}}$ distributions in the semileptonic regions in the tight and loose categories are constructed. The data for the three years are added together. The multijets background for the three years are combined in the likelihood while separate estimates are made for the $\mathrm{t\bar{t}}$+jets and the signals. The nuisance parameters in the likelihood correspond to various measurement and statistical uncertainties. The searched mass range for X lies between 900 GeV and 4 TeV and for Y between 60 and 600 GeV. The distributions of the data, the postfit backgrounds, and three representative signals are shown in Fig. 2. The estimated background is in good agreement with the observed data and the exclusion limits on the signal cross sections were calculated.

Upper limits are computed with a modified frequentist approach, using the $\mathrm{CL_s}$ criterion [62, 63] with the profile likelihood ratio used as the test-statistic with the asymptotic approximation. Expected and observed limits as a function of M_{X} and M_{Y} are shown in Fig. 3. The current analysis improves the limits by a factor of two over a similar search for a resonance

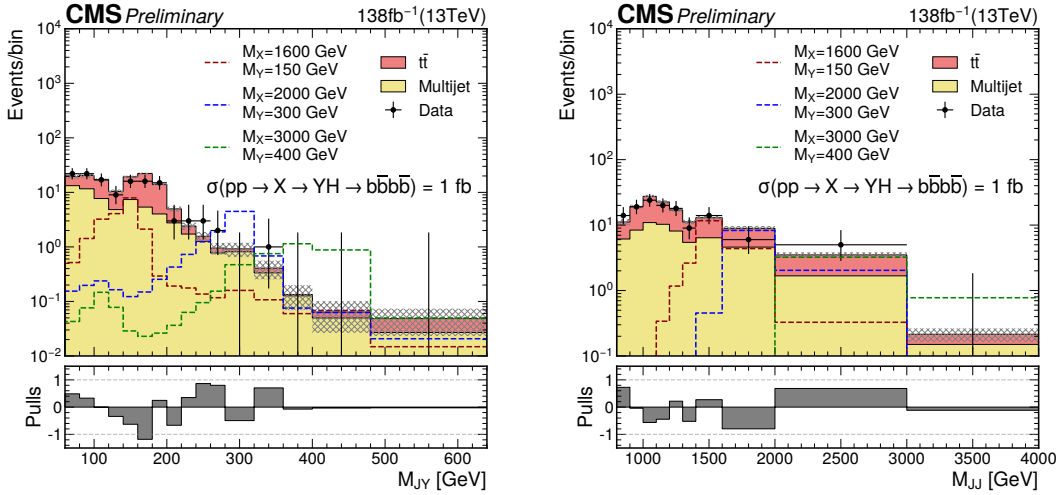

Figure 2: The M_J^Y (left) and M_{JJ} (right) distributions for the number of observed events (black markers) compared with the estimated backgrounds (filled histograms) in the signal region 1. The distributions expected from the signal under three M_X and M_Y hypotheses and assuming a cross section of 1 fb are also shown. The lower panels show the "Pulls" defined as $(\text{observed events} - \text{expected events})/\sqrt{\sigma_{obs}^2 + \sigma_{exp}^2}$, where σ_{obs} and σ_{exp} are the statistical and total uncertainties in the observation and the background estimation, respectively.

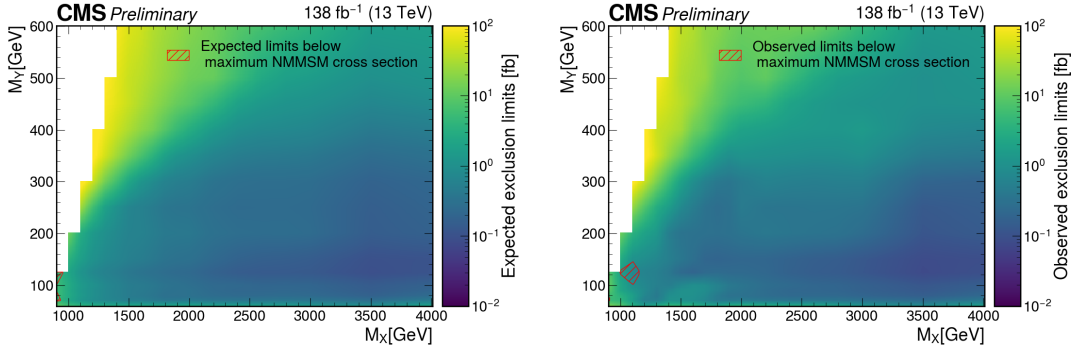

Figure 3: The 95% confidence level expected (left) and observed (right) upper limits on $\sigma(pp \to X \to YH \to b\bar{b}b\bar{b})$ for different values of M_X and M_Y.

decaying to two 125 GeV Higgs bosons [54]. The improvement is driven by the improved background rejection of the ParticleNet algorithm over the previous DeepAK8-MD boosted $H \to b\bar{b}$ identification algorithm [55].

8 Summary

A search for massive scalar resonances X and Y, where X decays to the lighter Y and the standard model Higgs boson H, is conducted in LHC proton-proton collision data collected by the CMS detector between 2016 and 2018 and corresponding to an integrated luminosity of 138 fb^{-1}. The considered decay modes of both Y and H are to a b quark-antiquark pairs each. Events are selected assuming very high Lorentz-boost of both the Y and the H, for which spe-

cialized jet substructure and identification techniques are used. The background, composed of multijets and $t\bar{t}$, is estimated using data control samples and Monte Carlo simulations. A binned likelihood fit to the data is performed using the reconstructed mass distributions of the X and Y candidates. Upper limits are set on the cross section of the process $X \to YH \to b\bar{b}b\bar{b}$ for assumed masses of X in the range 0.9–4 TeV and Y between 60–600 GeV. The results are interpreted in the context of the next-to-minimal supersymmetric standard model (NMSSM) scalar sector. This search significantly extends limit ranges of NMSSM scalars over previous analyses and places the most stringent cross section limits over much of the explored X and Y mass ranges to date.

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

[5] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[6] 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.

[7] 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.

[8] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons", *Phys. Rev.* **145** (1966) 1156, doi:10.1103/PhysRev.145.1156.

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

[10] D. Buttazzo et al., "Investigating the near-criticality of the Higgs boson", *JHEP* **12** (2013) 089, doi:10.1007/JHEP12(2013)089, arXiv:1307.3536.

[11] Y. Hamada, H. Kawai, and K.-y. Oda, "Bare Higgs mass at Planck scale", *Phys. Rev. D* **87** (2013), no. 5, 053009, doi:10.1103/PhysRevD.87.053009, arXiv:1210.2538. [Erratum: Phys.Rev.D 89, 059901 (2014)].

[12] D. R. T. Jones, "Comment on "Bare Higgs mass at Planck scale"", *Phys. Rev. D* **88** (2013), no. 9, 098301, doi:10.1103/PhysRevD.88.098301, arXiv:1309.7335.

[13] A. V. Bednyakov, B. A. Kniehl, A. F. Pikelner, and O. L. Veretin, "Stability of the Electroweak Vacuum: Gauge Independence and Advanced Precision", *Phys. Rev. Lett.* **115** (2015), no. 20, 201802, doi:10.1103/PhysRevLett.115.201802, arXiv:1507.08833.

[14] J. Wess and B. Zumino, "Supergauge Transformations in Four-Dimensions", *Nucl. Phys. B* **70** (1974) 390, doi:10.1016/0550-3213(74)90355-1.

[15] L. Randall and R. Sundrum, "A large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.* **83** (1999) 3370, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.

[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] 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.

[19] 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.

[20] 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.

[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] U. Ellwanger and M. Rodriguez-Vazquez, "Simultaneous search for extra light and heavy Higgs bosons via cascade decays", *JHEP* **11** (2017) 008, doi:10.1007/JHEP11(2017)008, arXiv:1707.08522.

[26] CMS Collaboration, "Search for a heavy Higgs boson decaying into two lighter Higgs bosons in the $\tau\tau$bb final state at 13 TeV", 2021. arXiv:2106.10361. Submitted to JHEP.

[27] T. Robens, T. Stefaniak, and J. Wittbrodt, "Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios", *Eur. Phys. J. C* **80** (2020) 151, doi:10.1140/epjc/s10052-020-7655-x, arXiv:1908.08554.

[28] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS", *Eur. Phys. J. C* **81** (Apr, 2021) 800, doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927.

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

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

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

[32] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[33] 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.

[34] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[35] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *JINST* **15** (2020) P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503.

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

[37] 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.

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

[39] A. Buckley et al., "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J. C* **75** (2015) 132, doi:10.1140/epjc/s10052-015-3318-8, arXiv:1412.7420.

[40] 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.

[41] S. Frixione, G. Ridolfi, and P. Nason, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.

[42] 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.

[43] 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.

[44] 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.

[45] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.

[46] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **07** (2018) 161, doi:10.1007/JHEP07(2018)161, arXiv:1802.02613.

[47] D. Krohn, J. Thaler, and L.-T. Wang, "Jet trimming", *JHEP* **02** (2010) 084, doi:10.1007/JHEP02(2010)084, arXiv:0912.1342.

[48] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[49] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, "Towards an understanding of jet substructure", *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.

[50] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft drop", *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.

[51] 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.

[52] H. Qu and L. Gouskos, "ParticleNet: Jet Tagging via Particle Clouds", *Phys. Rev. D* **101** (2020), no. 5, 056019, doi:10.1103/PhysRevD.101.056019, arXiv:1902.08570.

[53] CMS Collaboration, "Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques", CMS Detector Performance Note CMS-DP-2020-002, 2020.

[54] CMS Collaboration, "Search for resonant Higgs boson pair production in four b quark final state using large-area jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-B2G-20-004, CERN, Geneva, 2021.

[55] CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", *JINST* **15** (2020), no. 06, P06005, doi:10.1088/1748-0221/15/06/P06005, arXiv:2004.08262.

[56] E. Bols et al., "Jet Flavour Classification Using DeepJet", *JINST* **15** (2020) P12012, doi:10.1088/1748-0221/15/12/P12012, arXiv:2008.10519.

[57] R. G. Lomax and D. L. Hahs-Vaughn, "Statistical concepts: a second course". Taylor and Francis, Hoboken, NJ, 2012.

[58] CMS Collaboration, "Measurement of the top quark mass with lepton+jets final states using pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **78** (2018) 891, doi:10.1140/epjc/s10052-018-6332-9, arXiv:1805.01428.

[59] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II", *J. Phys. G* **43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.

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

[61] J. S. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra", in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*, p. 115. 2011. arXiv:1103.0354. doi:10.5170/CERN-2011-006.115.

[62] 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.

[63] 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.