Search for high-mass Higgs bosons in the final state with b-quarks with CMS Run 2 data

Jose Sánchez Fernán, University of Oviedo, Spain

September 11, 2022

Abstract

Some Beyond the Standard Model theories with an extended Higgs sector, such as the Minimal Supersymmetric Standard Model (MSSM) or Two-Higgs-Doublet models (2HDM), predict additional Higgs bosons. This project is about a search for the production of high-mass Higgs bosons in association with b-quarks. The analyzed data was collected by the CMS experiment in 2018.

Contents

1 Introduction

Some Beyond the Standard Model theories extend the Higgs sector by including more than one Higgs doublets. In particular, Two-Higgs-Doublet models (2HDM)[1] and one of its realizations, the Minimal Supersymmetric Standard Model (MSSM) [2], include 2 scalar doublets in the Higgs sector. After symmetry breaking, this results in 5 physical Higgs bosons: 2 charged scalars $H^{\pm}$, 2 neutral CP-even scalars h, H and 1 neutral CP-odd pseudoscalar A. By convention, h is lighter than H and it is usually associated with the Higgs boson discovered in 2012 [3, 4].

In certain scenarios, the coupling of these Higgs bosons to b-quarks is greatly enhanced. Therefore, the search for high mass Higgs bosons with b-quarks in the final state is of particular interest [5, 6]. In the analysis that this project contributes to, the $A/H \to b\bar{b}$ decay channel is considered, probing masses in the $300 - 1800\,\mathrm{GeV}$ range using data collected in 2018 with the CMS detector [7].

2 Signal and background processes

The main signal for this analysis is the production of a neutral high-mass Higgs boson in association with b-quarks and decaying into a pair of b-quarks. A diagram of this process is shown in Fig. 1. Therefore, 4 b-jets are expected in the final state. However, the fourth jet is usually too soft to be reconstructed. Thus, 3 b-tagged jets are required in this analysis. The 2 leading jets are associated with the Higgs daughters, so their reconstructed invariant mass is the main observable.

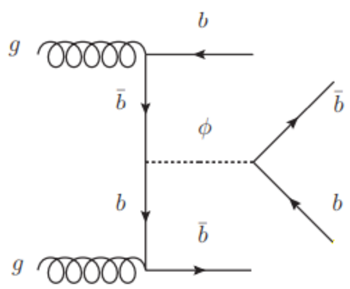

Figure 1: Feynman diagram of the main signal process.

The background is comprised of processes with at least 3 b-quarks or 2 b-quarks and a light flavour quark in the final state. The main contribution comes from QCD multi-jet production. Additionally, there is a small contribution from $t\bar{t}$ production. Example background processes are shown in Fig. 2.

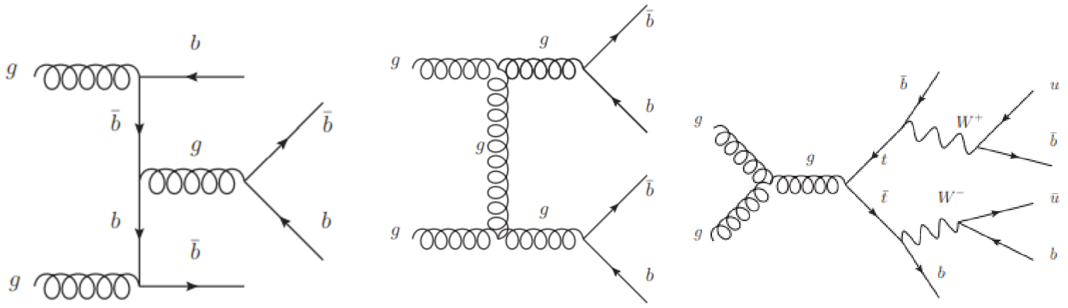

Figure 2: Feynman diagrams of main background processes: QCD multi-jet production (left and center), $t\bar{t}$ production (right).

3 Event selection

The trigger employed in this analysis selects events with at least 2 b-tagged jets by the DeepCSV algorithm [8]. The two leading jets are required to have $p_t \geq 116\,\text{GeV}$. In addition, other requirements regarding the spatial distribution of the jets are imposed.

For the offline event selection at least 3 jets are required. Moreover, the following kinematic cuts are applied:

- The two leading jets are required to have $p_t > 130\,\text{GeV}$. A lower threshold of $p_t > 40\,\text{GeV}$ is applied to the third jet.
- In order to improve b-tagging performance, a cut of $|\eta| \leq 2.2$ is applied to the three leading jets.
- Angular separation between each pair of the three leading jets is imposed by a cut of $\Delta R > 1.0$. This selection cut filters out the background from gluon splitting and improves the b-tagging by reducing contamination.
- A $\Delta\eta_{12} \leq 1.5$ cut is also applied to increase the signal to background ratio.

Regarding the tagging of b-jets, the DeepJet algorithm [9] is used for the offline selection. In the signal region (SR), the 3 leading jets have to pass the medium working point requirement. In the control region (CR), the 2 leading jets have to pass the medium working point, while the third jet must not pass the loose requirement.

4 Background model

The background model is obtained from the data in the CR. It is given by the following relation:

$$\text{SR Background} = \text{CR Parametrization} \times \text{Transfer factor}$$

where the transfer factor is a correction obtained from the simulation of QCD events in the SR and CR.

To obtain the parametrization of the CR, the distribution of the data is fit to an extended Novosibirsk distribution. However, the distribution of the data is not optimally described by a unique set of parameters across the full mass range. Therefore, for an improved description, the parametrization of the CR is performed in different overlapping fit ranges defined as:

- FR1: $260 - 550\,\mathrm{GeV}$
- FR2: $320 - 800\,\mathrm{GeV}$
- FR3: $380 - 2000\,\mathrm{GeV}$

5 Signal model

The expected signal is simulated to next at leading order for each of the mass points using the POWHEG generator [10]. After applying several corrections and applying the event selection to the simulated samples the signal distribution is obtained.

The signal distributions are parametrized with a double-sided Crystal Ball function. This function behaves as a Gaussian distribution for central values and as power laws in both of the tails. It depends on 6 parameters: μ and σ of the Gaussian section; α_1 and α_2, which determine where the change in behaviour happens for the left and right tails; n_1 and n_2, which determine the exponent of the respective power laws.

Different sub-ranges are used for the fit depending on the mass point. These are:

- SR1: $260 - 550\,\mathrm{GeV}$ $\quad m_{A/H} = 300, 350, 400\,\mathrm{GeV}$
- SR2: $320 - 800\,\mathrm{GeV}$ $\quad m_{A/H} = 450, 500, 600, 700\,\mathrm{GeV}$
- SR3: $380 - 2000\,\mathrm{GeV}$ $\quad m_{A/H} = 800, 900, 1000\,\mathrm{GeV}$
- SR4: $500 - 2000\,\mathrm{GeV}$ $\quad m_{A/H} = 1200, 1400, 1600, 1800\,\mathrm{GeV}$

I performed the fit for each of the mass points. As an example, figure 3 shows the resulting function for one mass point in each of the sub-ranges. The quality of the fit is lower for the 600 GeV and 700 GeV mass points, which is caused by the shape of the signal distribution deviating slightly from the fit function. In addition, the values obtained for α_1, α_2, n_1 and n_2 tend to have wide uncertainties. The reason for this could be that part of the respective tails lies outside of the sub-range of the fit.

6 Systematic Uncertainties

Corrections applied to the simulated signal constitute a source of systematic uncertainty in this analysis, since they can change the normalization or the shape of the signal distribution. During the project I studied the effects coming from:

- Pileup reweighting (PU): Matches the pileup profile of data and MC.
- Jet energy resolution (JER): Adds smearing to the p_t distribution of the jets to simulate the effects of the p_t resolution in the data.
- Jet energy scale (JES): Multiplicative factor that corrects the energy of the jets.
- Jet kinematic trigger efficiency (JKTE): Matches the efficiency of the kinematic trigger for data and MC.

In order to assess the effect of these systematic uncertainties, the signal distributions are obtained by applying the $\pm 1\sigma$ variations of the corrections instead of their central values. Then, the resulting histograms are compared to the ones obtained from the central values.

6.1 Pileup reweighting

The signal histograms obtained for the central and up and down variations of the pileup reweighting are shown in Fig. 4 for three representative mass points. As can be observed in the plot, the ratio between the histograms is approximately constant. This leads to the conclusion that this systematic uncertainty is not shape altering. Table 1 shows the relative change in normalization due to the variations for the same mass points. This normalization effect is small compared to other sources. As a consequence, the uncertainty arising from pileup reweighting is negligible.

Δnorm (%)	PU up	PU down
350 GeV	-0.33	0.24
600 GeV	-0.24	0.25
1200 GeV	-0.06	-0.04

Table 1: Relative deviation of the normalization with respect to the nominal value due to pileup reweighting variations.

6.2 Jet Energy Resolution and Jet Energy Scale

The signal histograms obtained for the central and up and down variations of the JER and JES corrections are shown in Fig. 5 and 6 for three representative mass points. Both of these uncertainties are shape altering.

In order to study how the variations affect each of the parameters associated with the signal, the fits are performed again for the variations while fixing different sets of parameters to their nominal values. This is done by letting only one parameter float at a time, letting all parameters float and letting only the mean and the sigma float.

The parameters α_1, α_2, n_1 and n_2 present large variations and wide uncertainties for many of the mass points, as in the fit for the central values. In contrast, the mean and the sigma are stable under the variations and have smaller deviations in general. Table 2 shows the relative deviations of the latter when fixing the rest of the parameters.

350 GeV	JER up	JER down	JES up	JES down
$\Delta\mu$ (%)	0.6	0.4	1.3	-0.3
$\Delta\sigma$ (%)	-5.2	-5.7	-4.5	-5.6
600 GeV				
$\Delta\mu$ (%)	0.15	-0.17	0.83	-0.83
$\Delta\sigma$ (%)	0.02	0.22	0.29	0.08
1200 GeV				
$\Delta\mu$ (%)	0.12	-0.11	0.59	-0.61
$\Delta\sigma$ (%)	-0.37	0.25	0.19	-0.18

Table 2: Relative deviation of the mean and sigma with respect to the nominal values due to JER and JES variations. α_1, α_2, n_1 and n_2 are fixed to their nominal values.

Regarding the normalization, JES has a stronger effect than the other systematic uncertainties. Relative deviations caused by JER and JES for the three representative mass points are shown in table 3.

Δnorm (%)	JER up	JER down	JES up	JES down
350 GeV	0.9	-1.2	5.6	-5.7
600 GeV	0.3	-0.3	1.6	-1.7
1200 GeV	0.1	-0.2	0.6	-0.6

Table 3: Relative deviation of the normalization with respect to the nominal value due to JER and JES variations.

6.3 Jet Kinematic Trigger Efficiency

The signal histograms obtained for the central and up and down variations of the JKTE correction are shown in Fig. 7 for three representative mass points. This uncertainty has a small shape altering effect on the lower mass points and does not alter the shape for the medium and high masses. However, it is considered as shape altering for all the mass points for simplicity.

The effects on the fit parameters due to these variations is generally small compared to jet energy corrections. Table 4 shows the relative deviations of the mean and the sigma when fixing the rest of the parameters.

350 GeV	JKTE up	JKTE down
$\Delta\mu$ (%)	-0.02	0.03
$\Delta\sigma$ (%)	0.06	-0.25
600 GeV		
$\Delta\mu$ (%)	-0.01	0.003
$\Delta\sigma$ (%)	0.17	-0.08
1200 GeV		
$\Delta\mu$ (%)	-0.004	-0.003
$\Delta\sigma$ (%)	-0.04	0.02

Table 4: Relative deviation of the mean and sigma with respect to the nominal values due to JKTE variations. α_1, α_2, n_1 and n_2 are fixed to their nominal values.

7 Cross section limits

As a last stage of the project, I obtained the expected 95% CL upper limits for $\sigma(b\bar{b}A/H)\mathrm{BR}(A/H \to b\bar{b})$ including the effect of systematic uncertainties on the signal model using Combine Tool [11] for the statistical inference procedure. The effects from systematic uncertainties included in the fit within the scope of this project are:

- JER effect on the sigma.
- JES effect on the mean and normalization.
- JKTE effect on the sigma.

The focus on the sigma and mean parameters is justified as these two parameters describe the peak of the signal distribution and carry most of the physical information.

8 Summary

This project focused on the study of systematic uncertainties within an ongoing analysis searching for high-mass Higgs bosons with b-quarks in the final state using 2018 CMS data.

The effect of pileup reweighting was found to be negligible. The effect of Jet Energy Scale, Jet Energy Resolution and Jet Kinematic Trigger Efficiency corrections on the signal shape were obtained. Finally, the studied systematic uncertainties were included in the statistical inference procedure in order to obtain upper limits for the production cross section of a neutral high-mass Higgs boson in association with b-quarks and decaying into a pair of b-quarks.

Acknowledgements

I would like to thank my supervisors Sandra Consuegra Rodríguez and Daina Leyva Pernia for their guidance and support throughout the program, as well as other members of the DESY CMS Higgs group for their valuable feedback. Moreover, I would like to thank the DESY Summer Student Program 2022 organisers for this unique opportunity.

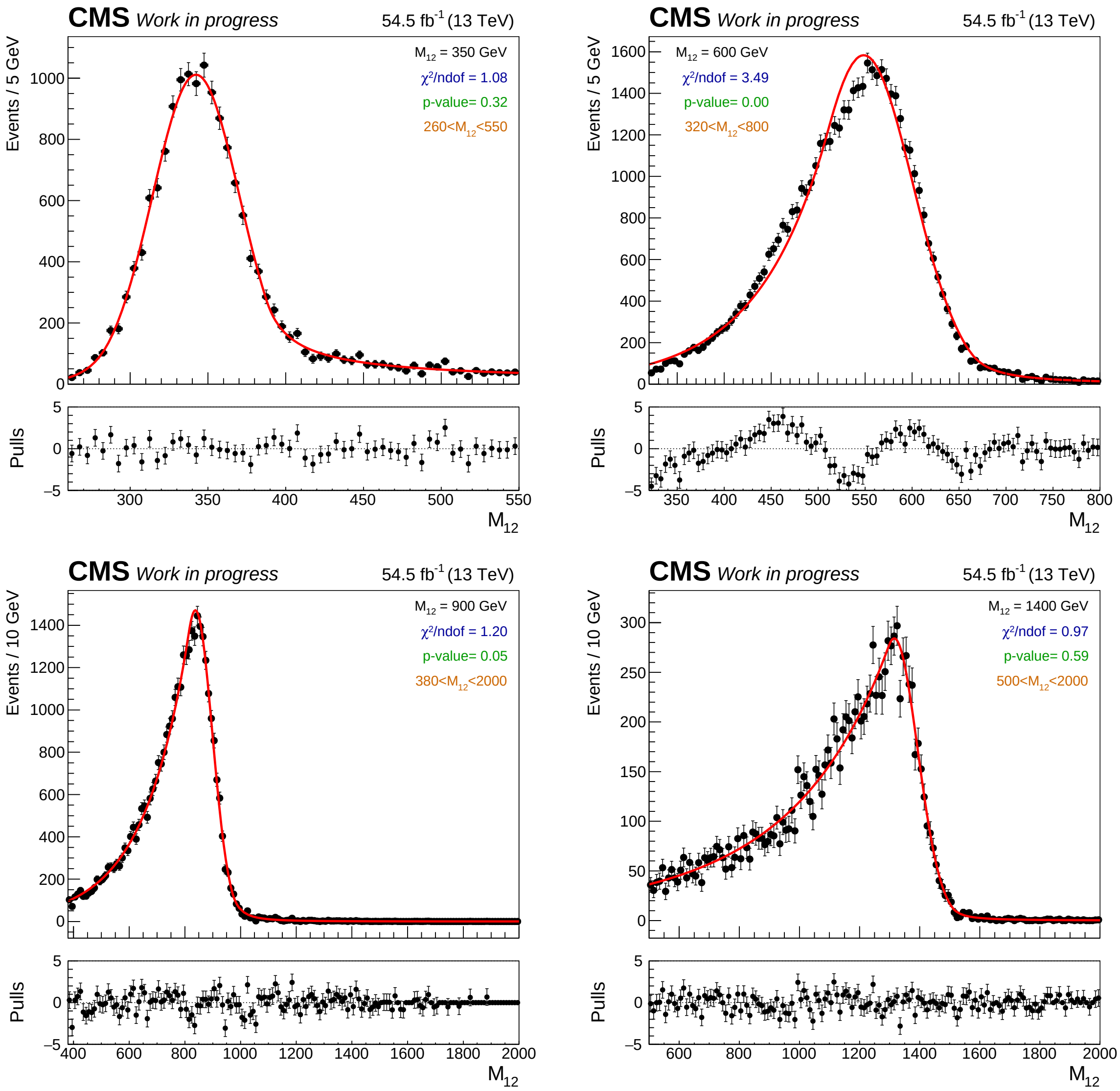

Figure 3: Distribution of the simulated signal and associated fit function for 4 mass points.

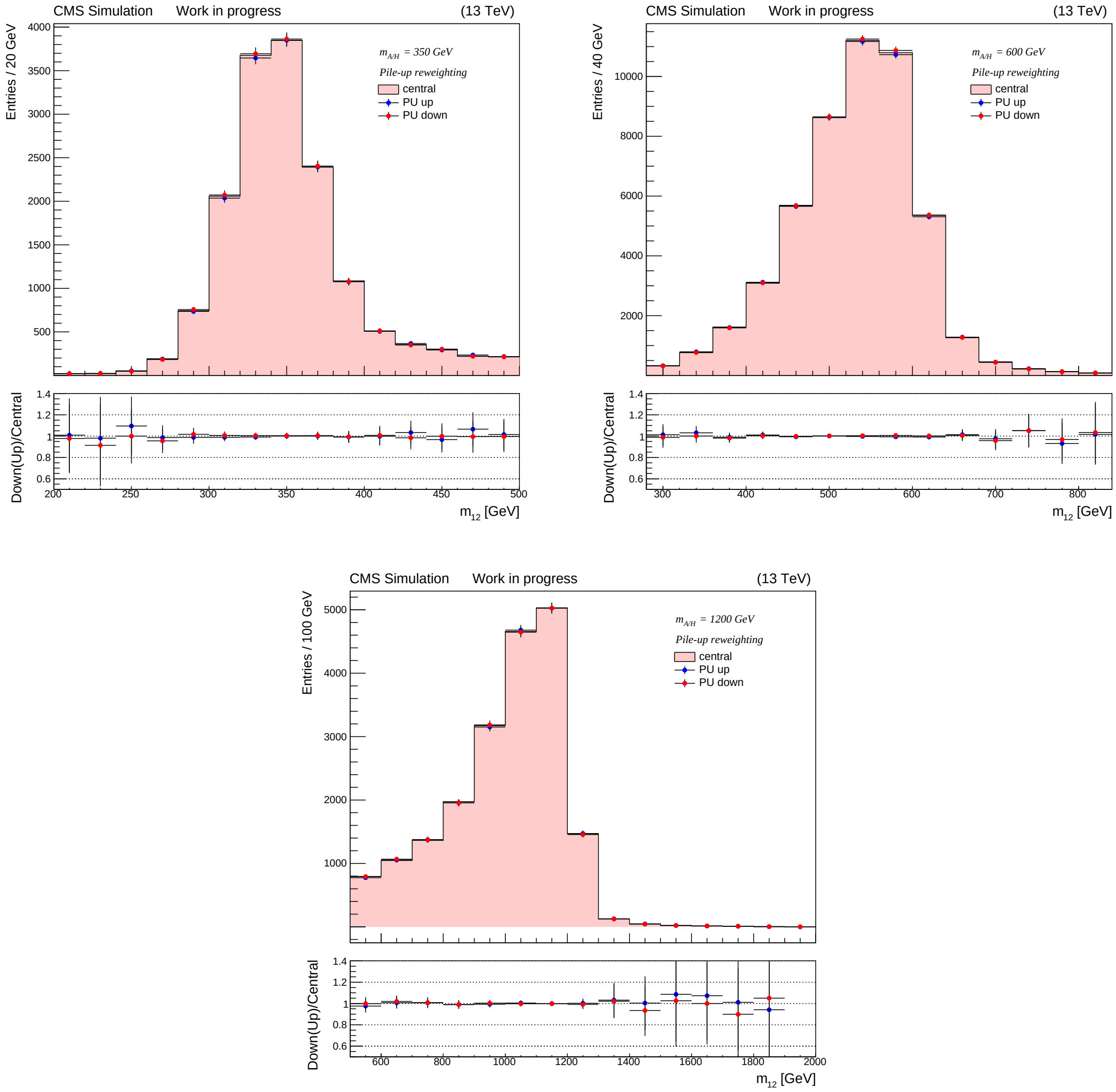

Figure 4: Distributions of the simulated signal obtained with the central values of the corrections as well as pileup reweighting up and down variations.

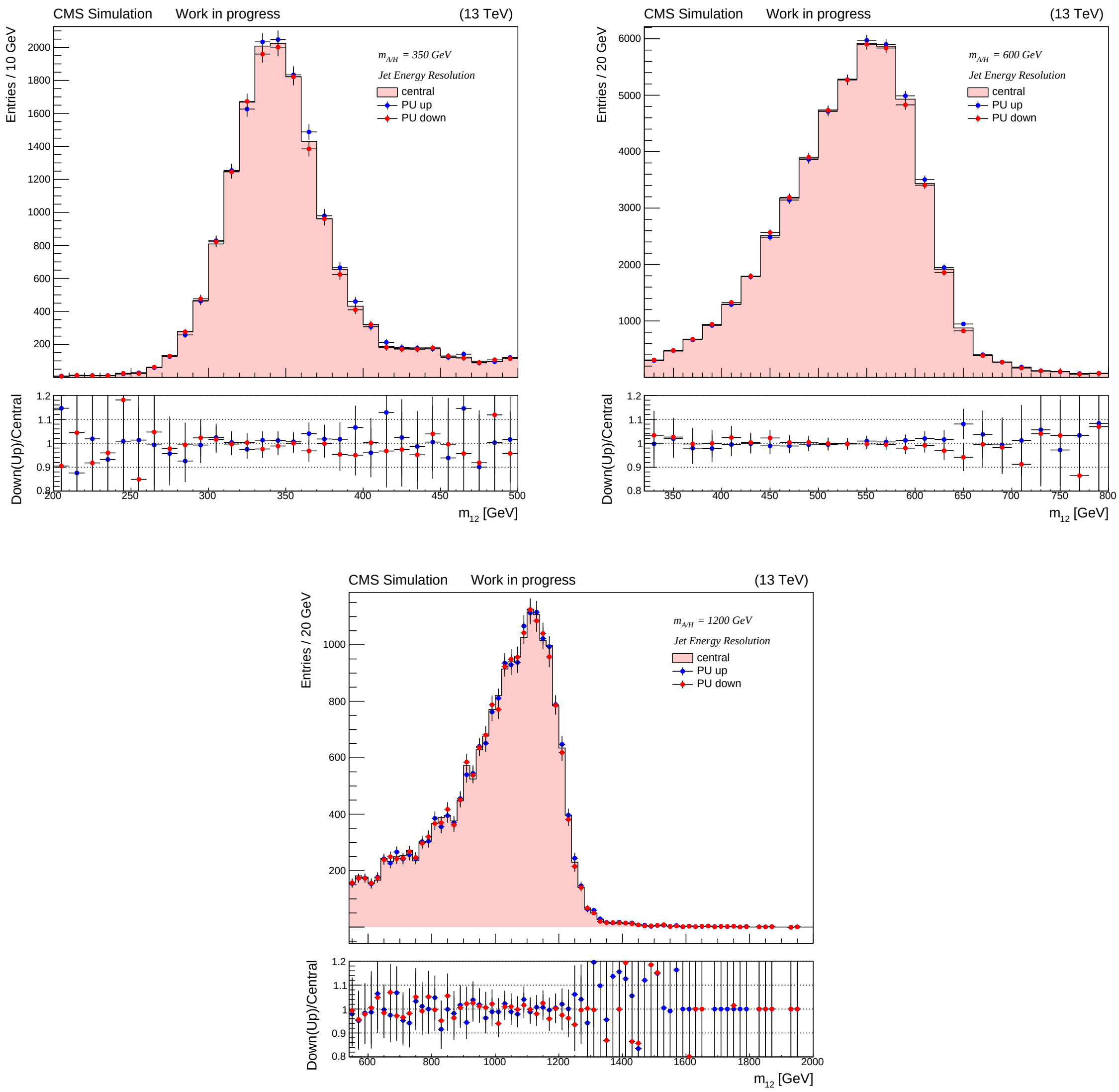

Figure 5: Distributions of the simulated signal obtained with the central values of the corrections as well as JER up and down variations.

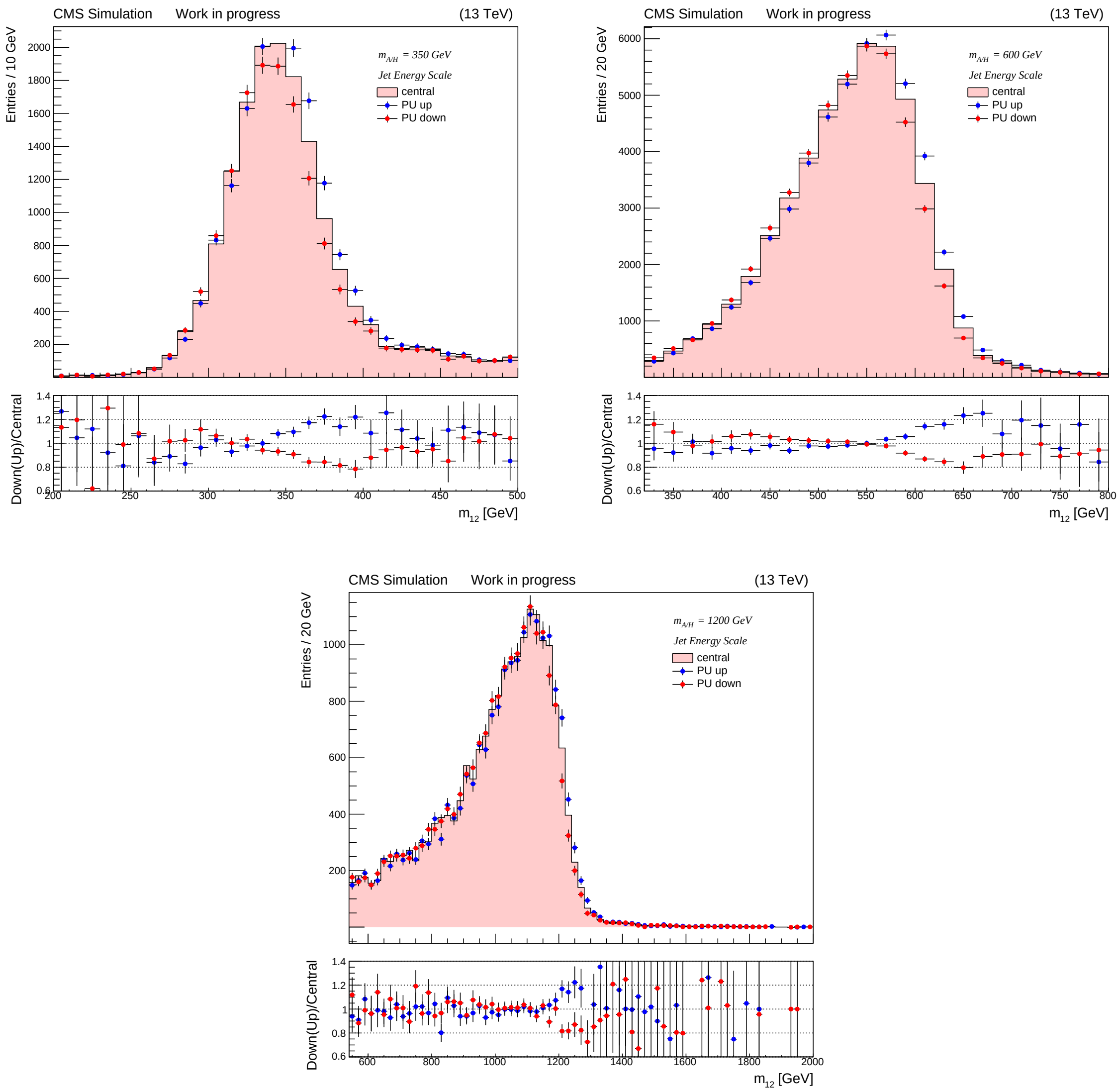

Figure 6: Distributions of the simulated signal obtained with the central values of the corrections as well as JES up and down variations.

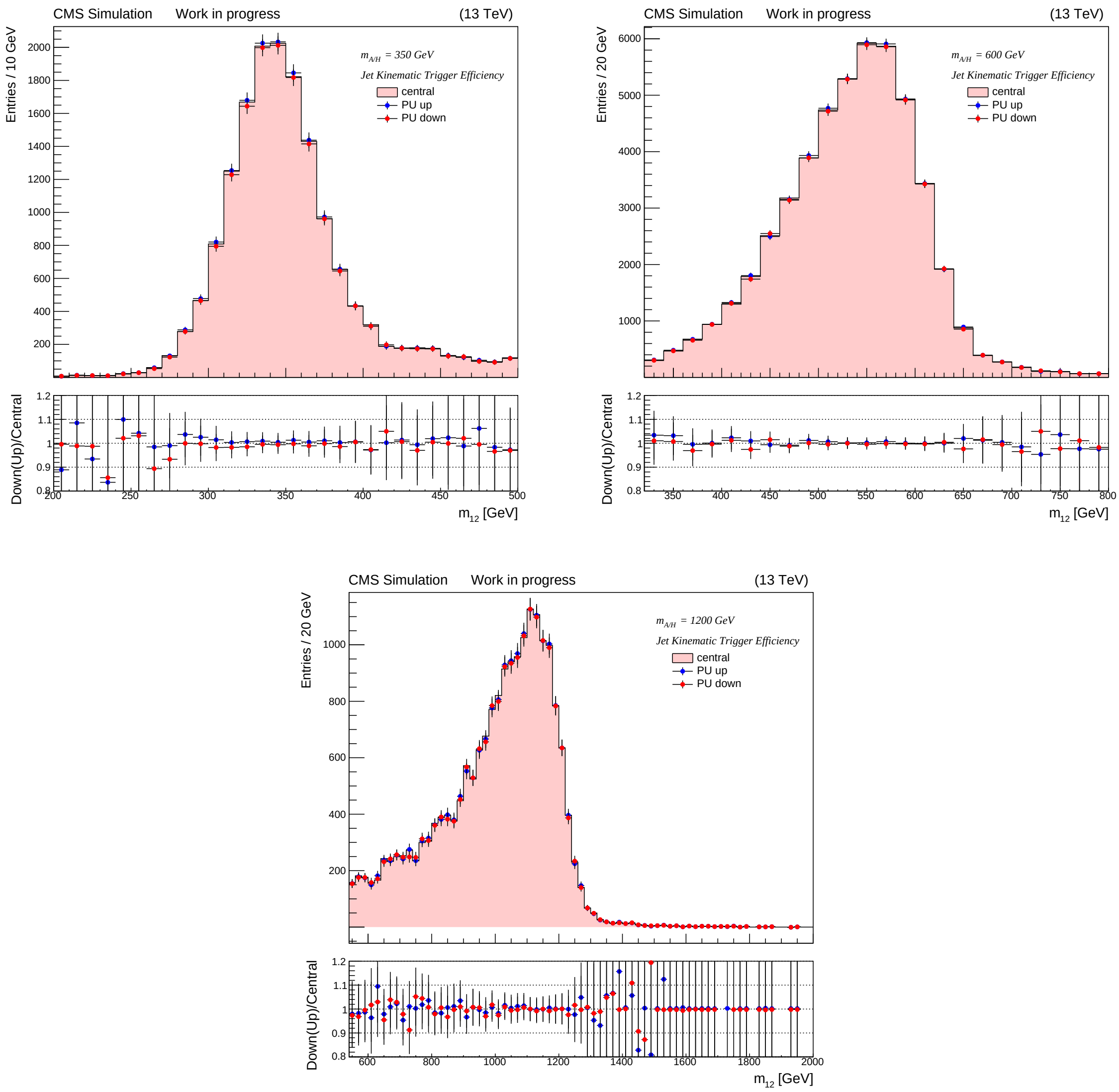

Figure 7: Distributions of the simulated signal obtained with the central values of the corrections as well as JKTE up and down variations.

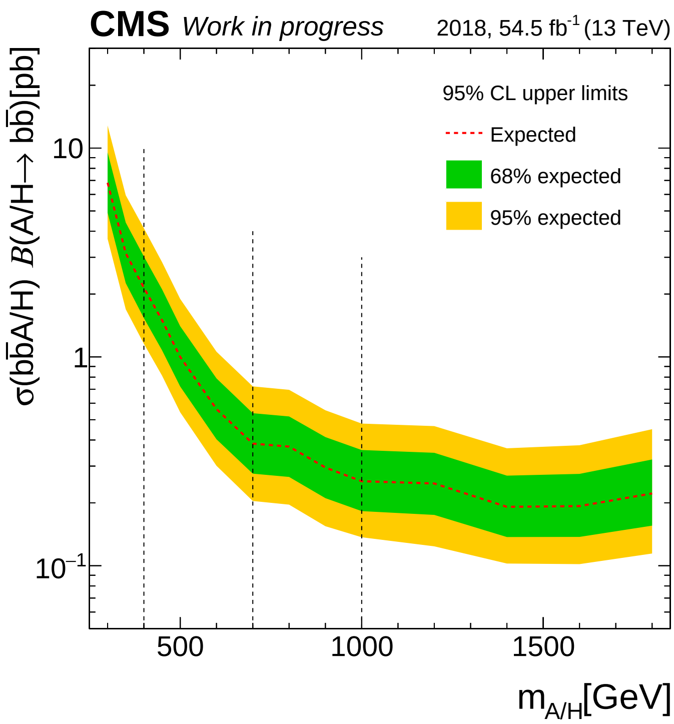

Figure 8: Expected 95% CL upper limits for the production of a neutral high-mass Higgs boson in association with b-quarks and decaying into a pair of b-quarks.

References

[1] G. Branco et al., “Theory and phenomenology of two-Higgs-doublet models”, Phys. Rept. 516 (2012) 1–102, doi:10.1016/j.physrep.2012.02.002, arXiv:1106.0034.

[2] D. Curtin et al., “Exotic decays of the 125 GeV Higgs boson”, Phys. Rev. D 90 (2014), no. 7, 075004, doi:10.1103/PhysRevD.90.075004, arXiv:1312.4992.

[3] 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.

[4] 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.

[5] Shevchenko, R., “Search for high mass higgs bosons in the final state with b-quarks with the CMS 2016 data”, Dissertation, Universität Hamburg, 2019, doi:10.3204/PUBDB-2019-01182.

[6] Asmuss, J. P., “Search for high-mass bosons of an extended Higgs sector in b quark final states using the 2017 data set of the CMS experiment”, Dissertation, Universität Hamburg, 2021, doi:10.3204/PUBDB-2021-03020.

[7] CMS Collaboration, “The CMS Experiment at the CERN LHC”, JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[8] D. Guest et al., “Jet Flavor Classification in High-Energy Physics with Deep Neural Networks”, Phys. Rev. D 94 (2016), no. 11, 112002, doi:10.1103/PhysRevD.94.112002, arXiv:1607.08633.

[9] E. Bols et al., “Jet flavour classification using DeepJet,” Journal of Instrumentation, vol. 15, no. 12, p. P12012, 2020.

[10] B. Jager, L. Reina and D. Wackeroth, “Higgs boson production in association with b jets in the POWHEG BOX,” Phys. Rev. D **93** (2016) no.1, 014030 doi:10.1103/PhysRevD.93.014030 [arXiv:1509.05843 [hep-ph]].

[11] ATLAS, CMS 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”, JHEP 08 (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.