CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2017/05/29

Inclusive search for the standard model Higgs boson produced in pp collisions at $\sqrt{s} = 13$ TeV using $H\to b\bar{b}$ decays

The CMS Collaboration

Abstract

A search for the standard model Higgs boson produced with high transverse momentum decaying to a bottom quark-antiquark pair has been performed using a data set of pp collisions at $\sqrt{s} = 13$ TeV collected with the CMS experiment at the LHC. The data sample corresponds to an integrated luminosity of 35.9 fb^{-1}. A high-transverse-momentum Higgs boson decaying to $b\bar{b}$ is reconstructed as a single jet and identified based on jet substructure and dedicated b-tagging techniques. The analysis strategy is validated with $Z\to b\bar{b}$ decays. The $Z\to b\bar{b}$ process is observed with a local significance of 5.1 standard deviations for the first time in the single jet topology. For a Higgs boson mass of 125 GeV, an excess of events is observed above the expected background with a local significance of 1.5 standard deviations. The measured cross section of $H(b\bar{b})$ production for $p_T > 450$ GeV is 74^{+51}_{-49} fb.

1 Introduction

In the standard model (SM) [1–3], the Brout-Englert-Higgs mechanism [4–6] is responsible for electroweak symmetry breaking and the mass of of all elementary particles. Although the Higgs boson has been discovered [7, 8], its observed properties and couplings are only measured with a precision at the level of 10% or worse [9]. In particular, the LHC Run 1 data was not sufficient to establish the coupling of the Higgs boson to bottom quarks, despite the dominant branching ratio of a SM Higgs boson (with a mass of 125.1 GeV) to a bottom quark-antiquark ($\mathrm{b\bar{b}}$) pair (58.1% [10]).

The traditional strategy to search for $\mathrm{H} \to \mathrm{b\bar{b}}$ decays at a hadron collider is to use events in which the Higgs boson is produced in association with a W or Z boson decaying leptonically, and recoiling with a large transverse momentum [11], in order to suppress overwhelming irreducible background from QCD multijet production of b quarks.

Recent theoretical and experimental developments [12] propose to use the production of a high-p_{T} Higgs boson in association with a high-p_{T} jet to resolve the long- and short-distance contributions to the gluon fusion process. Additionally, the boosted $\mathrm{H} \to \mathrm{b\bar{b}}$ channel provides an alternative approach to study the top Yukawa coupling beside the $\mathrm{t\bar{t}H}$ process.

In this note, the results of the first inclusive search for the SM Higgs boson with $\mathrm{H} \to \mathrm{b\bar{b}}$ decays are reported, using the data set of pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ collected by the CMS detector at the LHC in 2016, and corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. The main experimental difficulties for this search originate from the large cross section of background multijet events at low dijet mass. In this analysis, a high-transverse-momentum jet from initial-state radiation (ISR) produced in association with the Higgs boson, provides enough energy in the event to satisfy on-line requirements. Combinatorial background is reduced by requiring the resonance's decay products to be clustered in a single jet. We then search for a resonance in the jet mass distribution. The jet is required to have two-prong substructure and b-tagging consistent with the $\mathrm{H} \to \mathrm{b\bar{b}}$ signal. The dominant background from SM QCD multijet production are estimated by simultaneously fitting events that pass and fail the b-tagging requirement. The b-tagging algorithm is, by design, decorrelated from jet mass and p_{T}. This is one of the key features of this analysis, since it allows the use of the signal-depleted control region to predict the shape of the QCD background in the signal region.

2 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 superconducting 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 [13] 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. [13].

3 Simulated samples

Simulated samples of signal and background events are produced using various Monte Carlo (MC) event generators, with the CMS detector response modeled with using GEANT4 [14]. The Higgs boson signal samples are produced using the POWHEG+MiNLO [15, 16] event generator with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. The aMC@NLO [17] generator is used for the diboson, W+jets, Z+jets samples at leading order (LO) accuracy with MLM matching [18], while POWHEG is used to model the $\mathrm{t\bar{t}}$ and single-top processes [19–21]. For parton showering and hadronization the POWHEG and MADGRAPH samples are interfaced with PYTHIA8 [22, 23]. The PYTHIA8 parameters for the underlying event description are set to the CUETP8M1 tune [24] based on the Monash one [25]. The production cross sections for the diboson samples are rescaled to next-to-next-leading-order (NNLO) accuracy with the MCFM generator [26]. Top quark pair production is rescaled to the cross section computed with Top++ v2.0 [27] at NNLO. The cross sections for the W+jets and Z+jets samples are rescaled to include higher order QCD and electroweak (EWK) corrections and improve modeling of the high-p_{T} W and Z events [28–31]. The parton distribution functions (PDF) set used to produce the next-to-leading-order (NLO) samples is the NLO NNPDF3.0 set [32], while the LO NNPDF3.0 one is used for the LO samples.

Computing the differential cross section in Higgs p_{T} for the gluon fusion production mode and Higgs $p_{\mathrm{T}} > 450\,\mathrm{GeV}$ poses a number of challenges. At low Higgs p_{T}, the dominant contributions come from the application of higher order corrections which are large for loop-induced processes. The dominant correction at values of the Higgs p_{T} greater than approximately twice the mass of the top quark originates from the resolved top quark loop (finite top mass correction) [33]. The resolved top quark loop induces a deficit in the production of Higgs bosons at high p_{T} relative to the case where the loop is unresolved, known as the effective field theory (EFT) or $m_{\mathrm{t}} \to \infty$ approximation.

In the interest of comparing with other CMS results, the POWHEG generator with Higgs matrix elements up to 1 jet is used and tuned with the *h-fact* parameter set to 104.13 GeV in an attempt to match the generated spectrum from HRes [34–36]. The resulting tuned Higgs generation is normalized to the inclusive next-to-next-to-next-to-leading order ($\mathrm{N^3LO}$) accuracy. In addition, an alternative approach is considered to get the highest order possible differential Higgs p_{T} spectrum [37–40], while preserving the finite top mass correction [41]. To account for both the effects of higher order corrections and the resolved top loop, a multi-correction approach is adopted [42, 43], that can be summarized as

$$\begin{aligned}\mathrm{GF\,H(NNLO + m_t)} = \mathrm{Powheg(1\,jet}\ m_{\mathrm{t}} \to \infty) \times \frac{\mathrm{MG\,LO}\ 0-2\,\mathrm{jet}\ m_{\mathrm{t}}}{\mathrm{Powheg(1\,jet}\ m_{\mathrm{t}} \to \infty)} \times \\ \times \frac{\mathrm{NLO\ 1\,jet}\ m_{\mathrm{t}}}{\mathrm{LO\ 1\,jet}\ m_{\mathrm{t}}} \times \frac{\mathrm{NNLO\ 1\,jet}\ m_{\mathrm{t}} \to \infty}{\mathrm{NLO\ 1\,jet}\ m_{\mathrm{t}} \to \infty}\,. \end{aligned} \tag{1}$$

Samples are generated at LO for the 0, 1, and 2 jet Higgs production with MADGRAPH (MG) using the $\mathrm{loop_{SM}}$ model [44], with a jet threshold of 20 GeV and then showered with the CKKW-L scheme [10, 42, 45]. The higher order corrections to the loop incorporating a finite top mass are obtained by scaling the $0-2$ jet LO finite top mass sample by the approximate NLO/LO correction [41]. The latter correction is the approximate NLO finite top mass correction obtained by expanding the EFT in powers of $1/m_{\mathrm{t}}^2$ ($\mathrm{NLO^*}$) and it is found to be 2.0 ± 0.5 and roughly constant as a function of p_{T}. In the infinite top mass approximation, the NNLO to NLO correction is found to be 1.25 ± 0.15 and is also roughly constant across p_{T} [37, 38, 46, 47].

For Higgs $p_{\mathrm{T}} > 450\,\mathrm{GeV}$, the correction to the default POWHEG is found to be 1.27 ± 0.38, resulting in a cross section of 31.7 ± 9.5 for $\mathrm{H} \to \mathrm{b\bar{b}}$ in the gluon fusion production mode. An

uncertainty of 30% is assigned from the comparison of several different predictions at high p_T including: (a) aMC@NLO EFT, NLO merged with FxFx scheme for $0-2$ jets and corrected for the finite top mass corrections with approximate NLO [10, 41] and normalized to the inclusive N^{3}LO; (b) the LO finite top mass 2 jet ME showered with the generic LO showering scheme in PYTHIA. The generation scale is 20 GeV on the jet p_T with an additional 30 GeV minimun p_T requirement on the second jet relative to the first jet, with NLO and NNLO+1 jet scale factors applied. This is found to give a higher prediction; (c) Same as the one described in Eqn. 1 but with a larger merging scale of 100 GeV. The uncertainty associated with the modeling of the GF Higgs p_T spectrum is propagated to the overall normalization of the GF Higgs signal. In addition, the shape of the GF Higgs p_T distribution is allowed to vary depending on the Higgs p_T by up to 30% at 1000 GeV, without changing the overall normalization.

The p_T spectrum of the Higgs boson for the vector boson fusion (VBF) production mode is reweighted to account for N^{3}LO corrections to the cross section. The size of these corrections is within a few percent at high p_T [48] and they have negligible effect on the yield for this process for events with Higgs $p_T > 450$ GeV.

4 Event reconstruction and selection

The particle-flow event algorithm [49] is employed to reconstruct and identify each individual particle with an optimized combination of information from the various elements of the CMS detector. The algorithm identifies each reconstructed particle either as an electron, a muon, a photon, a charged hadron, or a neutral hadron. The missing transverse momentum vector is defined as the projection of the negative vectorial sum of the momenta of all particle-flow candidates identified in the event, and its magnitude is referred to as E_T^{miss}.

The particle-flow candidates are clustered into jets using the anti-k_T algorithm [50] with a distance parameter $R = 0.8$ (AK8 jets). To mitigate the effect of pileup, the pileup per particle identification (PUPPI) algorithm [51] is used to weight the particle-flow candidates prior to jet clustering based on the likelihood of originating from the primary vertex. Further corrections are applied to the jet energy as a function of jet pseudorapidity (η) and p_T to account for detector non-linearities.

To isolate the Higgs boson signal and overcome trigger restrictions, a high-p_T signal jet is required, which typically recoils against another high-p_T ISR jet. Combinations of several online selections are used, all requiring the total hadronic transverse energy in the event (H_T) or jet p_T to be above a given threshold. Extra requirements on the jet mass after removing remnants of soft radiation with the jet trimming technique [52] are added to reduce the H_T or p_T thresholds in order to improve signal acceptance. The online selection is fully efficient at selecting events offline with at least one AK8 jet with $p_T > 450$ GeV and $|\eta| < 2.5$.

Events containing identified and isolated electrons, muons, or taus with $p_T > 10$ GeV and $|\eta| < 2.5(2.4, 2.3)$ are vetoed to reduce backgrounds from SM electroweak processes. Since no real E_T^{miss} is expected for signal events, events with $E_T^{miss} > 140$ GeV are removed in order to further reduce top background contamination from $\bar{t}t$.

The decay products of a high-p_T $H \to b\bar{b}$ system are reconstructed as one AK8 jet. The highest p_T jet in the event is assumed to be the Higgs boson candidate, the H jet. The soft drop algorithm [53, 54] is used to remove soft and wide-angle radiation with a soft radiation fraction z greater than 0.1 and angular exponent parameter of $\beta = 0$. The soft drop jet mass m_{SD} peaks at the Higgs boson mass for signal events and reduces the masses of jets from background

quark- and gluon-initiated processes. Dedicated mass corrections [55] are derived from simulation and data in a region enriched with $t\bar{t}$ events with merged $W(q\bar{q})$ decays. They remove the residual dependence on the jet p_T and match the jet mass scale and resolution observed in data. In this note, the scaling variable for QCD jets ρ, defined as $\rho = \log(m_{SD}^2/p_T^2)$ [53] is used in the characterization of the correlation of jet substructure variables with the jet mass and p_T. Only events in the ρ range $-6.0 < \rho < -2.1$ are considered, to avoid instabilities at the edges of the distribution due to finite cone effects from the AK8 jet clustering (around $\rho \sim -2$), and to avoid the non-perturbative regime of the soft drop mass calculation (below $\rho \sim -6$). This requirement is fully efficient for the Higgs boson signal. The distributions of m_{SD} and ρ for data and simulation are shown in Fig. 1 for the full data sample, after the requirement on the jet p_T and the online selection.

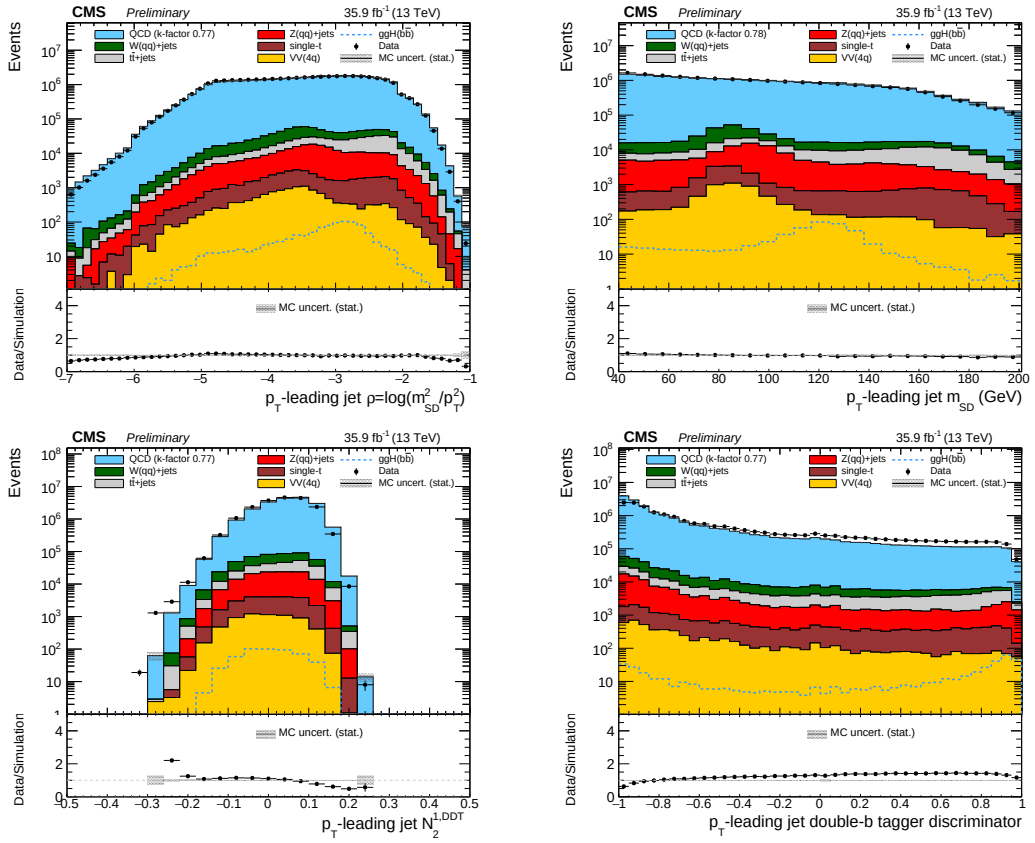

Figure 1: Data to simulation comparison of the ρ (top-left), m_{SD} (top-right), $N_2^{1,DDT}$ (bottom-left) and double-b tagger (bottom-right) variables, after the requirement on the jet p_T and the online selection.

The N_2^1 variable [56], which is based on a ratio of 2-point and 3-point generalized energy correlation functions (ECFs) [57], is exploited to determine how consistent a jet is with having a two-prong substructure. N_2^1 is computed from the jet constituents after applying the soft drop grooming to the jet. For a two-prong structure, signal jets have a stronger 2-point correlation than a 3-point correlation.

The N_2^1 observable provides excellent performance in discriminating two-prong signal jets from QCD background jets [56]. However, N_2^1 and many other similar variables [58] are correlated with the jet mass and p_T. The decorrelation procedure uses simulated QCD events and defines $N_2^{1,DDT}$ to be $N_2^1 - N_{2(26\%)}^1$, where $N_{2(26\%)}^1$ is the 26% quantile of the N_2^1 distribution as function of ρ and p_T. This ensures that the selection $N_2^{1,DDT} < 0$ yields a constant QCD background

efficiency of 26% across the whole ρ and p_T range considered in this search. The distribution is shown for data and simulated events in Fig. 1.

In order to select events for which the H jet is most likely to contain two b quarks, the double-b tagger algorithm [59] is used. The double-b tagger aims to fully exploit the presence of two b quarks inside an AK8 jet and their topology in relation to the jet substructure, namely the fact that the b hadron flight directions are strongly correlated with the energy flows of the two subjets. Several observables exploiting the distinctive properties of b hadrons are used as input variables to a multivariate algorithm, to distinguish between the H jet and the background from QCD jets. Events are selected by requiring the double-b tag discriminator value be greater than 0.9, which corresponds to about 1% efficiency for QCD jets and 33% for $H \to b\bar{b}$ signal. This value has been optimized to maximize the sensitivity of the search. In Fig. 1, the double-b tagger distribution is shown for data and simulated events.

Events are categorized depending on whether the jet has a double-b tag discriminator value greater than 0.9 (passing region) or not (failing region). In Fig. 2, the m_{SD} distribution is shown for simulated signal events in both the passing and failing regions. Besides the dominant gluon fusion process, other production mechanisms contribute to the SM Higgs boson signal yield after the event selection. All the Higgs boson production mechanism contributions are taken into account when extracting the signal contribution.

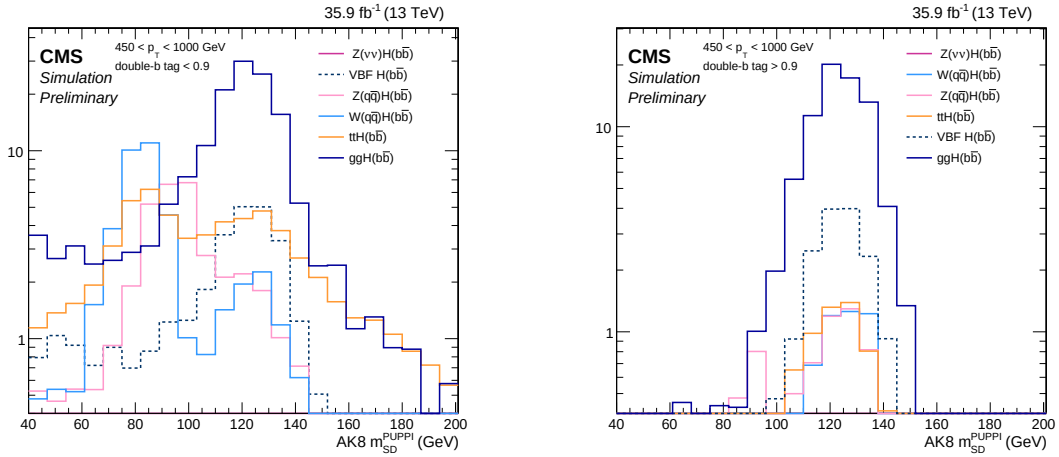

Figure 2: The m_{SD} distribution for simulated signal events after all the selection criteria in both the passing and failing regions. The yield for each process is normalized to its cross section.

5 Background estimate

The W/Z+jets backgrounds are modeled using MC simulation as are the H signal processes. The simulated W/Z+jets events are also corrected for NLO QCD and electroweak effects. Other electroweak processes are estimated from simulation and found to be negligible.

The contribution of $t\bar{t}$ production to the total SM background is estimated to be less than 3%. It is obtained from simulation corrected with scale factors derived from a $t\bar{t}$-enriched control sample, in which an isolated muon is required. The scale factors are treated as free parameters in the maximum likelihood fit used to extract the signal. They multiply the $t\bar{t}$ contribution in both the signal and control regions, correcting its overall normalization and the double-b mistag efficiency for jets originating from top quark decays.

The main background component, QCD multijet production, in the signal-enriched passing region is estimated from the signal-depleted failing region. Since the double-b tagger discriminator value and the jet mass are largely uncorrelated, the passing and failing regions have similar QCD jet mass distributions. A transfer factor $\mathcal{F}$ accounts for the residual difference in the QCD jet mass shape between the two regions and is determined by a fit to the data. This procedure is illustrated in Fig. 3.

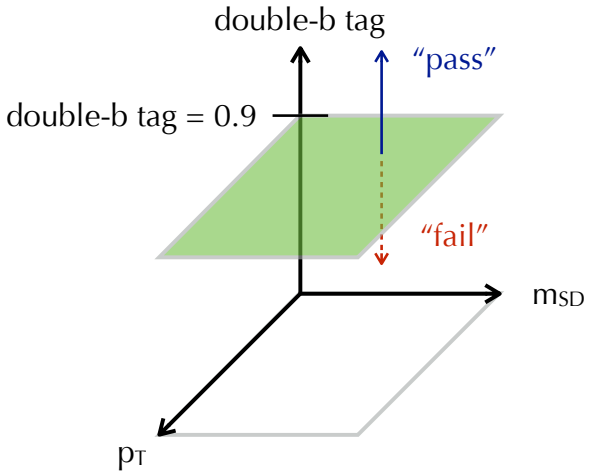

Figure 3: double-b tagger vs. jet p_T and m_{SD}

If the double-b tagger discriminator value were completely uncorrelated from the jet p_T and m_{SD}, the transfer factor would be constant. To account for deviations from this, $\mathcal{F}$ is Taylor-expanded as a polynomial in ρ and p_T, rather than m_{SD} and p_T, as the distribution of ρ is expected to be roughly invariant in all regions of p_T. For a given m_{SD} and p_T bin, the QCD yields in the passing and failing regions are related to each other through the transfer factor $\mathcal{F}(\rho, p_T) = \sum_{k,\ell} a_{k\ell} \rho^k p_T^\ell$. The QCD yield of the passing region in the i-th m_{SD} bin, with central value m_{SDi} and j-th p_T bin with value p_{Tj}, corresponding to the midpoint of the bin in logarithmic scale, is given by,

$$N_{\text{pass}}^{\text{QCD}}(m_{SDi}, p_{Tj}) = \epsilon^{\text{QCD}} \cdot \left(\sum_{k,\ell} a_{k\ell} \rho_{ij}^k p_{T_j}^\ell \right) \cdot N_{\text{fail}}^{\text{QCD}}(m_{SDi}, p_{Tj}) \,, \tag{2}$$

where ϵ^{QCD}, the polynomial coefficients $a_{k\ell}$, and the QCD contribution in each bin of the failing region, $N_{\text{fail}}^{\text{QCD}}(m_{SDi}, p_{Tj})$ are treated as free parameters determined by the fit to the data.

To determine the order of polynomial necessary to fit the data, an F-test is performed. Based on its results, a polynomial second order in ρ and first order in p_T was selected.

The m_{SD} distribution is binned in 23 bins of 7 GeV width from 40 GeV to 201 GeV and the p_T distribution is binned in six bins of increasing width from 450 GeV to 1 TeV. The signal, $\mathrm{t\bar{t}}$, and resonant electroweak backgrounds (W/Z) contributions are added as binned templates derived from MC to both the failing and passing regions.

6 Systematic uncertainties

The uncertainties on the background from QCD multijet estimate originate from the parametric uncertainties in the transfer factor fit.

The uncertainties from the $t\bar{t}$-related scale factors are automatically propagated to the signal extraction through the fit.

The nuisance parameters associated to the jet mass scale, jet mass resolution, and $N_2^{1,DDT}$ selection efficiency uncertainties of the W, Z, and $\mathrm{H}(b\bar{b})$ processes are 100% correlated. These uncertainties are constrained using a sample of merged W jets in semileptonic $t\bar{t}$ events in data. Using the same $N_2^{1,DDT}$ selection as in this search, distribution of W jet mass in the passing and failing regions in data and simulation are fit simultaneously to extract the tagging efficiency of a merged W jet and the jet mass scale and resolution in simulation and in data. The difference between hadronically decaying W and H is estimated by comparing PYTHIA and HERWIG showering algorithms, and found to be negligible.

The efficiency of the double-b tagger is measured in a data sample enriched in $b\bar{b}$ from gluon splitting [59].

These scale factors and their uncertainties determine the initial values for the W, Z, and signal shapes and they are then further constrained in the final fit in situ due to the presence of the W and Z peaks in the jet mass distribution. Finally, additional systematic uncertainties are applied to the W, Z, and $\mathrm{H}(b\bar{b})$ yields that are associated with higher-order corrections to the boson p_T distributions, jet energy scale and resolution [60], pileup modeling, and the integrated luminosity determination [61]. A quantitative summary of the systematic uncertainties considered is shown in Table 1.

In order to validate the background estimation method and associated systematic uncertainties, studies are performed on simulated samples injecting signal events and determining the bias in the measured signal cross section. No significant bias is observed in these studies.

Table 1: Systematic uncertainties and their relative size.

Systematic uncertainty source	Type (shape or normalization)	Relative size (or description)
QCD transfer factor	both	profile $a_{k\ell}$ and QCD normalization
Luminosity	normalization	2.5%
V-tag ($N_2^{1,DDT}$) efficiency	normalization	4.3%
Muon veto efficiency	normalization	0.5%
Electron veto efficiency	normalization	0.5%
Trigger efficiency	normalization	4%
Muon ID efficiency	shape	up to 0.2%
Muon isolation efficiency	shape	up to 0.1%
Muon trigger efficiency	shape	up to 8%
$t\bar{t}$ normalization SF	normalization	from 1μ CR: 8%
$t\bar{t}$ double-b mis-tag SF	normalization	from 1μ CR: 15%
W/Z NLO QCD corrections	normalization	10%
W/Z NLO EWK corrections	normalization	$15\% - 35\%$
W/Z NLO EWK ratio decorrelation	normalization	$5\% - 15\%$
double-b tagging efficiency	normalization	4%
Jet energy scale	normalization	up to 10%
Jet energy resolution	normalization	up to 15%
Jet mass scale	shape	shift m_{SD} peak by $\pm 0.4\%$
Jet mass resolution	shape	smear m_{SD} distribution by $\pm 9\%$
Jet mass scale p_T	normalization	0.4%/100 GeV (p_T)
Monte Carlo statistics	normalization	-
H p_T correction (gluon fusion)	both	30%

7 Results

The estimation of the SM background processes and the extraction of a potential signal from SM $H \to b\bar{b}$ are performed simultaneously. The resonant Z signal is used as a standard candle to simultaneously constrain the systematic uncertainties associated to it and the H boson. Results are obtained from a combined binned maximum likelihood fit to the m_{SD} distribution in data in the passing and failing regions of each p_T category, and in the $t\bar{t}$-enriched control region. The combined likelihood of the data for a given Higgs signal strength μ_H and Z signal strength μ_Z, $\mathcal{L}(\text{data}|\mu_H, \mu_Z)$, is given by the product of Poisson likelihoods in each bin multiplied by external constraints for the nuisance parameters.

Fig. 4 shows the m_{SD} distribution for data and measured SM background contributions in the passing and failing regions. Contributions from W and Z boson production are clearly visible in the data.

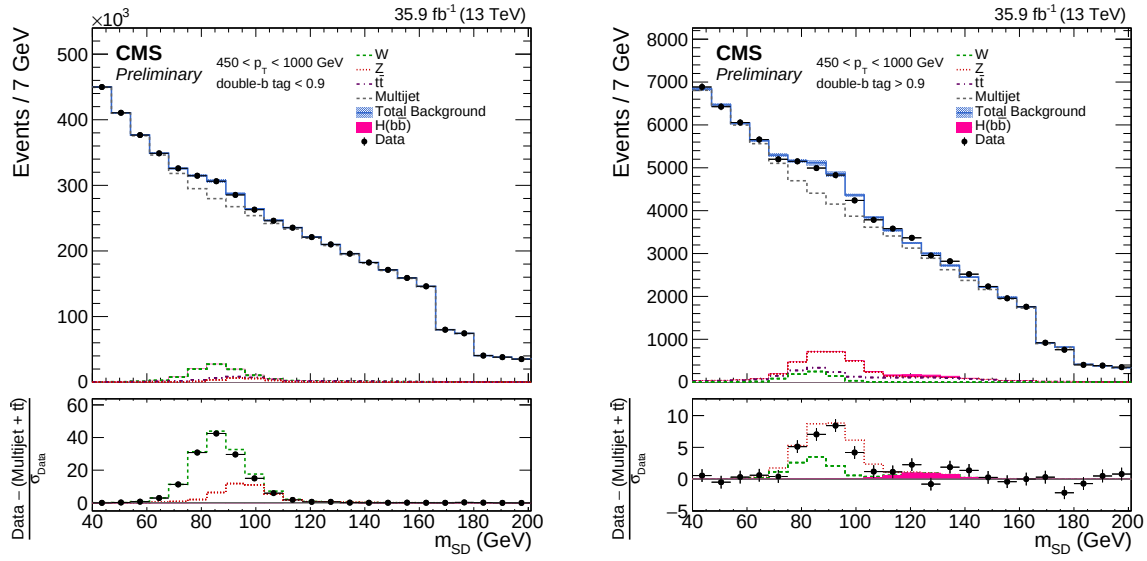

Figure 4: Post-fit m_{SD} distributions in data for the pass and fail regions and combined p_T categories by using a polynomial 2nd order in ρ and 1st order in p_T. The features at 166 GeV and 180 GeV in the m_{SD} distribution are due to the kinematic selection on ρ, which affects each p_T category differently.

The measured Z boson signal strength is $\mu_Z = 0.78^{+0.23}_{-0.19}$, which corresponds to an observed significance of 5.1σ with 5.8σ expected. This constitutes the first observation of the Z signal in the single-jet topology, further validating the substructure and b-tagging strategy for the Higgs boson search in the same topology. The measured cross section of the Z+jets process is $0.85^{+0.26}_{-0.21}$ pb, which is consistent, within the uncertainty on the measurement, with the SM. The measured H boson signal strength is $\mu_H = 2.3^{+1.8}_{-1.6}$ and includes the corrections to the p_T described in Sec. 3. The observed μ_H and the theoretical cross-section imply a measured cross-section of 74^{+51}_{-49} fb, which is consistent, within the stated uncertainty, with the SM. The observed (expected) significance is 1.5σ (0.7σ).

Tab. 2 summarizes the measured signal strengths and significances for the Higgs and Z boson processes. In particular, they are also reported for the case the corrections to the Higgs p_T spectrum are not applied. Fig. 5 shows the profile likelihood test statistic scan in data as function of the Higgs and Z signal strength parameters (μ_H, μ_Z).

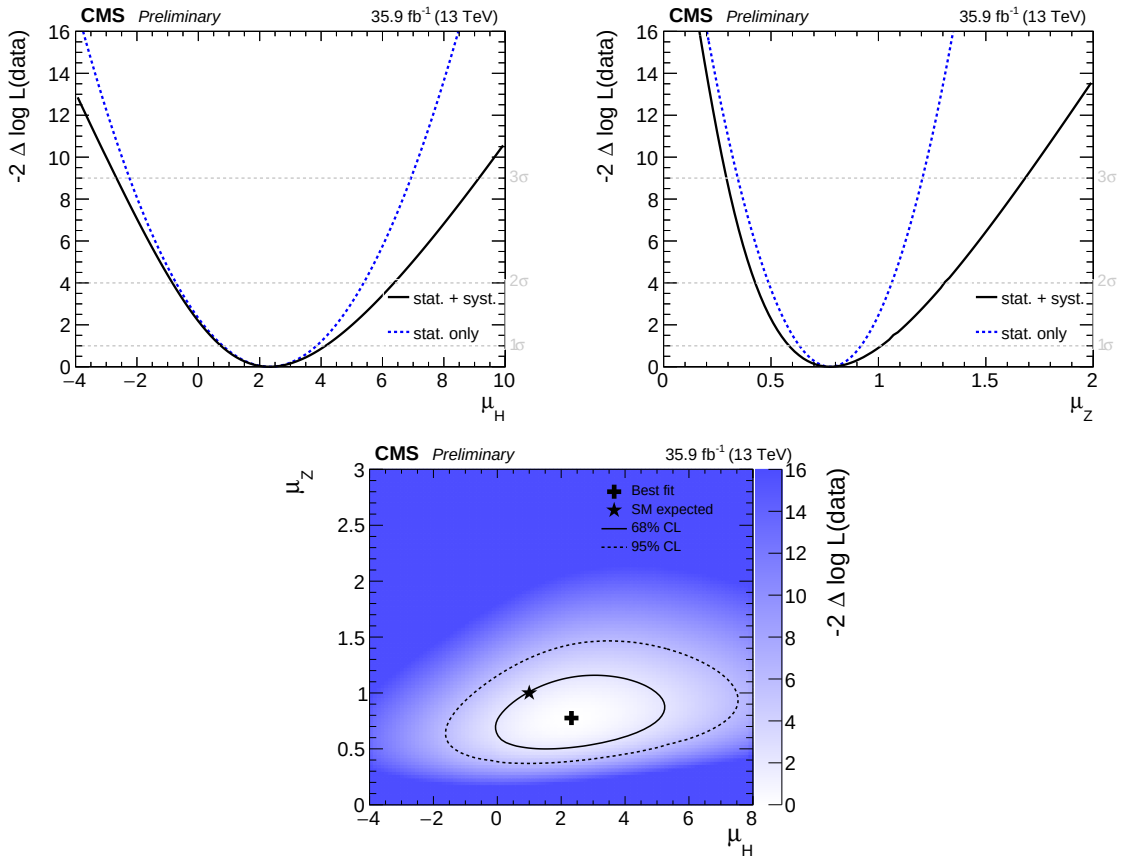

Figure 5: Profile likelihood test statistic $-2\Delta \log \mathcal{L}$ scan in data as a function of the H signal strength μ_{H} (upper left), Z signal strength μ_{Z} (upper right), and both signal strengths ($\mu_{\mathrm{H}}, \mu_{\mathrm{Z}}$) (lower).

	H	H no p_{T} corrections	Z
Observed best fit	$\mu_{\mathrm{H}} = 2.3^{+1.8}_{-1.6}$	$\mu'_{\mathrm{H}} = 3.2^{+2.2}_{-2.0}$	$\mu_{\mathrm{Z}} = 0.78^{+0.23}_{-0.19}$
Expected significance	0.7σ ($\mu_{\mathrm{H}} = 1$)	0.5σ ($\mu'_{\mathrm{H}} = 1$)	5.8σ ($\mu_{\mathrm{Z}} = 1$)
Observed significance	1.5σ	1.6σ	5.1σ

Table 2: Fitted signal strength and observed significance of the Higgs and Z signals.

8 Conclusions

An inclusive search for the standard model Higgs boson decaying to bottom quark-antiquark pairs with $p_T > 450\,\text{GeV}$ and reconstructed as a single jet has been presented using a data sample of proton-proton collisions collected by CMS corresponding to $35.9\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$. The Higgs jets are reconstructed with the anti-k_T algorithm with radius $R = 0.8$ and identified with the CMS double-b tag algorithm. The signal is then extracted on top of the falling QCD soft drop mass distribution (including contributions from W, Z, and top background processes) using an entirely data-driven QCD background prediction.

The analysis strategy is validated by extracting the signal strength of the Z+jets process. The Z+jets process is observed for the first time in the single-jet topology with a significance of 5.1σ.

The Higgs production is measured with a signal strength parameter of $\mu_H = 2.3^{+1.8}_{-1.6}$ and an observed significance of 1.5σ (0.7σ expected for the standard model Higgs) when including Higgs p_T spectrum corrections accounting for NLO and finite top mass effects. The measured cross section of the $H(b\bar{b})$ production for $p_T > 450\,\text{GeV}$ is $74^{+51}_{-49}\,\text{fb}$, which is consistent with the SM expectation within the uncertainty.

References

[1] A. Salam and J. C. Ward, "On a Gauge Theory of Elementary Interactions", *Nuovo Cim.* **19** (1961) 165–170, doi:10.1007/BF02812723.

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

[3] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

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

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

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particless", *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[7] 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) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[8] 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) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[9] The ATLAS and CMS Collaborations Collaboration, "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", Technical Report ATLAS-CONF-2015-044,CMS-PAS-HIG-15-002, CERN, Geneva, 2015.

[10] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", arXiv:1610.07922.

[11] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new Higgs search channel at the LHC", *Phys. Rev. Lett.* **100** (2008) 242001, doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.

[12] C. Grojean, E. Salvioni, M. Schlaffer, and A. Weiler, "Very boosted Higgs in gluon fusion", *JHEP* **05** (2014) 022, doi:10.1007/JHEP05(2014)022, arXiv:1312.3317.

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

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

[15] 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.

[16] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, "$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO", *JHEP* **10** (2013) 083, doi:10.1007/JHEP10(2013)083, arXiv:1306.2542.

[17] 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.

[18] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473–500, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[19] 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.

[20] 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.

[21] 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.

[22] Sjöstrand, Torbjörn and Mrenna, Stephen and Skands, Peter, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[23] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[24] 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.

[25] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[26] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC", *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10, doi:10.1016/j.nuclphysbps.2010.08.011, arXiv:1007.3492.

[27] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$", *Phys. Rev. Lett.* **110** (2013) 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[28] S. Kallweit et al., "NLO electroweak automation and precise predictions for W+multijet production at the LHC", *JHEP* **04** (2015) 012, doi:10.1007/JHEP04(2015)012, arXiv:1412.5157.

[29] S. Kallweit et al., "NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging", *JHEP* **04** (2016) 021, doi:10.1007/JHEP04(2016)021, arXiv:1511.08692.

[30] S. Kallweit et al., "NLO QCD+EW automation and precise predictions for V+multijet production", in *Proceedings, 50th Rencontres de Moriond, QCD and high energy interactions*, pp. 121–124. 2015. arXiv:1505.05704.

[31] J. M. Lindert et al., "Precise predictions for V+jets dark matter backgrounds", arXiv:1705.04664.

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

[33] U. Baur and E. W. N. Glover, "Higgs Boson Production at Large Transverse Momentum in Hadronic Collisions", *Nucl. Phys.* **B339** (1990) 38–66, doi:10.1016/0550-3213(90)90532-I.

[34] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, "Higgs boson production at the LHC: transverse momentum resummation effects in the H→2gamma, H→WW→lnulnu and H→ZZ→4l decay modes", *JHEP* **06** (2012) 132, doi:10.1007/JHEP06(2012)132, arXiv:1203.6321.

[35] M. Grazzini and H. Sargsyan, "Heavy-quark mass effects in Higgs boson production at the LHC", *JHEP* **09** (2013) 129, doi:10.1007/JHEP09(2013)129, arXiv:1306.4581.

[36] E. Bagnaschi and A. Vicini, "The Higgs transverse momentum distribution in gluon fusion as a multiscale problem", *JHEP* **01** (2016) 056, doi:10.1007/JHEP01(2016)056, arXiv:1505.00735.

[37] R. Boughezal et al., "Higgs boson production in association with a jet at next-to-next-to-leading order in perturbative QCD", *JHEP* **06** (2013) 072, doi:10.1007/JHEP06(2013)072.

[38] R. Boughezal et al., "Higgs boson production in association with a jet at NNLO using jettiness subtraction", *Phys. Lett. B* **748** (2015) 5–8, doi:10.1016/j.physletb.2015.06.055, arXiv:1505.03893.

[39] R. Boughezal et al., "Color singlet production at NNLO in MCFM", arXiv:1605.08011.

[40] X. Chen, T. Gehrmann, N. Glover, and M. Jaquier, "Higgs plus one jet production at NNLO", in *Proceedings, 12th International Symposium on Radiative Corrections (Radcor 2015) and LoopFest XIV (Radiative Corrections for the LHC and Future Colliders): Los Angeles, CA, USA, June 15-19, 2015*. 2016. arXiv:1604.04085.

[41] T. Neumann and C. Williams, "The Higgs boson at high p_T", *Phys. Rev. D* **95** (2017), no. 1, 014004, doi:10.1103/PhysRevD.95.014004, arXiv:1609.00367.

[42] M. Buschmann et al., "Mass Effects in the Higgs-Gluon Coupling: Boosted vs Off-Shell Production", *JHEP* **02** (2015) 038, doi:10.1007/JHEP02(2015)038, arXiv:1410.5806.

[43] R. Frederix, S. Frixione, E. Vryonidou, and M. Wiesemann, "Heavy-quark mass effects in Higgs plus jets production", *JHEP* **08** (2016) 006, doi:10.1007/JHEP08(2016)006, arXiv:1604.03017.

[44] V. Hirschi and O. Mattelaer, "Automated event generation for loop-induced processes", *JHEP* **10** (2015) 146, doi:10.1007/JHEP10(2015)146, arXiv:1507.00020.

[45] N. Greiner et al., "Phenomenological analysis of Higgs boson production through gluon fusion in association with jets", *JHEP* **01** (2016) 169, doi:10.1007/JHEP01(2016)169, arXiv:1506.01016.

[46] X. Chen, T. Gehrmann, E. W. N. Glover, and M. Jaquier, "Precise QCD predictions for the production of Higgs + jet final states", *Phys. Lett. B* **740** (2015) 147–150, doi:10.1016/j.physletb.2014.11.021, arXiv:1408.5325.

[47] R. Boughezal et al., "Higgs boson production in association with a jet at next-to-next-to-leading order", *Phys. Rev. Lett.* **115** (2015), no. 8, 082003, doi:10.1103/PhysRevLett.115.082003, arXiv:1504.07922.

[48] M. Cacciari et al., "Fully Differential Vector-Boson-Fusion Higgs Production at Next-to-Next-to-Leading Order", *Phys. Rev. Lett.* **115** (2015), no. 8, 082002, doi:10.1103/PhysRevLett.115.082002, arXiv:1506.02660.

[49] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", CMS Physics Analysis Summary CMS-PAS-PFT-10-001, CERN, 2010.

[50] 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.

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

[53] 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.

[54] 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.

[55] CMS Collaboration, "Jet algorithms performance in 13 TeV data", CMS Physics Analysis Summary CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[56] I. Moult, L. Necib, and J. Thaler, "New Angles on Energy Correlation Functions", *JHEP* **12** (2016) 153, doi:10.1007/JHEP12(2016)153, arXiv:1609.07483.

[57] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy Correlation Functions for Jet Substructure", *JHEP* **06** (2013) 108, doi:10.1007/JHEP06(2013)108, arXiv:1305.0007.

[58] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *JHEP* **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.

[59] CMS Collaboration Collaboration, "Identification of double-b quark jets in boosted event topologies", Technical Report CMS-PAS-BTV-15-002, CERN, Geneva, 2016.

[60] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) 11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[61] CMS Collaboration, "CMS Luminosity Measurement for the 2015 Data Taking Period", Technical Report CMS-PAS-LUM-15-001, CERN, Geneva, 2016.