Study of corrections for B-Tagged jets

Fabio Zoratti*
Scuola Normale Superiore
Università di Pisa

Supervisor
Roberval Walsh
DESY

Supervisor
Antonio Vagnerini
DESY

September 5, 2018

Abstract

A search for Higgs bosons that decay into a bottom quark-antiquark pair and are accompanied by at least one additional bottom quark is performed with the CMS detector. The final state considered in this analysis is particularly sensitive to signatures of a Higgs sector beyond the standard model, as predicted in the generic class of two Higgs doublet models (2HDMs). A study of jet corrections is performed to increase the overall signal over background ratio.

*fabio.zoratti@sns.it

Contents

1 Introduction

In the standard model (SM), a Higgs boson at a mass of 125 GeV has a large coupling to b quarks via Yukawa interactions. Its production in and subsequent decay into b quarks at the CERN LHC has been observed for the first time [1] after six years of its discovery [2, 3, 4, 5], despite the challenging high rate of heavy-flavour multijet production.

However, there are models beyond the SM that predict an enhancement of Higgs boson production in association with b quarks, which motivate further search for such processes, as done in [6, 7, 8].

Prominent examples of models beyond the SM are the two Higgs doublet model (2HDM) [9], which contains two scalar Higgs doublets, as well as one particular realization within the minimal supersymmetric extension of the SM (MSSM) [10]. These result in two charged Higgs bosons, $H^{\pm}$ and three neutral ones, jointly denoted as ϕ. Among the latter are, under the assumption that CP is conserved, one CP-odd (A), and two CP-even (h, H) states, where h usually denotes the lighter CP-even state. For the purpose of this analysis, the boson discovered in 2012 with a mass near 125 GeV [2, 3, 4, 5] is interpreted as h, whose mass is thus constrained to the measured value. The two heavier neutral states, H and A, are the subject of the search presented here.

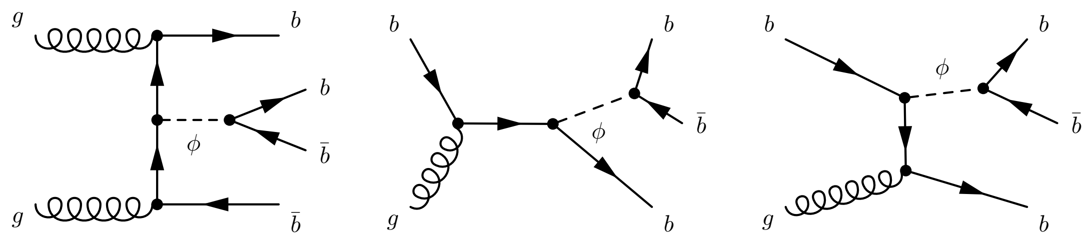


Figure 1: example Feynman diagrams for the signal process.

2 Motivation

A resonance should be observed in a first approximation as a Breit-Wigner on an invariant mass plot. Unfortunately, the jet energy resolution [11] is several orders of magnitude larger than the natural width of the Higgs boson [12]. Calculating better the jet parameters can help to sharpen the shape of the resonance.

One of the most common decay modes of b quarks is via weak interaction, where a W boson is produced. This W boson can decay in a charged lepton-neutrino pair, so an important aspect is correctly taking into account the missing energy in an event where at least a neutrino is present in one of the leading jets. The missing energy will shift the peak to the left of the expected position. Jet energy regression [13] is a method to correct the energy and p_T, even not knowing the four-momentum of these weakly interacting particles.

Another kind of correction we can make is taking into account possible final state radiation jets, as explained in Section 5.4.

The aim of this study is looking at the effects of these corrections on the sensitivity, as known as $\frac{\text{Signal}}{\sqrt{\text{Bkg}}}$, to improve our capability of discovering SUSY particles or to exclude their presence.

3 Event reconstruction and simulation

The CMS [14] (Compact Muon Solenoid) apparatus is together with ATLAS [15] one of the two LHC detectors designed to probe the mass range of the Higgs boson and investigate new high energy physics. Its central feature is a superconducting solenoid of 6 m internal diameter, producing a magnetic field of 3.8 T. Within the field volume are the silicon tracker, the crystal electromagnetic calorimeter and the brass/scintillator hadron calorimeter. The silicon strip detector includes two cylindrical barrel detectors and two endcap systems. Muons are detected in a gas-ionisation system inserted in the steel return yoke. Detailed information about the CMS detector can be found in Ref. [14]. The CMS detector uses a right-handed coordinate system, with the nominal interaction point identified as the origin, the x axis directed to the centre of the ring and the z axis aligned with the counterclockwise direction of the beam. Given the polar angle θ measured from the z axis and the azimuthal angle ϕ corresponding to the radial direction in the $x - y$ y plane, the pseudorapidity is defined as:

$$\eta = \ln(\tan(\theta/2))$$

The acceptance in pseudorapidity is $|\eta| < 2.5$ for the tracker and $|\eta| < 2.4$ for the muon system. The other key kinematic variable is the transverse momentum p_T (the component perpendindicular to the beam axis) which can be converted into its Cartesian components p_x, p_y and p_z by means of the following transformations

$$p_x = p_T \cos\phi; \quad p_y = p_T \sin\phi; \quad p_z = p_T \sinh\eta$$

from which it follows that $|p| = p_T \cosh\eta$. The p_T resolution for muon candidates detected by the combination of the muon and tracker system within a range of $5 - 100\,\text{GeV/c}^2$ is of about $1 - 3\%$. The parameter for the separation of hadron jets R is obtained by adding in quadrature the changes in the azimuthal angle and the pseudorapidity η:

$$\Delta R(\text{jet}_i, \text{jet}_j) = \sqrt{\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2}$$

For this study, the signal and background samples were produced using PYTHIA 8[16]. The CMS detector response was modelled using GEANT4 toolkit [17].

A particle-flow algorithm [18] aims to reconstruct and identify all particles in the event, i.e. electrons, muons, photons, and charged and neutral hadrons, with an optimal combination of all CMS detector systems.

The reconstructed vertex with the largest value of summed physics-object p_T^2 is taken to be the primary pp interaction vertex. The physics objects chosen are those that have been defined using information from the tracking detector, including jets, the associated missing transverse momentum, which is taken as the negative vector sum of the p_T of those jets, and charged leptons.

Jets are clustered from the reconstructed particle-flow candidates using the anti-k_T algorithm [19] with a distance parameter of 0.4. Each jet is required to pass dedicated quality criteria to suppress the impact of instrumental noise and misreconstruction. Contributions from additional pp interactions within the same or neighbouring bunch crossing (pileup) affect the jet momentum measurement. To mitigate this effect, charged particles associated with other vertices than the reference primary vertex are discarded before jet reconstruction [20], and residual contributions (e.g. from neutral particles) are accounted for using a jet-area based correction [21].

4 Trigger and event selection

A major challenge to this search is posed by the huge hadronic interaction rate at the LHC. This is addressed with a dedicated trigger scheme [22], specially designed to suppress the multijet background. Only events with at least two jets in the range of $|\eta| < 2.3$ are selected.

Different cuts are applied for different kind of events. For this report, we are going to call semileptonic events (SL) all events in which at least one of the 2 leading jets includes one neutrino[1] and full hadronic events (FH) the events in which no neutrino is present. The two leading jets are required to have $p_T >$ $100\,\mathrm{GeV(FH)}/40\,\mathrm{GeV(SL)}$, and an event is accepted only if the absolute value of the difference in pseudorapidity, $\Delta\eta$, between any two jets fulfilling the p_T and η requirements is less than or equal to 1.5. The tight online requirements on the opening angles between jets are introduced to reduce the trigger rates while preserving high efficiency in the probed mass range of the Higgs bosons. At trigger level, b jets are identified using the DeepCSV [23, 24] algorithm at the medium working point. This working point features a 1% probability for light-flavour jets

[1] Actually we are checking the presence of a charged lepton, like electron, muon or tau, the neutrino is not detected but must be present because of the topology of a W boson decay.

(attributed to u, d, s, or g partons) to be misidentified as b jets. At least two jets in the event must satisfy the online b tagging criteria.

For the semileptonic channel, our selection is more strict: we require a tight muon identification and a p_T greater than 12 GeV for this charged lepton, to use at best the CMS muon trigger.

Our signal region will be made up of events where there is another b-tagged jet satisfying a p_T cut of 40 GeV(30 GeV) for the FH(SL) channel. Our control region will be composed of events in which the third jet satisfies a reverse b-tag requirement.

Furthermore, a fourth jet veto is applied in order to suppress the $t\bar{t}$ background. If there is a fourth jet, its p_T has to be smaller than 40 GeV(FH)/30 GeV(SL). In every event we require that both the two leading jets are matched with trigger objects, otherwise, the event is rejected.

5 Corrections

There are four kinds of corrections applied to this study.

- Smearing
- B-Tag scale factors
- Jet energy regression
- Final state radiation (FSR)

5.1 Smearing

This is a correction that has to be applied only on simulated data. Previous measurements showed that the jet energy resolution [11] (JER) in the real data is worse than in the simulation and the jets in MC need to be smeared to describe the data. There are 2 different methods to smear the reconstructed data. Both of them are in fact a scale factor for the p_T of the jet, so the final p_T^{smeared} will simply be

$$p_T^{\text{smeared}} = c_{\text{JER}} p_T$$

where c_{JER} will be defined in Equation 1 and Equation 2. The recommendation is following the "hybrid" method: when matching particle-level jet is found, the scaling method should be used; otherwise, the stochastic smearing should be applied.

Scaling method

$$c_{\mathrm{JER}} = 1 + (s_{\mathrm{JER}} - 1)\frac{p_T - p_T^{\mathrm{ptcl}}}{p_T} \tag{1}$$

where p_T is its transverse momentum, p_T^{ptcl} is the transverse momentum of the corresponding jet clustered from generator-level particles, and s_{JER} is the data-to-simulation core resolution scale factor. Factor c_{JER} is truncated at zero, i.e. if it is negative, it is set to zero. This method only works if a well-matched particle-level jet is present and can result in a large shift of the response otherwise. The following requirements are imposed for the matching:

$$\Delta R < R_{\mathrm{cone}}/2 \qquad |p_T - p_T^{\mathrm{ptcl}}| < 3\sigma_{\mathrm{JER}} p_T$$

Here R_{cone} is the jet cone size parameter [2] and σ_{JER} is the relative p_T resolution as measured in simulation.

Stochastic method The other approach, which does not require the presence of a matching particle-level jet, is the stochastic smearing. In this case, corrected jet four-momentum is rescaled with a factor

$$c_{\mathrm{JER}} = 1 + \mathcal{N}(0, \sigma_{\mathrm{JER}})\sqrt{\max(s_{\mathrm{JER}}^2 - 1, 0)} \tag{2}$$

where σ_{JER} and s_{JER} are the relative p_T resolution in simulation and data-to-simulation scale factors, and $N(0, \sigma)$ denotes a random number sampled from a normal distribution with a zero mean and variance σ^2. As before, scaling factor c_{JER} is truncated at zero if negative. This method only allows to degrade the resolution.

5.2 B-Tag scale factors

To emulate real data when studying the MC, we should take into account the inefficiency of b-tagging [23]. This is done by applying a weight to every event. The weights are retrieved from a database and every histogram is created using them. This will result in a factor which has the order of magnitude of $\frac{1}{2}$ on all the histograms after these scale factors.

5.3 Jet energy regression

A Deep Neural Network (DNN) approach has been used to improve the reconstruction of the jet energy and p_T, as explained in Ref. [13]. This correction

[2] In our case, 0.4 for AK4 jets

should affect more the events in which one of the two leading jets includes a lepton because the aim of that approach was taking into account the missing energy of the neutrino.

In the framework I used, NanoAOD, this correction is already calculated and saved as a jet variable, so no further computations are required for the application of this correction.

5.4 Final state radiation

It may happen that a b quark irradiates a gluon before hadronizing, generating a second jet, that would be not included in the bjet. This effect may shift the peak on the invariant mass plot to the left and also modify the shape. This FSR moves the event from the right tail to the left tail, increasing the asymmetry of the signal. A simple algorithm is applied to check whether if a gluon jet has been irradiated.

For both the leading bjets a ΔR matching is performed over all the soft jets with quark-gluon-likelihood discriminant [25] less than 0.5. If there is a soft jet close enough ($\Delta R < 0.8$), then this soft jet is considered part of the leading bjet and their four-momentum p^μ are summed.

6 Signal model

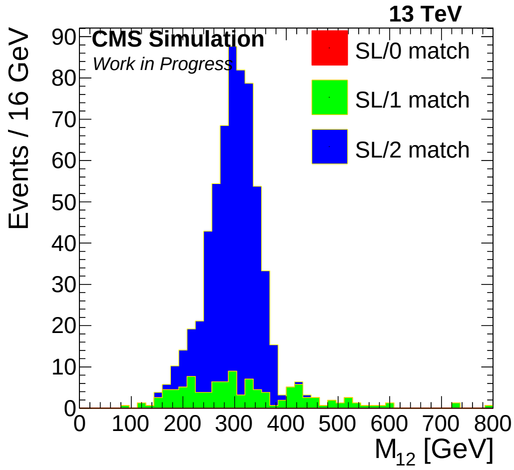
6.1 Correct matching of Higgs daughters

On the Monte Carlo, we can match jets with generator level jets and go back to generator particle level, reconstructing the main vertex and checking if we are correctly choosing the Higgs daughter particles as first two leading jets. In Figure 2 and Figure 3 you can see the invariant mass of the two leading jets for the MC of mass $M_\Phi = 350\,\mathrm{GeV}$. The events are classified by the number of Higgs daughter presents in the two leading jets. As you can see, the majority of events have a correct matching, but the situation is very different for the mass point $M_\Phi = 120\,\mathrm{GeV}$, where there is a complete mismatch, as you can see in Figure 4 and Figure 5. This is mostly due to trigger effects: we expect to see a useful number of events only at masses that are roughly twice the p_T cut we are applying. In the semileptonic case out cut is at $40\,\mathrm{GeV}$, which means that we do not expect correct events at masses lower than $80\,\mathrm{GeV}$. For the full hadronic channel is dramatically worse, because we do not expect events lower than $200\,\mathrm{GeV}$ and our mass point is $120\,\mathrm{GeV}$.

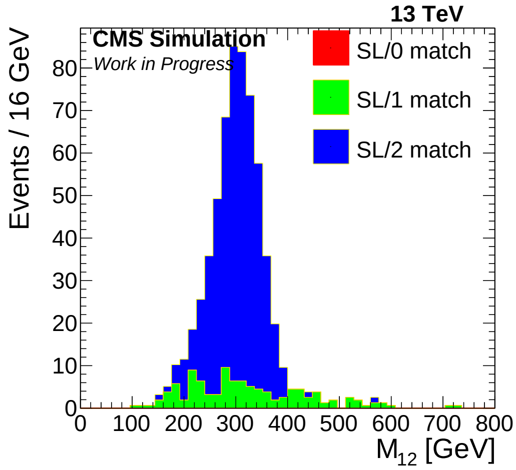
6.2 Application of all corrections

We can qualitatively describe the following effects for the corrections. I will make more quantitative statements in Section 6.4.

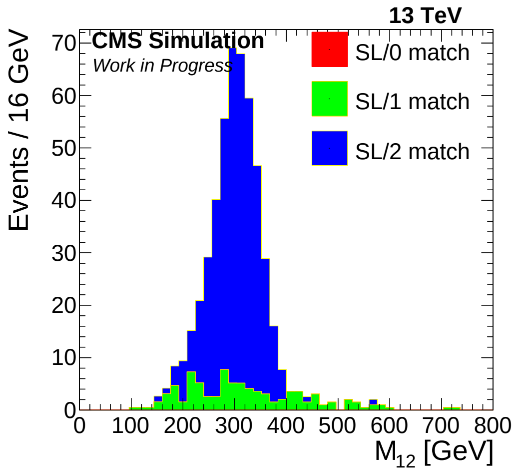
- The smearing of the simulated data widens the shape, making it a bit more symmetric.
- B-Tag scale factors seem not to modify the shape of the curve but only to make a global common scale factor of order 0.5 for each bin.
- Jet energy regression shifts the position of the peak to the right and symmetrizes the shape a bit.
- Final state radiation correction shifts the peak to the right, less than the jet energy regression, but symmetrizes the shape more than the previous one.
- The combination of effects of jet energy regression and FSR seems to keep the positive effects of both.



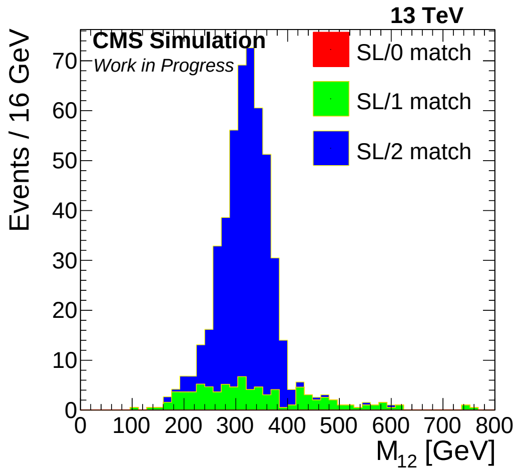
(a) No corrections.



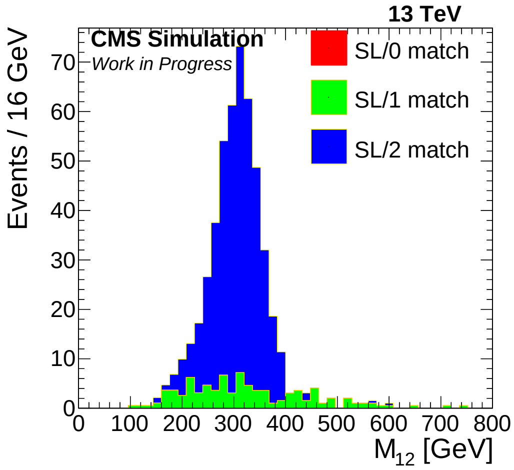
(b) Smearing.



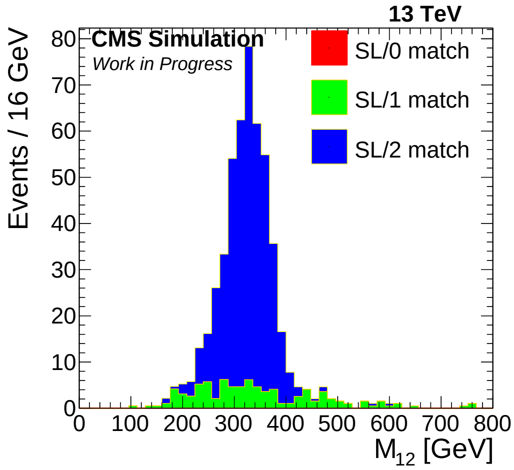
(c) B-Tag scale factors.



(d) Regression.

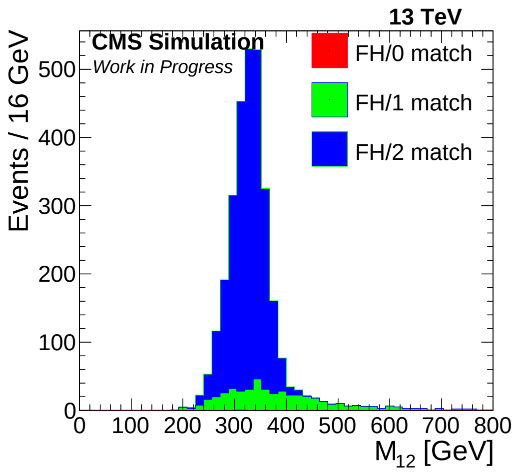


(e) FSR.

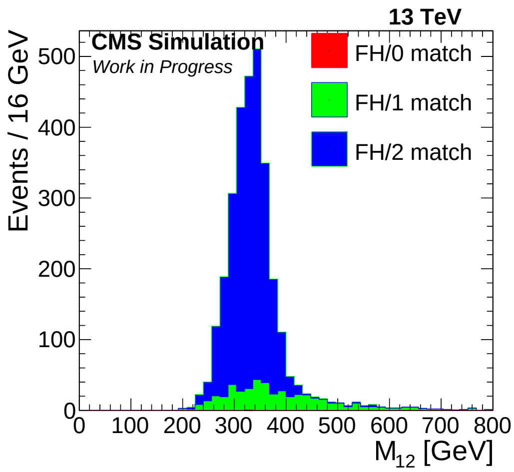


(f) FSR and regression.

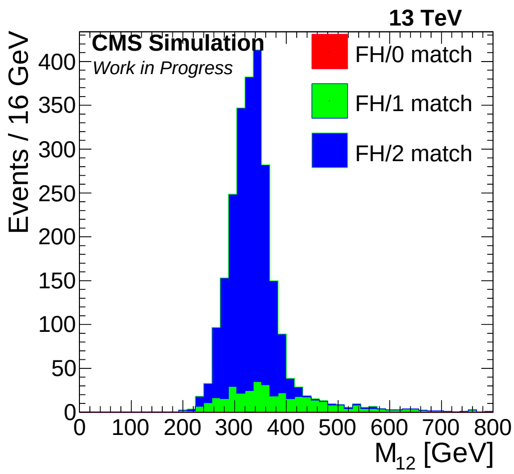
Figure 2: MC for mass point $M_\Phi = 350\,\mathrm{GeV}$. Effect of the corrections on the semileptonic channel.



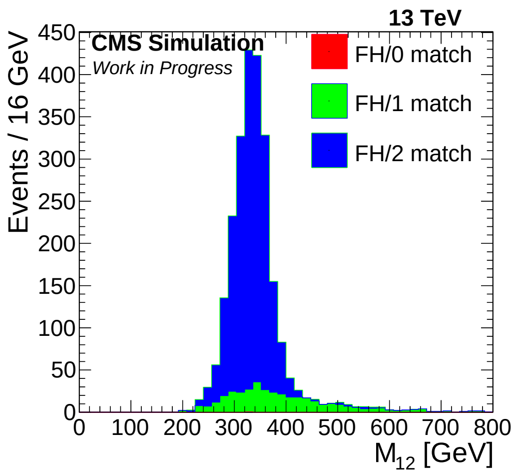
(a) No corrections.



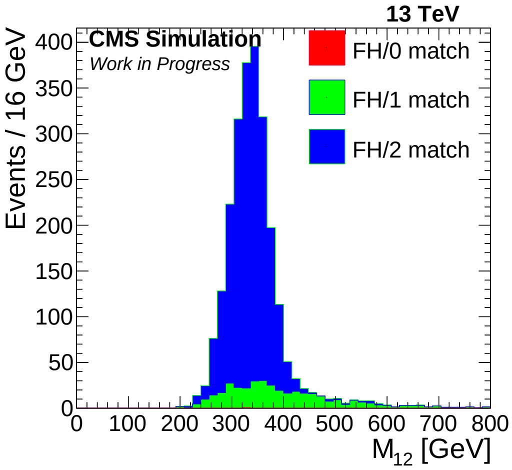
(b) Smearing.



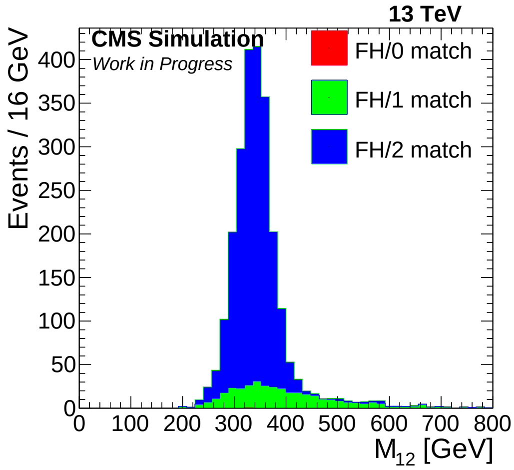
(c) B-Tag scale factors.



(d) Regression.

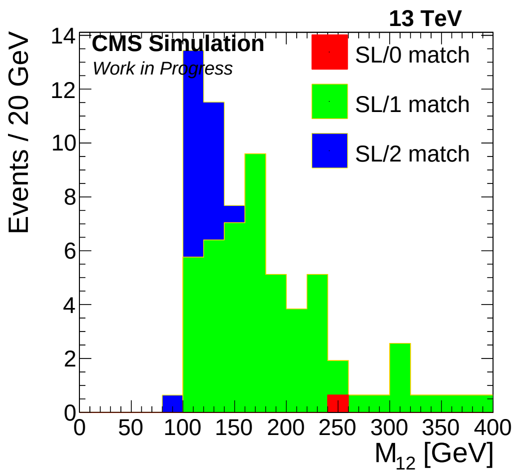


(e) FSR.

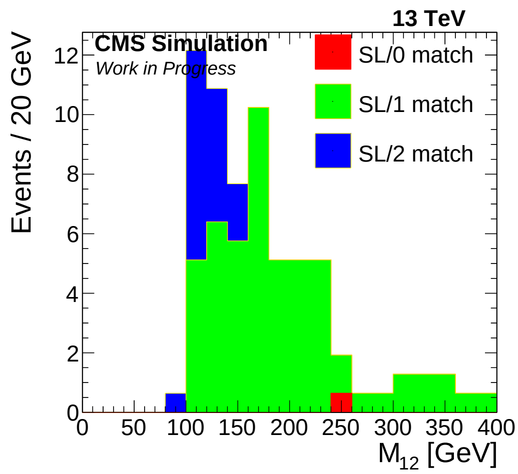


(f) FSR and regression.

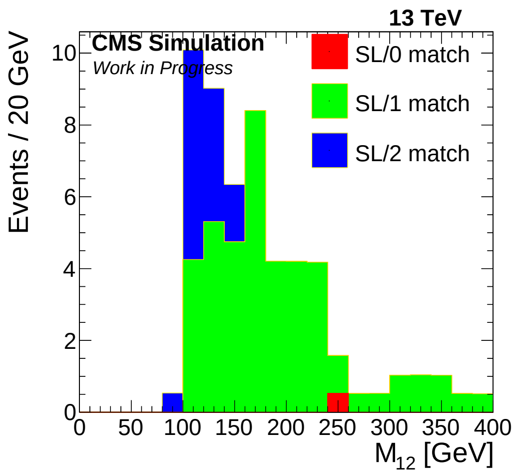
Figure 3: MC for mass point $M_\Phi = 350\,\mathrm{GeV}$. Effect of the corrections on the full hadronic channel.



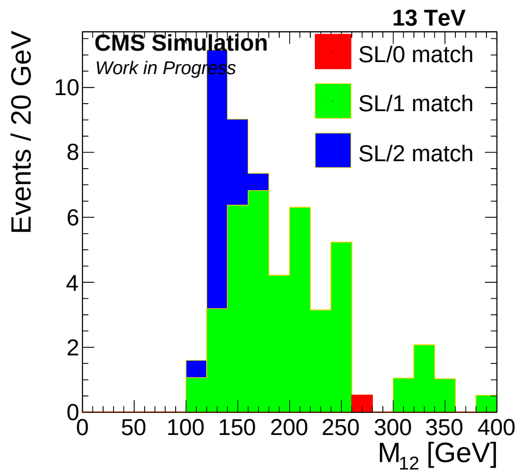
(a) No corrections.



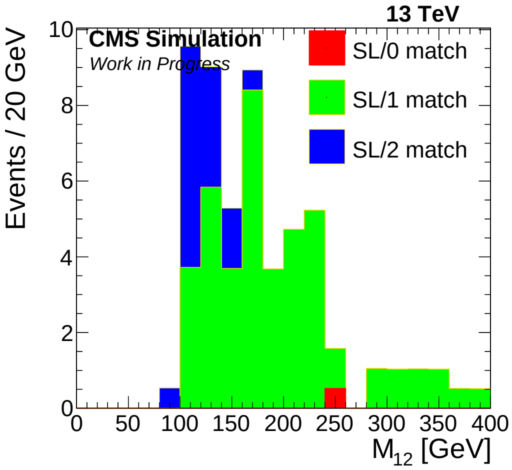
(b) Smearing.



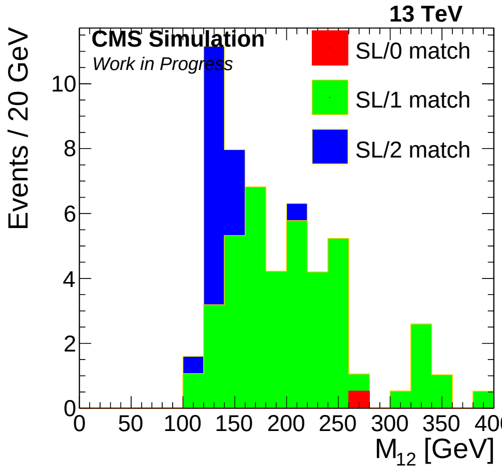
(c) B-Tag scale factors.



(d) Regression.

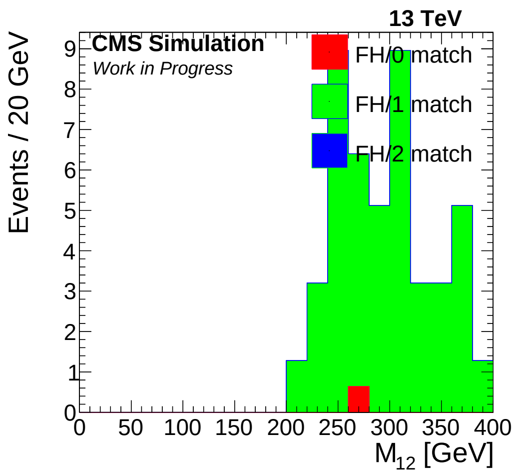


(e) FSR.

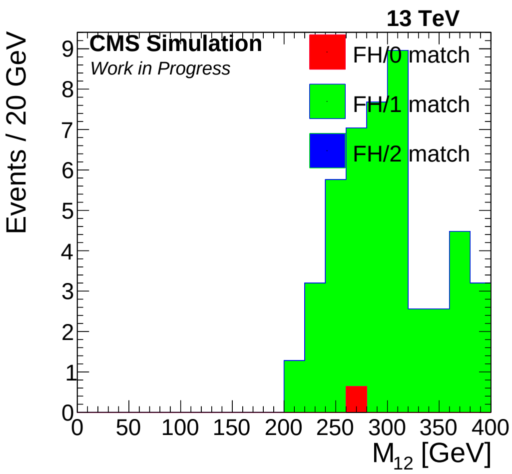


(f) FSR and regression.

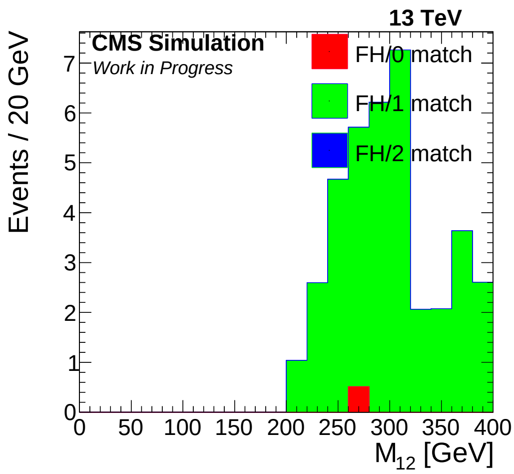
Figure 4: MC for mass point $M_\Phi = 120\,\mathrm{GeV}$. Effect of the corrections on the semileptonic channel.



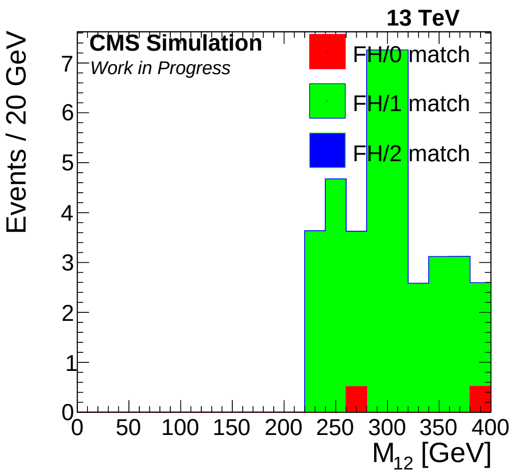
(a) No corrections.



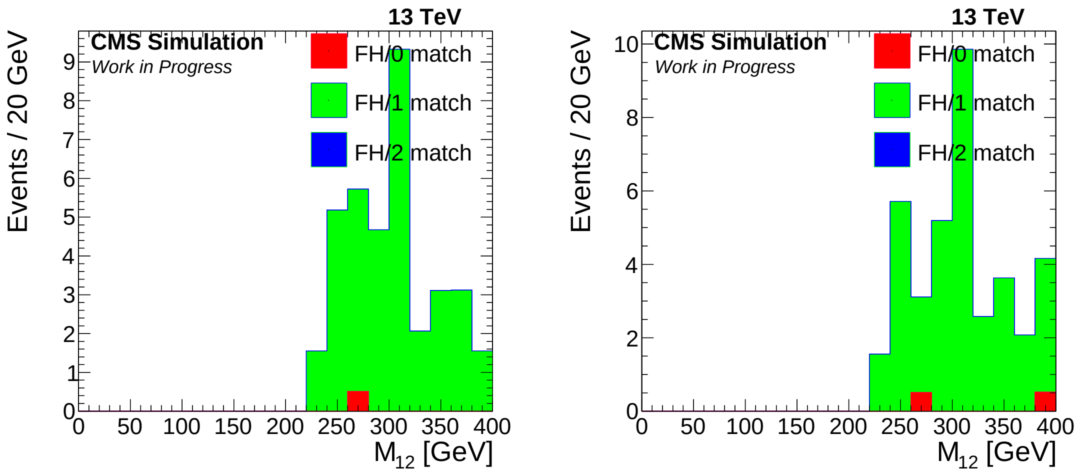
(b) Smearing.



(c) B-Tag scale factors.



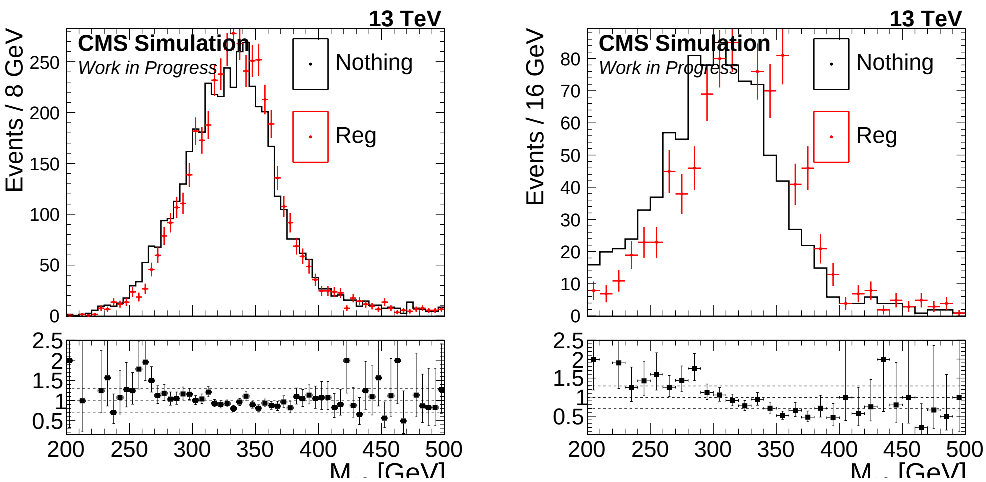
(d) Regression.



(e) FSR. (f) FSR and regression.

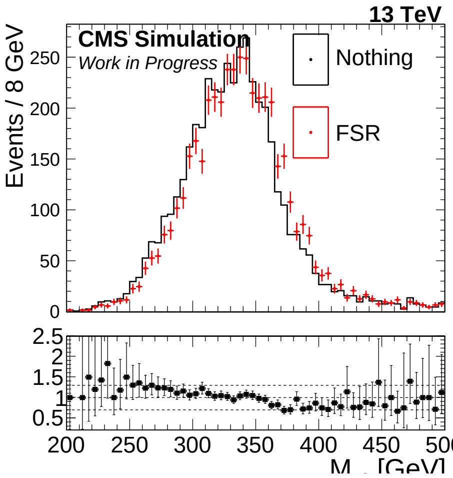
Figure 5: MC for mass point $M_\Phi = 120\,\text{GeV}$. Effect of the corrections on the full hadronic channel.

6.3 Direct comparison of MC before and after corrections

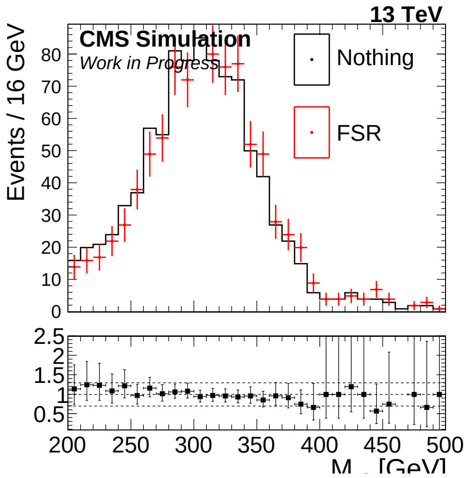


(a) full hadronic channel. (b) semileptonic channel.

Figure 6: ratioplot for mass point $M_\Phi = 350\,\text{GeV}$. Black curve has applied only smearing and b-tag SF. Red one has also jet energy regression.

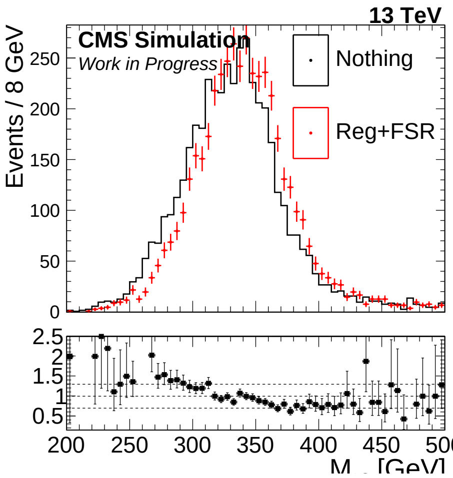


(a) full hadronic channel.

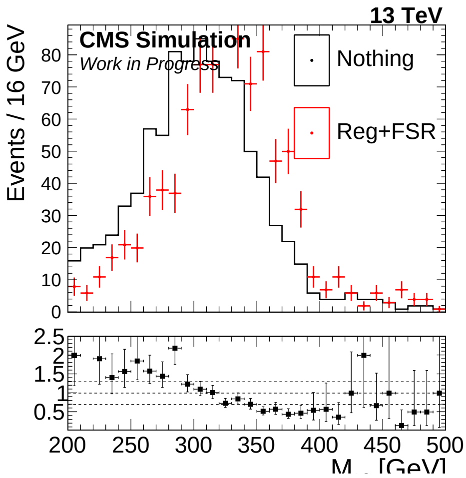


(b) semileptonic channel.

Figure 7: ratioplot for mass point $M_\Phi = 350\,\text{GeV}$. Black curve has applied only smearing and b-tag SF. Red one has also final state radiation.

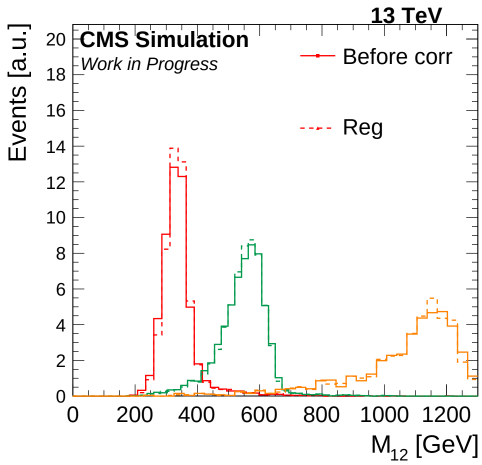


(a) full hadronic channel.

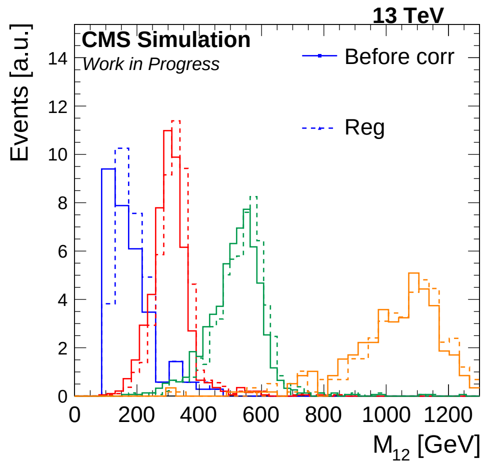


(b) semileptonic channel.

Figure 8: ratioplot for mass point $M_\Phi = 350\,\text{GeV}$. Black curve has applied only smearing and b-tag SF. Red one has also jet energy regression and FSR.

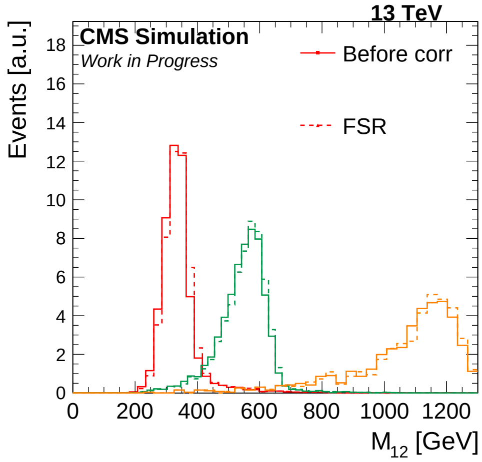


(a) full hadronic channel.

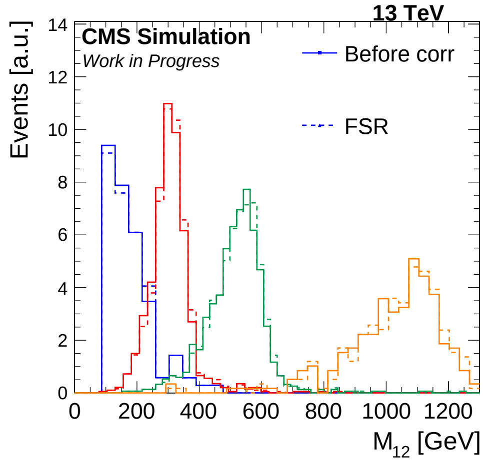


(b) semileptonic channel.

Figure 9: normalized signal before and after jet energy regression only for mass points $M_\Phi = 120\,\mathrm{GeV}$, $M_\Phi = 120\,\mathrm{GeV}$, $M_\Phi = 600\,\mathrm{GeV}$, $M_\Phi = 1200\,\mathrm{GeV}$.

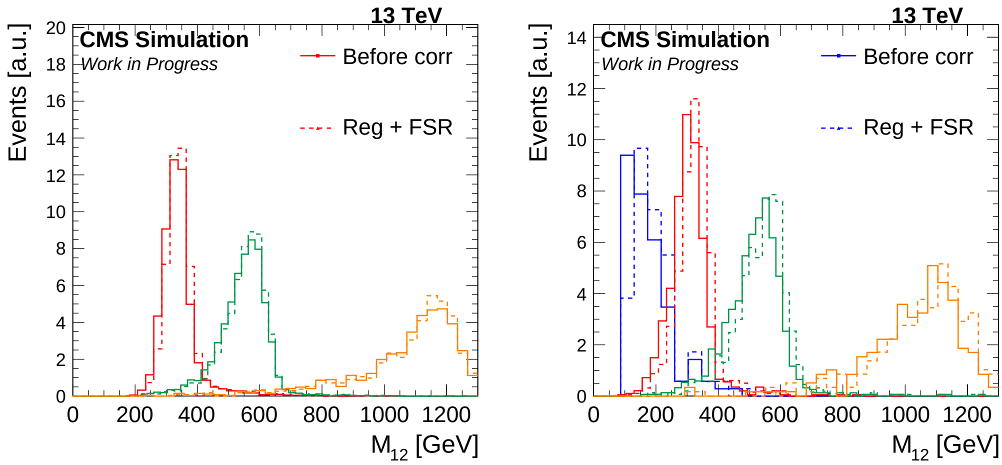


(a) full hadronic channel.



(b) semileptonic channel.

Figure 10: normalized signal before and after final state radiation only for mass points $M_\Phi = 120\,\mathrm{GeV}$, $M_\Phi = 120\,\mathrm{GeV}$, $M_\Phi = 600\,\mathrm{GeV}$, $M_\Phi = 1200\,\mathrm{GeV}$.



(a) full hadronic channel.

(b) semileptonic channel.

Figure 11: normalized signal before and after jet energy regression and final state radiation for mass points $M_\Phi = 120\,\mathrm{GeV}$, $M_\Phi = 120\,\mathrm{GeV}$, $M_\Phi = 600\,\mathrm{GeV}$, $M_\Phi = 1200\,\mathrm{GeV}$.

6.4 Signal shape fit

I decided to try to fit the signal shape for the following two reasons:

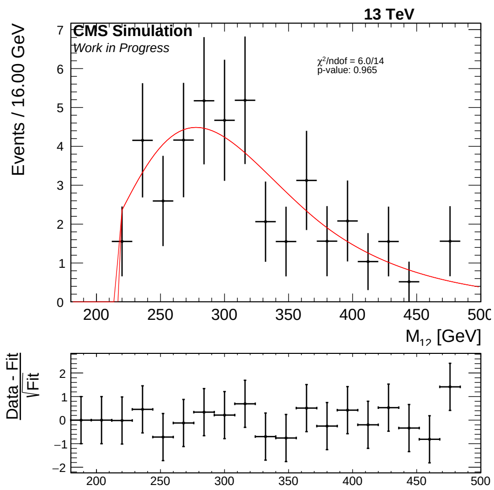
- giving quantitative effects of the corrections.
- use a shape approach for the combine tool to calculate upper limits on cross section per branching ratio of our process.

The shape of our signal is an asymmetric peak. I decided to use the Bukin function, defined in Appendix A

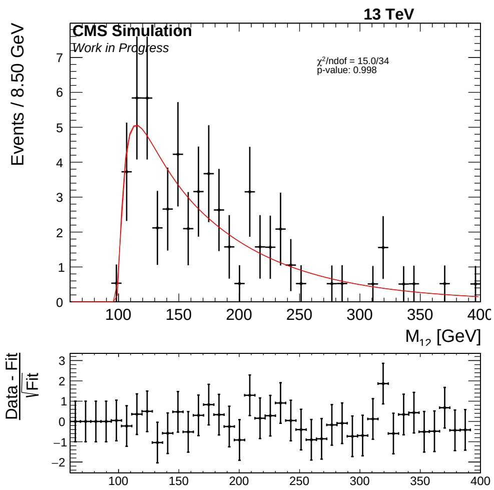
In Table 1 the important parameters of the fitting function are reported for both channels, full hadronic and semileptonic. The parameter x_p has to be interpreted as the position of the peak, σ_p^2 is not the variance of the shape but can be still considered a measure of its width and ξ is an asymmetry parameter. The more it's close to zero, the more symmetric the shape is. We can see the following effects

- The application of jet energy regression moves a lot the position peak towards the expected position in the semileptonic channel ($\approx 5\%$) and improves a lot the resolution for both channels, by a factor of $\approx 10\%$, as was expected from the previous studies [13]. We are also glad to see that the main effect is on the semileptonic channel, as expected.

- Application of FSR correction move slightly the peak to the right but much less than the regression. The main improvement is in the symmetrization of the shape. FSR modifies the ξ parameter closer to zero, which is the symmetric value.
- The combination of both corrections seem to keep all the positive aspects of the two taken individually.

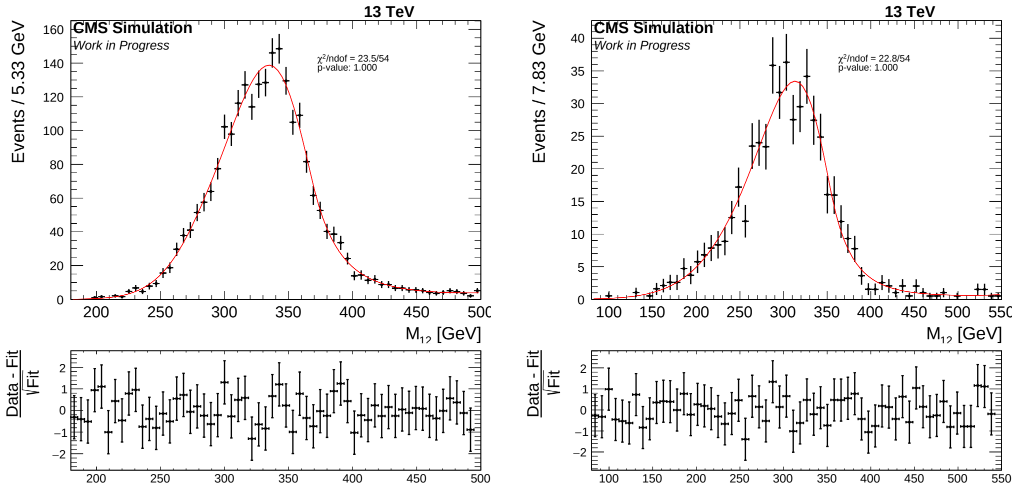


(a) full hadronic channel.



(b) semileptonic channel.

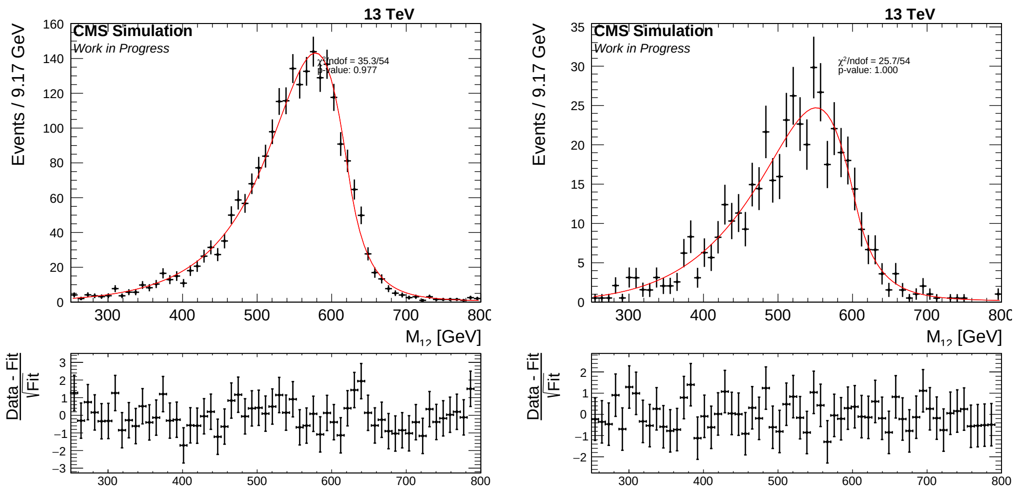
Figure 12: fit of MC for mass point $M_\Phi = 120\,\text{GeV}$. Smearing and b-tag scale factors applied.



(a) full hadronic channel.

(b) semileptonic channel.

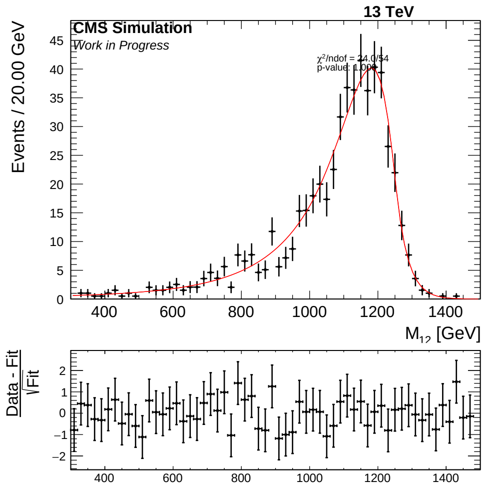
Figure 13: fit of MC for mass point $M_\Phi = 350\,\mathrm{GeV}$. Smearing and b-tag scale factors applied.



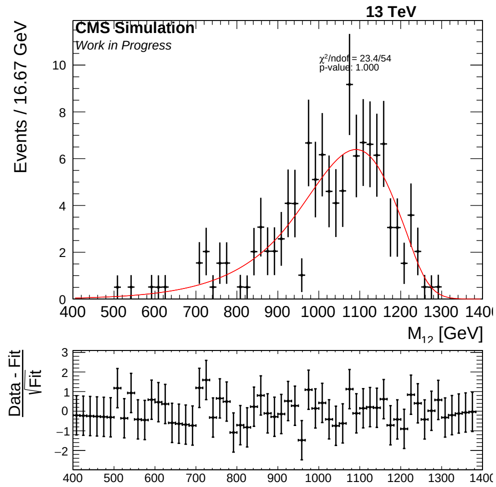
(a) full hadronic channel.

(b) semileptonic channel.

Figure 14: fit of MC for mass point $M_\Phi = 600\,\mathrm{GeV}$. Smearing and b-tag scale factors applied.



(a) full hadronic channel.



(b) semileptonic channel.

Figure 15: fit of MC for mass point $M_\Phi = 1200\,\mathrm{GeV}$. Smearing and b-tag scale factors applied.

Correction	x_p [GeV]	σ_p [GeV]	Asymmetry ξ
Smearing + B-Tag	335 ± 2	34 ± 1	-0.14 ± 0.03
Reg	337 ± 1	30.9 ± 0.9	-0.14 ± 0.03
FSR	339 ± 2	34 ± 1	-0.12 ± 0.03
Reg + FSR	341 ± 2	31.5 ± 0.9	-0.09 ± 0.03

(a) full hadronic.

Correction	x_p [GeV]	σ_p [GeV]	Asymmetry ξ
Smearing + B-Tag	313 ± 4	43 ± 3	-0.20 ± 0.06
Reg	330 ± 4	40 ± 3	-0.19 ± 0.06
FSR	316 ± 4	42 ± 3	-0.17 ± 0.06
Reg + FSR	333 ± 4	39 ± 3	-0.16 ± 0.06

(b) semileptonic.

Table 1: important parameters of fitting function for $M_\Phi = 350\,\mathrm{GeV}$.

7 Background model

7.1 Corrections effect on the control region

Corrections have to be applied also to the control region. The dataset corresponds to an integrated luminosity of $35.6\,\mathrm{fb}^{-1}$ [26] at a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ taken in 2017 at the LHC. In Figure 16, Figure 17 and Figure 18 you

Correction	x_p [GeV]	σ_p [GeV]	Asymmetry ξ
Smearing + B-Tag	578 ± 3	52 ± 2	-0.28 ± 0.04
Reg	578 ± 3	49 ± 1	-0.27 ± 0.03
FSR	584 ± 3	51 ± 2	-0.28 ± 0.03
Ref + FSR	583 ± 3	48 ± 1	-0.26 ± 0.03

(a) full hadronic.

Correction	x_p [GeV]	σ_p [GeV]	Asymmetry ξ
Smearing + B-Tag	551 ± 7	63 ± 5	-0.30 ± 0.08
Reg	569 ± 7	58 ± 4	-0.29 ± 0.08
FSR	555 ± 7	63 ± 5	-0.29 ± 0.08
Reg + FSR	574 ± 6	57 ± 4	-0.30 ± 0.07

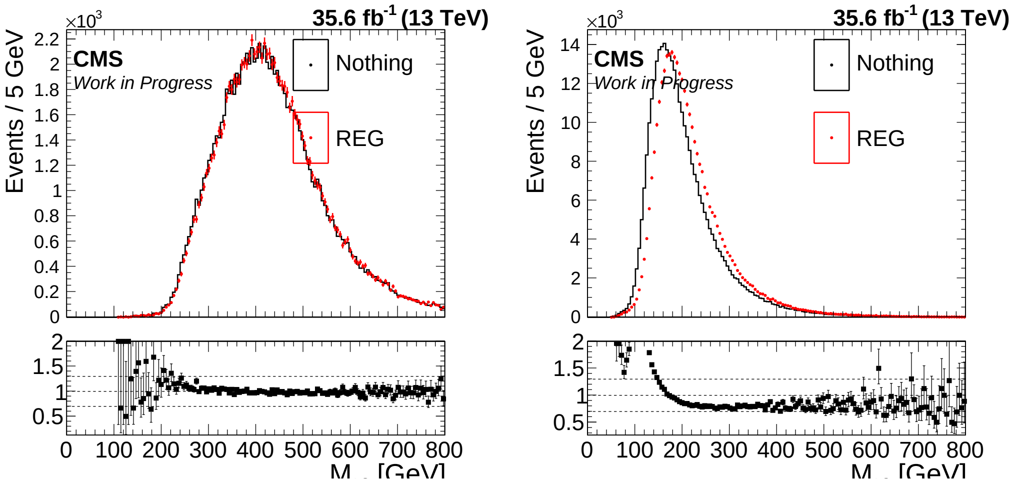
(b) semileptonic.

Table 2: important parameters of fitting function for $M_\Phi = 600\,\text{GeV}$.

can see the comparison between the same channel with and without correction.

As for the signal region, we see that the most of the impact of the jet energy regression is only on the semileptonic channel, where the difference is really visible, while in the full hadronic one there is only a little shift to the right of the peak. On the other side, the FSR correction seems to affect more the full hadronic channel than the semileptonic one.

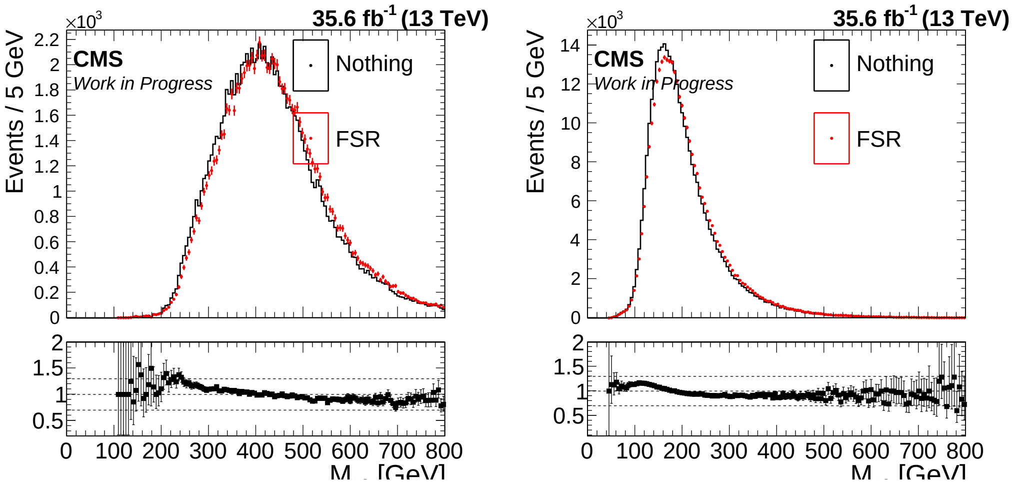
The effects of both corrections seem to factorize, meaning that they sum without interfering.



(a) full hadronic channel.

(b) semileptonic channel.

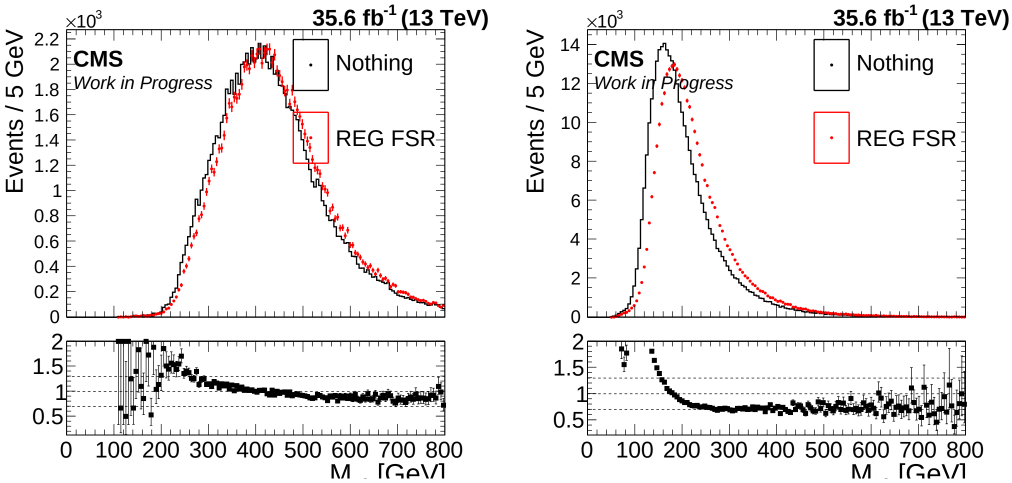
Figure 16: ratioplot. Black curve has no correction applied. Red one has jet energy regression.



(a) full hadronic channel.

(b) semileptonic channel.

Figure 17: ratioplot. Black curve has no correction applied. Red one has final state radiation.



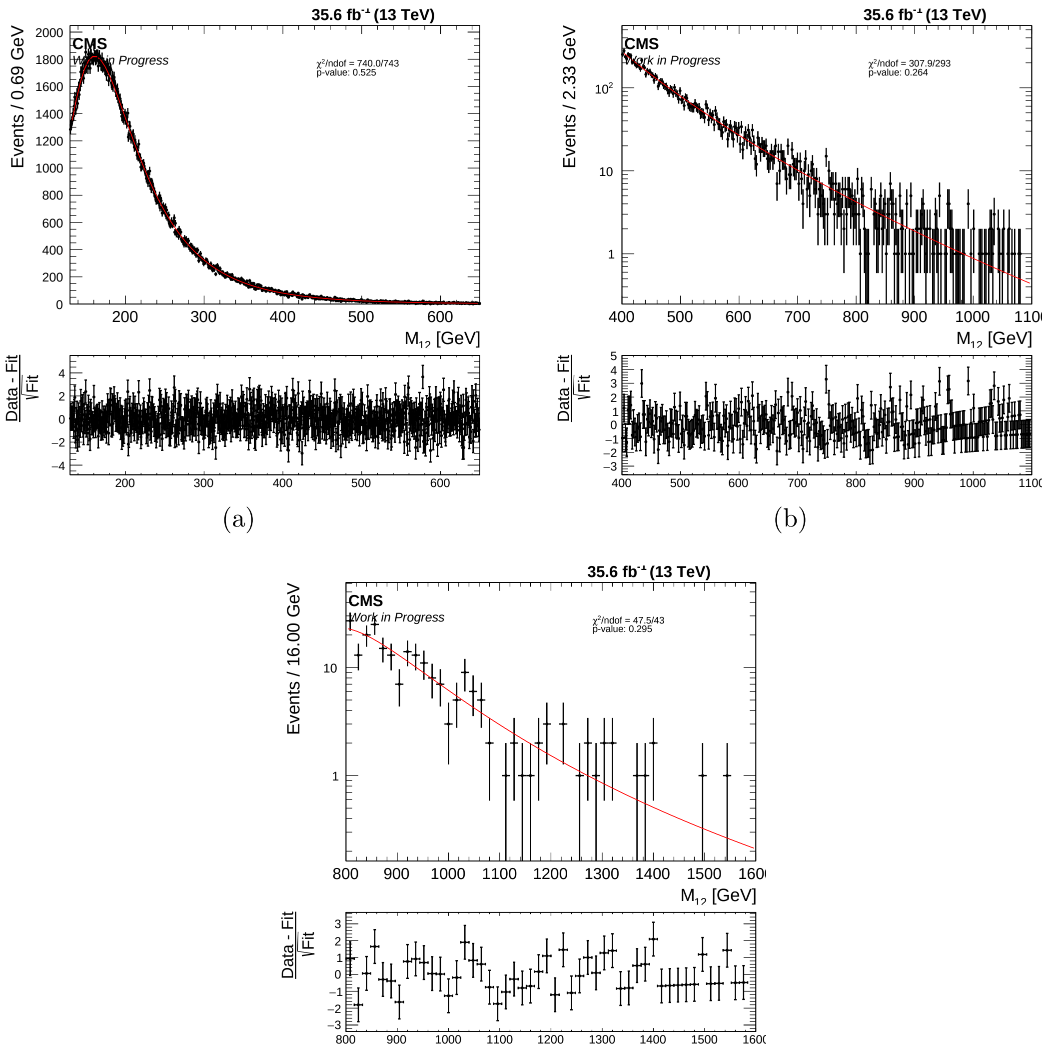
(a) full hadronic channel.

(b) semileptonic channel.

Figure 18: ratioplot. Black curve has applied only smearing and b-tag SF. Red one has also jet energy regression and FSR.

7.2 Fitting the control region

For the leptonic case, I used the function defined in Equation 3 in Appendix B. This function is an extension of the Novosibirsk function originally used to describe a Compton spectrum. I was inspired by the article in Ref. [8] for this choice.



(c)

Figure 19: fit of the semileptonic channel control region for the 3 ranges with no correction.

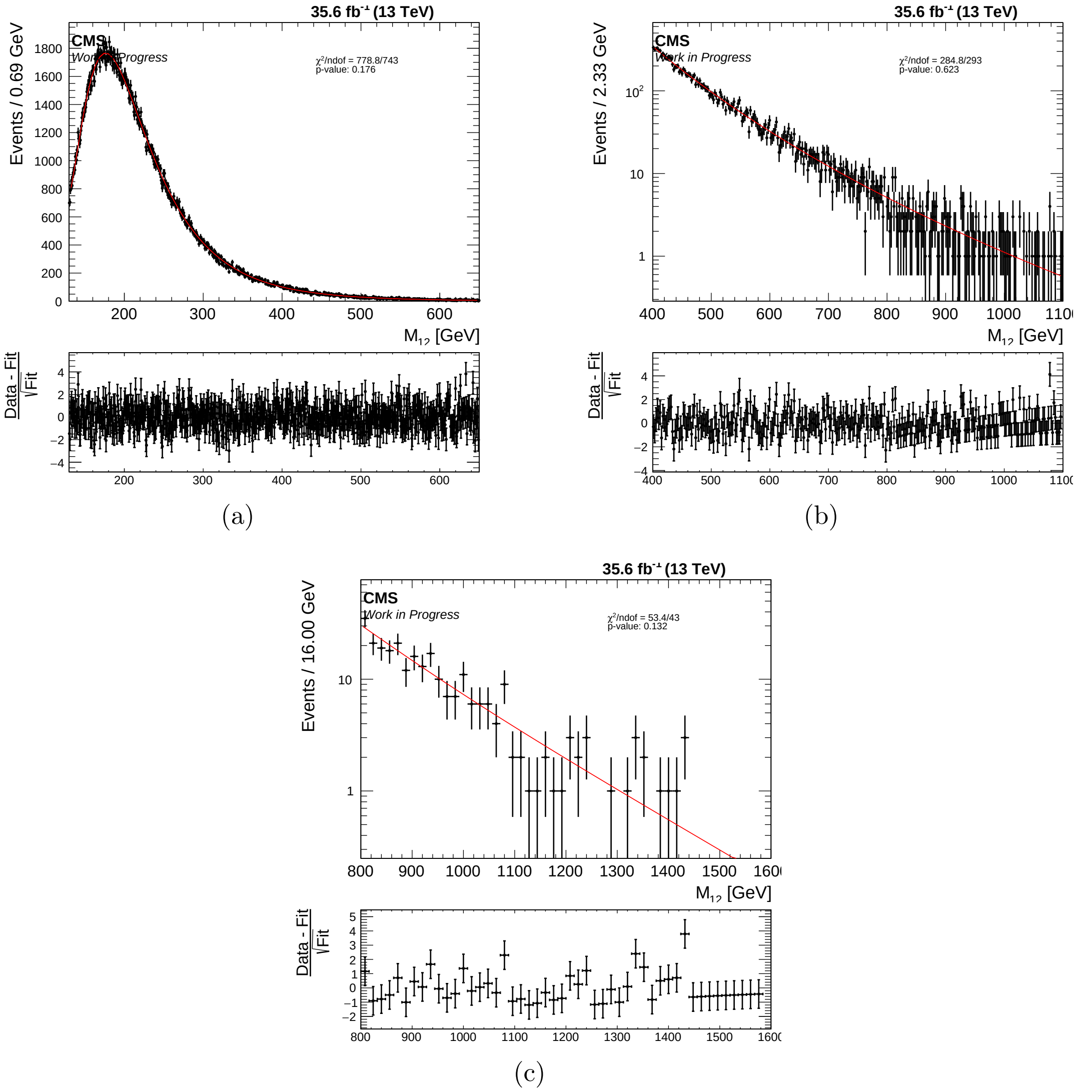
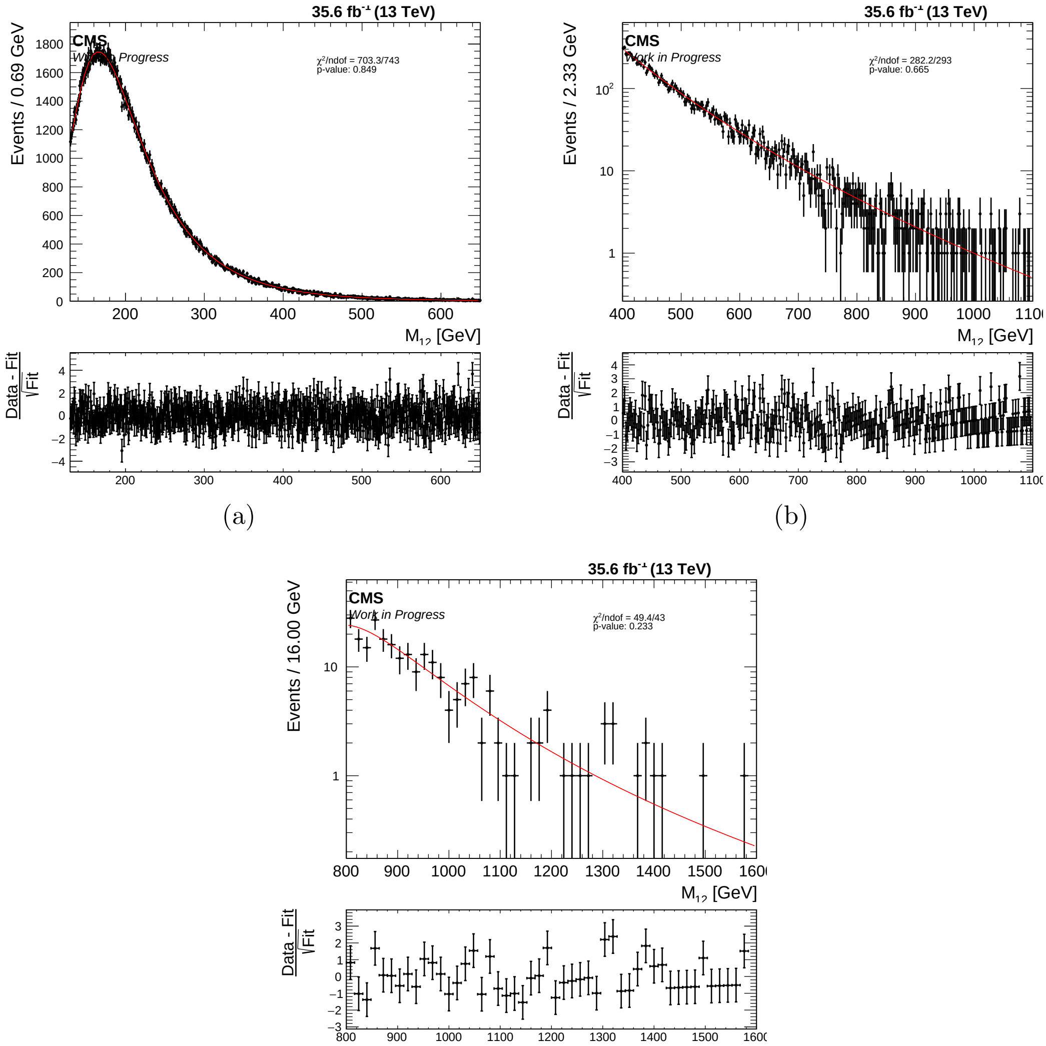


Figure 20: fit of the semileptonic channel control region for the 3 ranges with jet energy regression only.



(a) (b) (c)

Figure 21: fit of the semileptonic channel control region for the 3 ranges with FSR correction only.

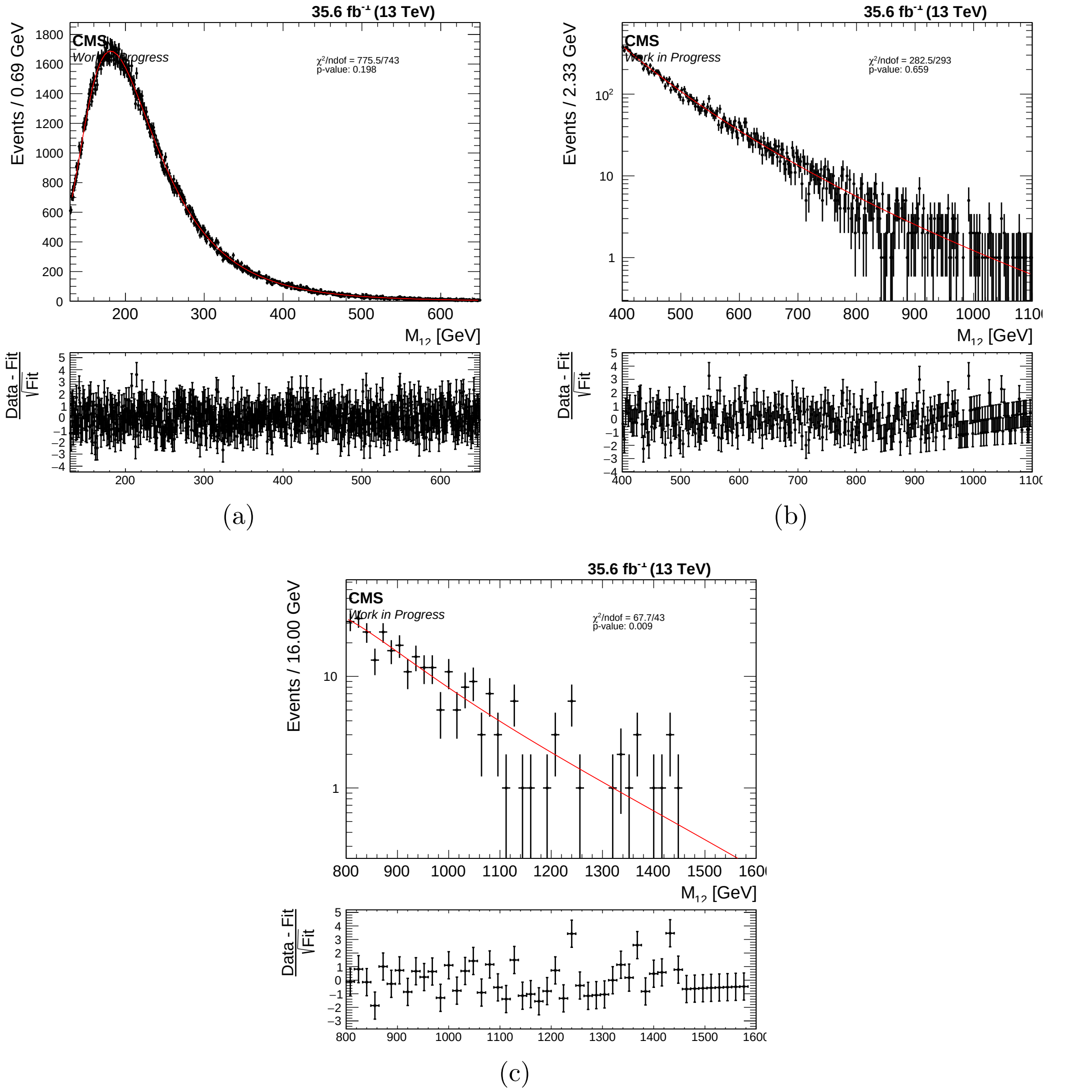


Figure 22: fit of the semileptonic channel control region for the 3 ranges with regression and FSR corrections.

8 Cross section upper limits

8.1 Template approach

The CombinedLimit software has been used to calculate our sensitivity. In the first instance, I used a simple template approach, providing histograms. Our aim is to maximize the Signal/$\sqrt{\text{Bkg}}$, so I show here in the following plots the limits we can infer on the cross-section σ multiplied by the branching ratio after each correction.

For this simple analysis, I took into account only one systematic uncertainty, on the total integrated luminosity [26], of 2.3% because our aim was only to compare the results with and without all our corrections. We expect to obtain much wider error bands and also the medium value can be rigidly shifted to higher cross sections, but this shouldn't affect the difference between the limits with and without the corrections.

In Figure 23 and Figure 24 we can see the real effect of these corrections on our sensitivity.

- Semileptonic channel improves in the low mass region up to 15% with the regression correction.
- FSR correction neither seems to improve our sensitivity for the semileptonic channel nor for the full hadronic one, but the hope is not lost because it still helps to symmetrize the shape of our signal, so we hope that with a shape approach our results will improve.
- The combination of both effects seems to preserve the benefits of the regression.

I also decided to make a comparison between the limits with and without the fourth jet veto, written in Table 3, Table 4, Table 5 and Table 6. In all these situations except for the semileptonic channel at the mass point $M_\Phi = 350\,\text{GeV}$, there is an improvement, which reaches $\approx 25\%$, that justifies our decision to apply also this kind of kinematic cut. In very rare cases, like in Table 6, our sensitivity worsen after all the corrections, but this may be due to the presence of a very small background in this mass regime. Further investigation is required for this high-mass range.

Correction level	-1σ [pb]	Mean [pb]	$+1\sigma$ [pb]
Nominal	51.686	71.875	100.238
Reg	41.349	57.5	80.191
FSR	52.675	73.25	102.156
Reg + FSR	41.169	57.25	79.614

(a) without fourth jet veto.

Correction level	-1σ [pb]	Mean [pb]	$+1\sigma$ [pb]
Nominal	49.799	69.25	96.577
Reg	40.63	56.5	78.796
FSR	50.698	70.5	98.321
Reg + FSR	39.462	54.875	76.748

(b) with fourth jet veto.

Table 3: limits on $\sigma \times \mathrm{BR}$ for semileptonic channel, mass point $M_\Phi = 120\,\mathrm{GeV}$

Correction level	-1σ [pb]	Mean [pb]	$+1\sigma$ [pb]
Nominal	2.955	4.109	5.731
Reg	2.88	3.984	5.573
FSR	3.023	4.203	5.862
Reg + FSR	2.845	4.030	5.621

(a) without fourth jet veto.

Correction level	-1σ [pb]	Mean [pb]	$+1\sigma$ [pb]
Nominal	3.157	4.391	6.123
Reg	2.849	4.016	5.664
FSR	3.202	4.453	6.21
Reg + FSR	3.079	4.281	5.988

(b) with fourth jet veto.

Table 4: limits on $\sigma \times \mathrm{BR}$ for semileptonic channel, mass point $M_\Phi = 350\,\mathrm{GeV}$

Correction level	-1σ [pb]	Mean [pb]	$+1\sigma$ [pb]
Nominal	0.716	0.996	1.385
Reg	0.699	0.973	1.356
FSR	0.699	0.973	1.356
Reg + FSR	0.683	0.949	1.324

(a) without fourth jet veto.

Correction level	-1σ [pb]	Mean [pb]	$+1\sigma$ [pb]
Nominal	0.508	0.711	0.997
Reg	0.503	0.699	0.975
FSR	0.508	0.707	0.989
Reg + FSR	0.497	0.691	0.967

(b) with fourth jet veto.

Table 5: limits on $\sigma \times \mathrm{BR}$ for full hadronic channel, mass point $M_\Phi = 350\,\mathrm{GeV}$.

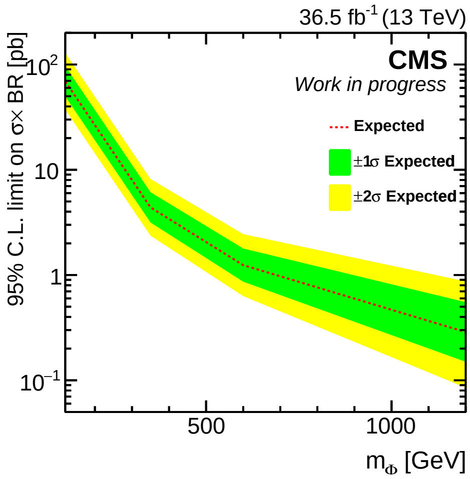
Correction level	-1σ [pb]	Mean [pb]	$+1\sigma$ [pb]
Nominal	0.075	0.139	0.253
Reg	0.066	0.123	0.226
FSR	0.064	0.122	0.227
Reg + FSR	0.097	0.176	0.309

(a) without fourth jet veto.

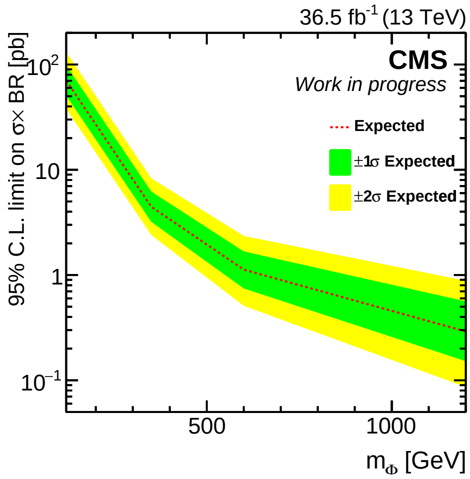
Correction level	-1σ [pb]	Mean [pb]	$+1\sigma$ [pb]
Nominal	0.06	0.115	0.22
Reg	0.063	0.119	0.226
FSR	0.068	0.129	0.245
Reg + FSR	0.079	0.148	0.28

(b) with fourth jet veto.

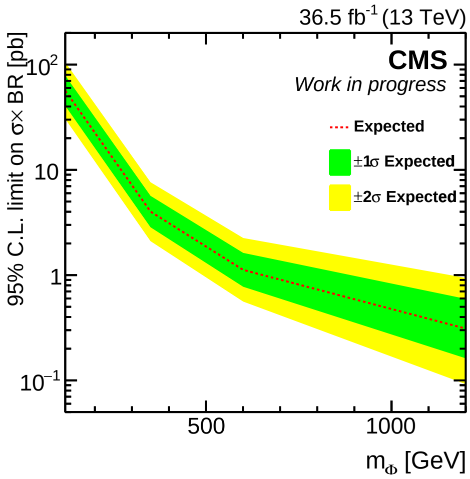
Table 6: limits on $\sigma \times \mathrm{BR}$ for full hadronic channel, mass point $M_\Phi = 1200\,\mathrm{GeV}$.



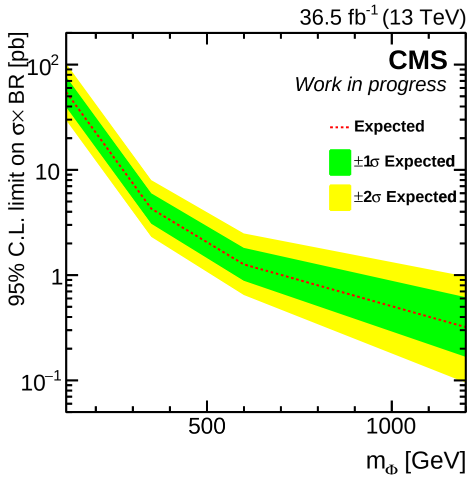
(a) No correction.



(b) Only FSR correction.

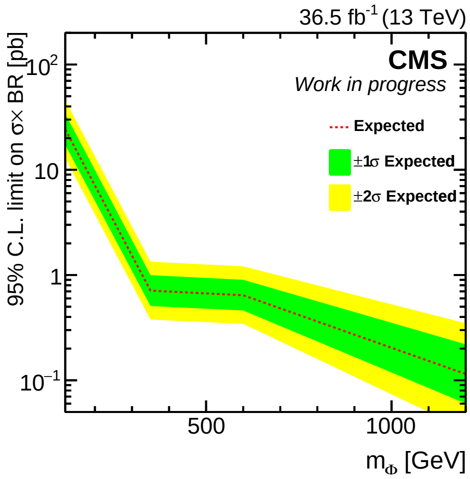


(c) Only jet energy regression correction.

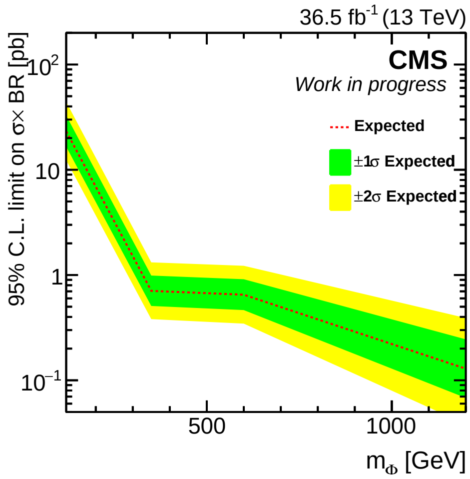


(d) Both FSR and JER correction.

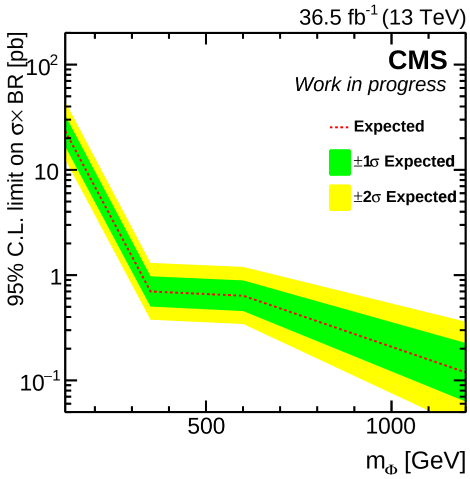
Figure 23: limits on $\sigma \times$ BR for semileptonic channel.



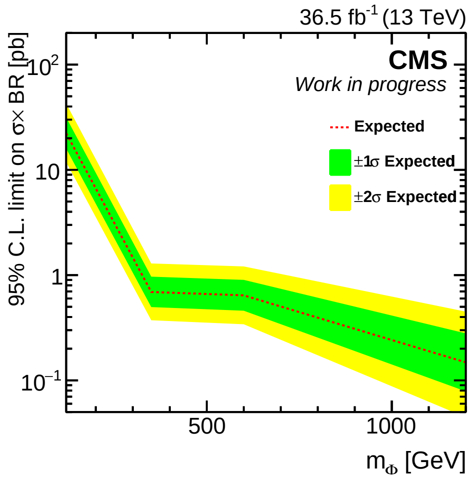
(a) No correction.



(b) Only FSR correction.



(c) Only jet energy regression correction.



(d) Both FSR and JER correction.

Figure 24: limits on $\sigma \times$ BR for full hadronic channel.

8.2 Shape approach

With the fits obtained in Section 6.4 and Section 7.2 we can try a different approach to obtain the upper limits on the cross-section. As before, I took into account only a few systematic uncertainties for this study because our aim is to compare the results before and after the corrections, not to obtain an absolute value. The uncertainties were

- uncertainty of 2.3% on the total integrated luminosity.

- uncertainty of 5% on the online b-tag.

unfortunately, this is still a work in progress so I decided not to include these plots in this report because I first want to check the correctness of them.

9 Summary

A study of the effect of jet energy regression and final state radiation correction has been performed. The data analyzed correspond to an integrated luminosity of $35.6\,\mathrm{fb}^{-1}$, recorded in proton-proton collisions at center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ at the LHC. The first correction seems to be very effective in the semileptonic channel, moving the peak in the control region towards the expected position by a factor of 5%. For the full hadronic channel, the effect is much less evident but we can still claim that this moves the peak to the right place and improves the jet energy resolution up to 10%. Final state radiation correction symmetrizes the shape in the control region and moves slightly the peak to the expected mass. The former correction improves a lot our sensitivity in cross section in the lower mass regime, from $\approx 120\,\mathrm{GeV}$ to $\approx 350\,\mathrm{GeV}$ and the latter one does not modify significantly the situation.

Acknowledgments

I would like to thank all the people from my hostel's kitchen for the great time I spent with them and my supervisors Antonio and Roberval for having constantly followed my work and for their helpful advice.

Appendices

A Definition of Bukin function

The Bukin function is defined as:

$$f(M_{12}) = A_p \exp\left[-\ln 2 \frac{\ln^2\left(1+\sqrt{2}\xi\sqrt{1+\xi^2}\frac{M_{12}-x_p}{\sqrt{\ln 2}\sigma_p}\right)}{\ln^2\left(1+2\xi(\xi-\sqrt{1+\xi^2}\right)}\right]$$

if $x_1 < M_{12} < x_2$,

$$f(M_{12}) = \frac{A_p}{2} \exp\left[\pm \frac{\xi\sqrt{1+\xi^2}(M_{12}-x_i)\sqrt{2\ln 2}}{\sigma_p \ln(\sqrt{1+\xi^2}+\xi)\left(\sqrt{1+\xi^2}\mp\xi\right)^2} + \rho_i \left(\frac{M_{12}-x_i}{x_p - x_i}\right)^2\right]$$

else. The parameters x_p, σ_p, ρ_1, ρ_2, ξ are free parameters, $\rho_i = \rho_1$, $x_i = x_1$ if $M_{12} < x_1$ and $\rho_i = \rho_2$, $x_i = x_2$ if $M_{12} > x_2$, where $x_{1,2}$ are defined as below:

$$x_{1,2} = x_p + \sigma_p\sqrt{2\ln 2}\left(\frac{\xi}{\sqrt{1+\xi^2}} \mp 1\right)$$

B Definition of extended Novosibirsk function

The extended Novosibirsk function is defined as the product of two different functions

$$p(M_{12}) = f(M_{12}) \cdot g(M_{12}) \tag{3}$$

where $f(M_{12})$ is an activation function defined as below

$$f(M_{12}) = \frac{1}{2}\left(\text{erf}[p_0(M_{12}-p_1)]+1\right)$$

and $g(M_{12})$ as follows

$$g(M_{12}) = p_2 \exp\left(-\frac{1}{2\sigma_0^2}\ln^2\left[1-\frac{M_{12}-p_3}{p_4}p_5 - \frac{(M_{12}-p_3)^2}{p_4}p_5 p_6\right] - \frac{\sigma_0^2}{2}\right)$$

where p_2 is a normalization parameter, p_3 the peak value of the distribution, p_4 and p_5 are the parameters describing the asymmetry of the spectrum, and p_6 is the parameter of the extended term. The variable σ_0 is defined as:

$$\sigma_0 = \frac{2}{\xi}\sinh^{-1}(p_5\xi/2), \quad \text{where } \xi = 2\sqrt{\ln 4}$$

and the erf function used earlier is obviously

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\mathrm{d}t$$

References

[1] CMS Collaboration. Observation of Higgs boson decay to bottom quarks. *Phys. Ref. Lett.*, 2018. arXiv:1808.08242.

[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. doi:10.1016/j.physletb.2012.08.021. arXiv:1207.7235.

[3] 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. doi:10.1016/j.physletb.2012.08.020. arXiv:1207.7214.

[4] 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*, 2013. doi:10.1007/JHEP06(2013)081. arXiv:1303.4571.

[5] CMS Collaboration. Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP 11 (2017) 047*, 2017. doi:10.1007/JHEP11(2017)047. arXiv:1706.09936.

[6] CMS Collaboration. Search for a Higgs boson decaying into a b-quark pair and produced in association with b quarks in proton–proton collisions at 7 TeV. *Phys. Lett. B*, 722(4):207 – 232, 2013. doi:10.1016/j.physletb.2013.04.017. arXiv:1302.2892.

[7] CMS Collaboration. Search for a light pseudoscalar Higgs boson produced in association with bottom quarks in pp collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 11(010), 2017. doi:10.1007/JHEP11(2017)010. arXiv:1707.07283.

[8] CMS Collaboration. Search for beyond the standard model Higgs bosons decaying into a $b\overline{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 1808(113), 2018. doi:10.1007/JHEP08(2018)113. arXiv:1805.12191.

[9] G. C. Branco et al. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rep. 516*, 2012. doi:10.1016/j.physrep.2012.02.002. arXiv:1106.0034.

[10] H. P. Nilles. Supersymmetry, supergravity and particle physics. *Phys. Rep. 110*, 1984. doi:10.1016/0370-1573(84)90008-5.

[11] CMS Collaboration. Jet algorithms performance in 13 TeV data. *CMS Physics Analysis Summary CMS-PAS-JME-16-003*, 2016.

[12] Vernon Barger, Muneyuki Ishida, Wai-Yee Keung. Total Width of 125 GeV Higgs Boson. *Phys. Rev. Lett. 108, 261801 (2012)*, 2012. doi:10.1103/PhysRevLett.108.261801.

[13] CMS Collaboration. Deep neural network based simultaneous b-jet energy correction and resolution estimator. *CMS notes*, 2017.

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

[15] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST 3 S08003*, 2008. doi:10.1088/1748-0221/3/08/S08003.

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

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

[18] 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.

[19] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP 04 (2008) 063*, 2008. doi:10.1088/1126-6708/2008/04/063. arXiv:0802.1189.

[20] CMS Collaboration. Pileup removal algorithms. *CMS Physics Analysis Summary CMS-PAS-JME-14-001*, 2014.

[21] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659(119), 2008. doi:10.1016/j.physletb.2007.09.077. arXiv:0707.1378.

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

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

[24] D. Guest et al. Jet flavor classification in high-energy physics with deep neural networks. *Phys. Rev. D 94 (2016) 112002*, 2016. doi:10.1103/PhysRevD.94.112002.

[25] CMS Collaboration. Quark-gluon Jet Discrimination At CMS. arXiv:1409.3072.

[26] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13\,\mathrm{TeV}$. *CMS Physics Analysis Summary CMS-PAS-JME-17-004*, 2017.