University of California
Santa Barbara

Constraints on the Higgs boson decay width from off-shell production decay into Z-boson Pair in simulated data

A dissertation submitted in partial satisfaction
of the requirements for the degree

Bachelor of Science
in
Physics

by

Jiahong (Jerry) Ling | 凌嘉鸿

Committee in charge:

Professor Claudio Campagnari, Chair
Doctor Sathya Guruswamy

June 2020

The Dissertation of Jiahong (Jerry) Ling | 凌嘉鸿 is approved.

Doctor Sathya Guruswamy

Professor Claudio Campagnari, Committee Chair

May 2020

Constraints on the Higgs boson decay width from off-shell production decay into Z-boson Pair in simulated data

by

Jiahong (Jerry) Ling | 凌嘉鸿

Acknowledgements

The research presented in this senior thesis would not been possible if not for the many people's guidance, encouragement, and help. Foremost, I would like to thank my research adviser Dr. Claudio Campagnari for his constant support and mentorship, both in research and my student life in the past almost 3 years. He lead me into the field of High Energy physics and had provided personal attention to my progress along the way. He is also the one who suggested the research topic to me as a step up from what I have been doing.

At the same time, I would also like to thank Dr. Ulascan Sarica, who is the primary correspondent in physics analysis for this research at UCSB. It would not be possible for me to undertake such a complex analysis if not for his dedication and careful teaching.

Additionally, I would like to acknowledge the help from Nick Amin, who guided me personally in the summer of 2017 when I first joined the group. Nick has also helped me tremendously in both physics and technical issues and have provided many useful tools for physics and computing. In the same regard, I would like to thank Bennett Marsh and Sicheng Wang for their help and guidance in the past years.

Abstract

Constraints on the Higgs boson decay width from off-shell production decay into Z-boson Pair in simulated data

by

Jiahong (Jerry) Ling | 凌嘉鸿

The discovery of the Standard Model (SM) Higgs boson at the Large Hadron Collider (LHC) was a major achievement of the experimental particle physics community in the 21st century. Though a fair portion of the physics analysis focus has shifted to Supersymmetry (SUSY) physics, the direct search of SUSY models has yield null result so far. Meanwhile, the many properties of Higgs are still to be measured.

In this analysis, we present constraints on the decay width of Higgs boson, Γ_{H}, by using the on-shell and off-shell decay rates of Higgs to a pair of Z bosons and both Z's decay to a pair of electrons or muons. A total number of 2 (ee or $\mu\mu$) $\times$ 4 ($N_{\mathrm{jets}} = 0, 1, 2, 3+$) = 8 channels are considered and used for limits fitting. The result represent the expected constraints using the physics events and CMS experiment detector response simulated data via Monte Carlo methods (MC). The data and expected results correspond to run 2018 which has an integrated luminosity of $137.15\,\mathrm{fb}^{-1}$ at center-of-mass energy of 13 TeV.

Contents

Chapter 1

Introduction

This chapter contains a brief introduction of the experimental aspect of LHC physics, the CMS detector, and the physics behind the off-shell methods for constraining Higgs decay width. Finally, a very brief technical account for the Monte Carlo (MC) events production is also presented.

1.1 Physics at LHC

After the discovery of Standard Model (SM) Higgs boson in 2012, the Large Hadron Collider (LHC) went through a series of upgrades during Long Shutdown 1 (2013-2015) and finally restarted with center-of-mass energy reaching 13 TeV. Though the last particle promised by the SM has been discovered, no Supersymmetry (SUSY) particle has been spotted so far.

The null results has not discouraged physicists from other attempts of probing the beyond the Standard Model (BSM) physics in other ways. Higgs boson has a special role in the process mainly due to the Higgs field, which the particle originates from, couples to all massive particles and could be a window to probe BSM the physics, as discussed in Section 1.3.

The LHC has been operating at the same energy for the past 3 years (2016-2018) wtih the delivered luminosity ramping up in each year[1]. In this analysis, we will be using mostly simulation that were produced corresponding to Run period 2018, but scaled to $137.15\,\mathrm{fb}^{-1}$ to simulate the expected result by using the entire Run 2 data, for the Compact Muon Solenoid (CMS) detector.

1.2 The CMS Detector

The CMS is one of the two general purpose detectors built around LHC (the other one is ATLAS). It is 'compact' only when comparing to the ATLAS detector and is in fact heavier than the latter. Though we are not using the data accumulated by the CMS detector, the simulated events are reconstructed based on the material interaction and electronics responses of the real detector using the Geant4 package[2].

The defining feature of the CMS detector is the solenoid (as in its name) around the beam line, in the middle of the detctor. This superconducting solenoid produces a $4\,\mathrm{T}$ magnetic field and is $12.5\,\mathrm{m}$ long free bore[3]. This feature allows the track of any charged particle (given the momentum is not ridiculously high) to be bent when penetrating the detector layers and leave behind hit points that provide information regarding the particle's electric charge.

One important kind of 'final-state' particles used in this thesis is charged lepton, which includes electrons and muons (half-life of tauon is too short to interact with the detector directly); 'final-state' here means they are the end products of decay chain and are recorded by detector components directly. If we concentrate on the red and blue lines in Fig. 1, we can see that for the electrons (red), they leave a few hits that resemble a curved track within the Silicon Tracker layer and are stopped at the Electromagnetic

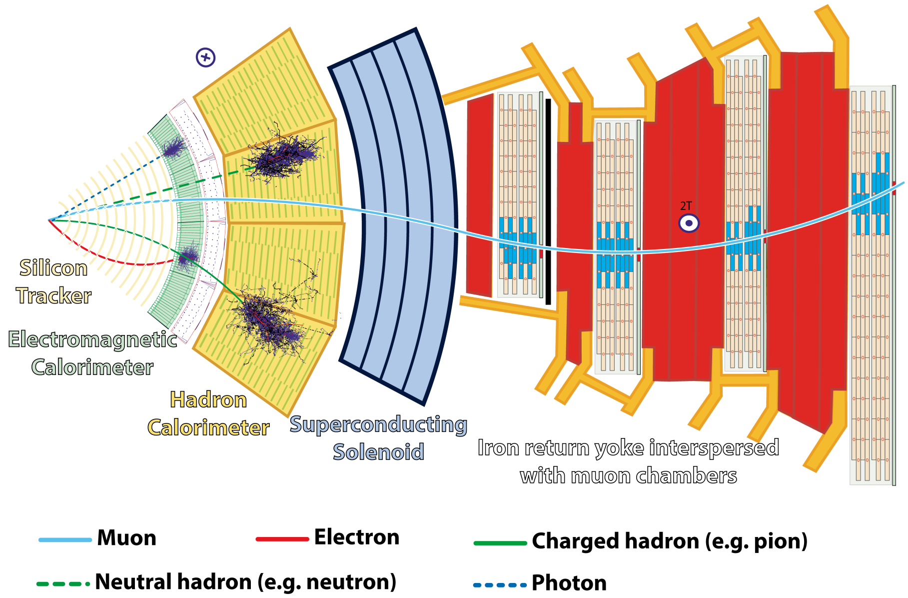


Figure 1: A section slice of the CMS detector where we can see how particles with different physical properties interact differently with the detector.

Calorimeter (ECAL).

For the muons (blue line in Fig. 1), they largely go through all the interior layers unhinged due to their high mass; pay close attention to the 'S'-shaped curve, this is due to the opposite magnetic field outside of the superconducting solenoid. Although information such as momentum relies on hits on the silicon trackers, the hits in the muon chambers and a matching trajectory is needed for the object to be reconstructed as a muon. Notice how the muon chambers occupy more than half of the detector by size, such design allows the CMS detector to excel in muon measurements and is the primary reason for the overall design.

Zooming out from the individual components of the detector, the complete (physics object) reconstruction process is complex and sometimes requires information from multiple parts of the detector, this algorithm is called the Particle Flow (PF)[4]. While the energy of simple and neutral objects such as photon is directly measured by the ECAL (with correction for zero-suppression), electron measurements need the information from the inner tracker (to determine charge and momentum), energy at the ECAL, and also sum of the energy of photons produced by bremsstrahlung compatible with the electron's track.

For gluons and quarks that come out of the interaction vertex, due to quark confinement, the detector is only be able to 'see' a narrow 'spray' of final state (stable) particles who are clustered into objects called 'jets'. There are different algorithms and criterias to combine a collection of measurements into a single jet, in the CMS at the moment, the anti-k_t clustering algorithm is used. In short, the algorithm would cluster objects in a cone (meaning it has a fixed $\eta - \phi$ space) originated from the vertex according to some parameters. The algorithm is resilient to QCD effects that would cause jets to split, such as shown in Fig. 2, where previous generation of algorithm may result in misidentification.

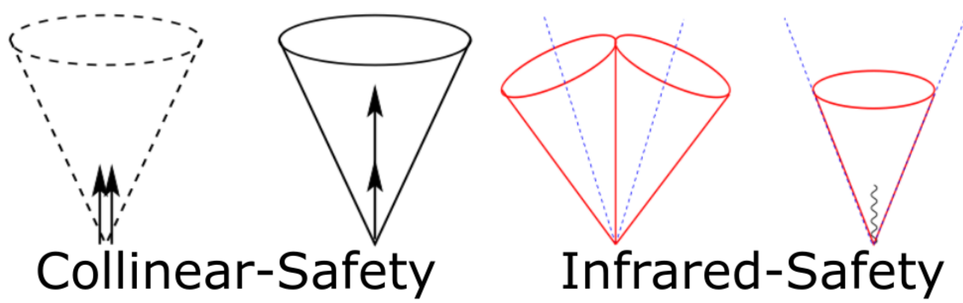


Figure 2: A illustration of two kinds of QCD effects in jets[1]

In this analysis, we use the AK4 jets which means that the algorithm is given a $R = 0.4$ parameter, another common choice is $R = 0.8$ which is preferred in some cases for boosted topologies due to a larger 'opening angle'.

[1] `https://twiki.cern.ch/twiki/bin/viewfile/Sandbox/Lecture?rev=1;filename=Philipp_Schieferdeckers_Lecture.pdf#page=3`

1.3 The Higgs boson and off-shell methods

The importance of Beyond the Standard Model (BSM) physics is self-evident, though the SM is one of the most precise theories in physics we ever had, the model requires many 'inputs' from experiment for parametrization, and, it cannot account for phenomena such as neutrino oscillation with its original form. While the direct searches in the past few years have all yielded null results, many indirect probings are being conducted as well.

As mentioned above, the Higgs boson has a special place as it can be seen as a 'bridge' between the SM and the BSM (in some models) as some SUSY particles can decay into it or decayed from it , or simply have a BSM production Feynman diagram that leads to more a higher Higgs bosons production rate than the SM expects. In any of these cases, the basic properties of the SM Higgs boson remain important.

Many properties are well measured, such as it's mass and spin[5], others, however, have not entered the realm of 'precision' physics so far. One of them is the decay width of the Higgs boson, which is of course associated to the particle's half-life. The SM predicts that the Higgs boson to have a decay width of 4.07 MeV. The problem is that the energy resolution of the CMS detector, which is around $\mathcal{O}(1)$GeV for the di-photon or 4-leptons final states, is not even remotely close to that prediction.

One of the ways suggested by the theorists[6, 7] to indirectly measure this decay width is called the off-shell method. When a Higgs boson decays into two vector bosons (VV) (in the scope of this thesis, two Z bosons), one of the them has to be off-shell, or the Higgs boson is off-shell, this is because $m_{\mathrm{H}} \approx 125\,\mathrm{GeV}$ is smaller than the mass of two W's($\approx 160\,\mathrm{GeV}$) or two Z's($\approx 182\,\mathrm{GeV}$). The Higgs boson decay branching ratio is coupled to the daughter particles' mass thus this cross section $\sigma_{\mathrm{H}\to\mathrm{VV}}$ is enhanced as the mass of Higgs boson gets closer to the on-shell VV mass.

In fact, the production of Higgs boson from a pair of vector boson is also related to the decay width of Higgs boson Γ_{H} via the 'propagator'[8]. In terms of the differential cross section:

$$\frac{\mathrm{d}\sigma_{\mathrm{vv}\to\mathrm{H}\to\mathrm{VV}}}{\mathrm{d}q_{\mathrm{H}}^2} \sim \frac{g_{\mathrm{vvH}}^2 g_{\mathrm{HVV}}^2}{\left(q_{\mathrm{H}}^2 - m_{\mathrm{H}}^2\right)^2 + m_{\mathrm{H}}^2\Gamma_{\mathrm{H}}^2} \tag{1}$$

where the two g on the right-hand side are the couplings for production from VV and decay to VV respectively. Integrating this equation near the on-shell mass of the SM Higgs boson or in the tail region (above the mass of VV), one can translate the differential cross section into event rate that can be (in theory) measured in experiments:

$$\begin{aligned} N_{\mathrm{vv}\to\mathrm{H}\to\mathrm{VV}^*}^{\text{on-shell}} &\sim \frac{g_{\mathrm{vvH}}^2 g_{\mathrm{HVV}}^2}{m_{\mathrm{H}}\Gamma_{\mathrm{H}}} \sim \mu_{\mathrm{vvH}} \\ N_{\mathrm{vv}\to\mathrm{H}^*\to\mathrm{VV}}^{\text{off-shell}} &\sim \frac{g_{\mathrm{vvH}}^2 g_{\mathrm{HVV}}^2}{(2m_{\mathrm{V}})^2} \sim \mu_{\mathrm{vvH}} \cdot \Gamma_{\mathrm{H}} \end{aligned} \tag{2}$$

The key takeaway is that, up to some correction factors, the event rate of off-shell Higgs scales linearly respect to the Higgs decay width Γ_{H} which allows indirect measurement of the width itself.

In this thesis, we will focus on the $\mathrm{H} \to \mathrm{ZZ} \to 2\ell 2\nu$ channel at the CMS detector using simulated (MC) events. The work conducted here is part of the ongoing analysis within a group formed by UCSB HEP group, Université Libre de Bruxelles, and Beihang University under the CMS Collaboration[2]. The analysis items and methods included in this thesis is a subset of what will be in the official analysis and is a part of the final measurement. More detail on what approximation has been taken in order to obtain a preliminary expected result is discussed in the next few sections.

[2] Internally, CMS AN-20-081

1.4 Background and signal simulation

The following list contains the MC samples (by physical process) used in this thesis, of the first two contain (off-shell) Higgs boson in the intermediate state:

- ggZZ offshell: Gluon fusion $gg \rightarrow \mathrm{H} \rightarrow \mathrm{ZZ}$
- VVZZ offshell: Vector Boson Fusion (VBF) into Higgs
- qqZZ, qqWZ, qqWW
- DY: Drell-Yan process
- TT: $t\bar{t}$, including samples with additional vector boson (TTW/TTZ) or photon + jets (TTGJets).

The MC samples for all processes except DY are produced in RunIIAutumn18MiniAOD-102X, DY sample is produced in RunIISummer16MiniAODv3 94X and scaled appropriately afterwards.

Various programs are used in the long chain of simulated events production. POWHEG MELA v2 [9] is used for the signal simulation. MadGraph [10] is used to generated NLO samples, Pythia [11] for parton showering and NNPDF 3.0[12] sets are used for the parton distribution functions.

Before diving into the procedure in which the signal samples are generated separately and subsequently combined with a re-weighting, it would be appropriate to give an account for the general idea behind the 'event weight' and its significance.

As described in the beginning of this section, events correspond to different physical processes are generated in different MC consigurations at different 'order' of the QCD/QED physics. Naively, one would imagine a process where the MC can directly simulate the physics at the LHC at a given center-of-mass energy. Unfortunately this

Example of some important weights		
Name	abbr.	Description
Generator weight	GEN wgt	Given by MC event generator
Pile-up weight	PU wgt	Correction for the pile-up effect
Matrix element weight	ME wgt	From generator that uses ME Likelihood approach
K-factor	Kfactor	Correction for LO cross section of QCD processes

Table 1: An incomplete list of weights used in the MC events used.

is neither efficient nor possible: not possible because some physical processes (especially QCD ones) are non perturbative and post hoc procedures are needed to 'add' physical object into the simulation. Not to mention that the SUSY physics we are searching for does not have a 'true model'.

It is also not efficient because the processes an analysis concerns (for example, in all SUSY searches) usually have a tiny (if not 0) cross sections compare to other common ones that can be found at $\sqrt{S} = 13\,\text{TeV}$ at the LHC. And it would be a waste of computing resources to generate the common processes over and over again.

In reality, Monte Carlo events are each given many 'weights' (Tab. 1), so that we don't have to generate uninteresting processes, and at the same time, for the events that lack in number (results in poor statistics in the distribution), one can optionally generate extension events set for it. Also, this enables the generation of 'unknown' processes which can be used to constrain possible new physics in a likelihood fit (against null hypothesis).

In this thesis, we explicitly use K-factors and scale-up and scale-down of it to obtain Electromagnetic systematical uncertainty in qqZZ/qqWZ/qqWW backgrounds. For the signal processes involving Higgs boson, a special treatment is given to merge and obtain high statistics sample from multiple 'raw' samples with different 'true' Higgs mass corresponding to the m_{H} term in the denominator on the right hand side of Eq. 1. This approach is also necessary for generation of off-shell (Higgs) decays. The procedures used and the resultant combined signal samples are discussed in the Sec. 2.2.

1.5 Uncertainties

A limited number of experimental and theoretical uncertainties in both signal and background processes are discussed here. Although dedicated to MC-only analysis, the leading theoretical uncertainties are considered:

Source	Uncertainty	Affected processes
Integrated luminosity	2.5%	GGH, VBF,qqZZ,qqWZ
Non-resonant background estimation	10%	TT, qqWW
Electro-weak	1σ	qqZZ, qqWZ
Higgs branching ratio	2%	GGH, VBF
GluonGluon background	Parametric	ggZZ

Table 2: Summary of systematic uncertainties considered in this thesis and their magnitude as well as processes affected by them.

Most of the uncertainties are simply experimental, for example the luminosity. The Electro-weak uncertainties are obtained by scaling up and down on the corresponding samples and Non-resonant background comes from separate studies that look at charged leptons consistent withing the Z boson mass peak. The GluonGluon background uncertainty is parametric respect to the $M_{\mathrm{T}}^{\mathrm{ZZ}}$ defined below.

Chapter 2

Methods

In this chapter, a description of event selection is given, as well as definitions of key physical variables and how they are used to select events. Then, a procedure regarding how the signal sample is manipulated to produce a high statistics off-shell Higgs sample is presented. Finally, the binning of variables used to obtain the results is determined and defined.

2.1 Event selection and physical variables

Proton bunches cross each at a rate of about 400 MHz in the beam line of the LHC, naturally, not all of these crossings are recorded due to both technical limitation of the electronics as well as the fact that the vast majority of these crossings don't produce inelastic collision that is energetic enough to be interesting.

After the selection of Level 1 (L1) trigger and the higher level trigger (HLT), less than 1000 events per second are permanently recorded and would go to off-line, full reconstruction. Among these, we only select the ones that passes certain HL triggers:

- HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8

- HLT_IsoMu24
- HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
- HLT_DoubleEle25_CaloIdL_MW
- HLT_DoublePhoton70
- HLT_Ele32_WPTight_Gsf
- HLT_Photon200

Most of the trigger names are self-explanatory, and are the starting point of the analysis. The purpose of the triggers is to insure events in the corresponding samples would contain the physics we wanted (since run data samples are divided according to what triggers they passed).

The next step is to 'add' composite variables that is significant to the physics of interests as well as making base-line cuts on the events. The jets are all AK4 jets unless mentioned otherwise, and a special type of jets, called B-tagged jets are produced using DeepFlavour algorithm.[13] We shall also define a few of the uncommon variables in the list, and physical motivations are given in the following paragraphs.

After mandating passage of the set of triggers above, based on more delicate physics reasons, the baseline cuts are:

- No ak4-jet b-tagged jet
- Both leptons have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$
- $\left|\Delta\phi_{\ell\ell\_E_{\mathrm{T}}^{\mathrm{miss}}}\right| > 1.0$
- $\left|\Delta\phi_{\ell\ell\mathrm{Jets}\_E_{\mathrm{T}}^{\mathrm{miss}}}\right| > 2.5$

- $|m_{\ell\ell} - 91.2\text{GeV}| < 15\text{GeV}$: the signal process consists of $Z \to \ell\ell$, we require the di-lepton system has a mass that is consistent within the Z mass peak.

- $p_{\mathrm{T}}^{\ell\ell} > 55\text{GeV}$: the Drell–Yan (DY) process creates a lot of backgrounds events, but their di-leptons go back-to-back with expected value of this variable close to 0.

- $E_{\mathrm{T}}^{\mathrm{miss}} > 125\text{GeV}$: the signal process creates true $E_{\mathrm{T}}^{\mathrm{miss}}$ with neutrinos, this cut also reduce bkg such as DY.

- $\min\left|\Delta\phi_{\mathrm{j}\_E_{\mathrm{T}}^{\mathrm{miss}}}\right| > 0.25$

In addition, to reduce the detector effects at high pseudo rapidity angels:

- $\eta_{\mu} < 2.4$

- $\eta_{e} < 2.5$

A lot of the cuts are related to angles between various physical objects presented in the reconstruction. The reason is simple: the signal events, where Higgs goes to ZZ and one Z goes to 2 charged leptons and the other goes to 2 neutrinos, ideally would have the two Z's 'back-to-back' in Higgs boson's rest frame, leading to a large angel in the transverse plane.

Furthermore, in background events which do not contain this kinematic feature, the correlation in the directions of $E_{\mathrm{T}}^{\mathrm{miss}}$ and the transverse momentum of the leptons is weaker.

To use this kinematic feature to increase signal to background ratio, we define $\left|\Delta\phi_{\ell\ell\_E_{\mathrm{T}}^{\mathrm{miss}}}\right|$ as the azimuthal angel (perpendicular to the beam line) between the di-lepton system and the transverse missing energy. In the signal events that produce 0 jet, this variable should be π. The cut is lowered to 1.0 due to the finding that in the (not so rare) cases where there are jet(s) recoiling against the ZZ system, this variable can go quite low.

This leads to the next variable on the list, $\left|\Delta\phi_{\ell\ell\mathrm{Jets}\_E_{\mathrm{T}}^{\mathrm{miss}}}\right|$, which is almost the same except that we add all jets' momentum into the di-lepton system to account for the events that have produced jets, which in turn would cause the angel be lower in a multi-body final state.

Finally, $\min\left|\Delta\phi_{\mathrm{j}\_E_{\mathrm{T}}^{\mathrm{miss}}}\right|$ is the minimum azimuthal angel difference between any of the jet (that passes cuts) and the $E_{\mathrm{T}}^{\mathrm{miss}}$, it exist because jets are one of the most difficult physical objects measure, they often create so-called instrumental $E_{\mathrm{T}}^{\mathrm{miss}}$ due to jet mis-measurements and it can be quite large in magnitude. However, such mis-measurements often yield large $E_{\mathrm{T}}^{\mathrm{miss}}$ in the direction of the original jet. This cuts requires angular separation since in signal process, the jet recoils against ZZ system.

We also define variables that are not cut on, but are used for fitting:

$$D_{jj}^{VBF} = \frac{P_{SM}^{VBF}(\vec{\Omega})}{P_{SM}^{gg}(\vec{\Omega}) + P_{SM}^{VBF}(\vec{\Omega})} \tag{3}$$

As introduced in[14]. In short, this variable (discriminator) is sensitive to the VBF physics and the correlation between (angels and mass of) the outgoing jets resulted from the VBF topology. At the same time:

$$M_{\mathrm{T}}^{\mathrm{ZZ}} = \left[\sqrt{p_{\mathrm{T},\ell\ell}^{2} + m_{2\ell}^{2}} + \sqrt{{E_{\mathrm{T}}^{\mathrm{miss}}}^{2} + m_{\mathrm{Z}}^{2}}\right]^{2} - \left[\vec{p}_{\ell\ell} + \vec{E}_{\mathrm{T}}^{\mathrm{miss}}\right]^{2} \tag{4}$$

is defined based on the hypothesis that the $E_{\mathrm{T}}^{\mathrm{miss}}$ is comprised of mainly the two neutrinos from one of the Z bosons. We shall see the usefulness of this variable in channels where not enough jets are present to construct the DJJVBF variable.

On top of the cuts stated in the beginning, several $E_{\mathrm{T}}^{\mathrm{miss}}$ filter and lepton isolated tracks veto are present in the underlying analysis framework to reject events with pathological reconstruction.

2.2 Signal samples re-weighting

Extra attention was given to the off-shell Higgs sample used in this thesis and two different kinds of re-weighting of the simulated events are applied in to produced a MC sample with wide mass spectrum way beyond the mass of Higgs ($\approx 125\,\mathrm{GeV}$). We use the gluon fusion Higgs (ggH) sample to illustrate the procedures, the same procedures are applied to the VBF samples as well.

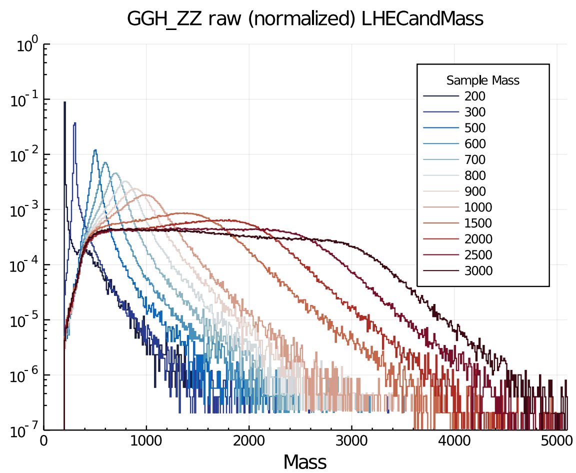
We start by generating separate samples with different Higgs mass (LHECandMass). This is the mass of Higgs terms appears in the propagator on the R.H.S of Eq. 1 as mentioned before. The raw distribution of the true mass in different samples (without any weight) is shown in Fig. 3 (left). As expected, the peak of the distribution moves to the right as the mass of the sample becomes larger, at the same time, the 'peak' of samples with very large mass are wider because the lower edge of the peak is dominated by an underlying exponential 'tail'. We also see that for some lower mass samples (200, 300, 400 etc.), they have a cut-off beyond $\mathrm{M} \approx 2500$ which means they have 0 statistics beyond that mass range. After applying the GEN, PU, and ME weights given by their individual MC process and JHUGen MELA, as shown in Fig. 3, we see that they are consistent with each others' line shape. However, it is clear that:

(i) Lower mass samples have cut-offs in the tail region

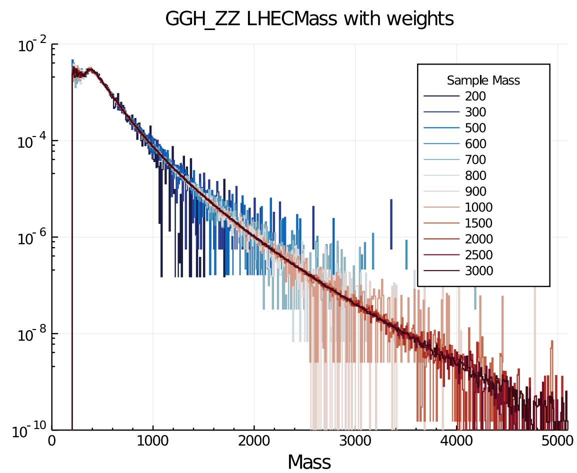
(ii) Samples have poor statistics in mass windows that are far from their true mass (as listed in the legend).

The second point is best illustrated by the wide spikes of lower mass samples near their cut-offs, as well as the visible fluctuations of high-mass samples in the mass region (don't be fooled by the visual, the plot is in semi-log scale).

The goal of the combination of samples is to use all the events, but with a correction



(a)



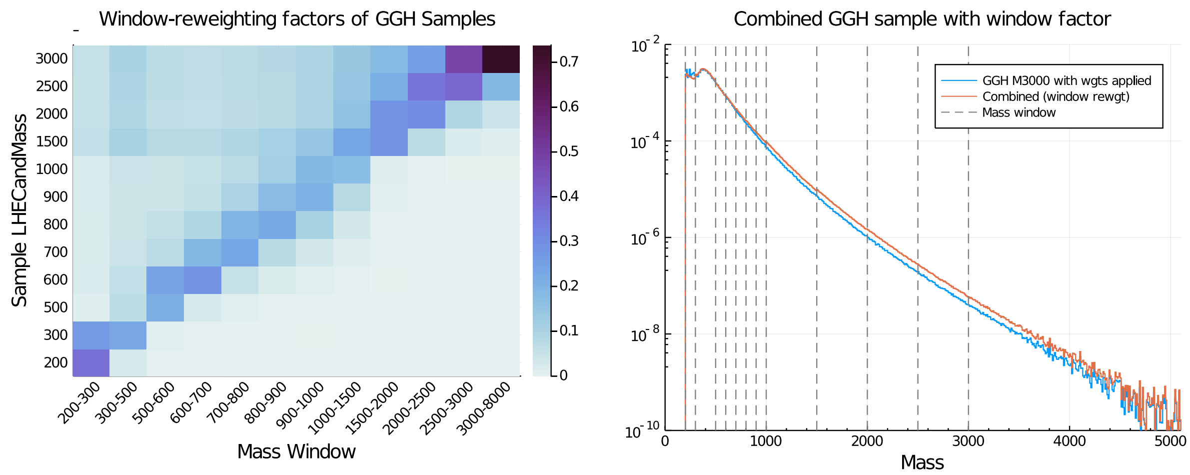
(b)

Figure 3: Normalized distributions of LHECandMass before (left) and after (right) applying the weights. Together they show a need to combine samples for a wide-range, high statistics signal sample. Bin size = 10 GeV.

weight such that each sample has a higher weight in the region where they posess good statistics. Of course while keeping the overall normalization stays unchanged. To do this, we pick a list of 'mass windows' with edges sitting on the true masses of the samples, and we define effective number of events $\mathrm{N_{eff}} = \frac{(\sum \mathrm{wgts})^2}{\sum(\mathrm{wgts}^2)}$ within each mass window. Here, the wgts corresponds to the product of PU wgt, GEN wgt, K-factor, and ME weight for the GGH sample in consideration. For a specific GGH sample i_0 and its events fall in a mass window j, $\mathrm{N_{eff}^{i_0 j}}$ is first obtained and a re-weighting factor can be computed:

$$\mathrm{wgt}^{i_0 j}_{\mathrm{window}} = \frac{\mathrm{N}^{i_0 j}_{\mathrm{eff}}}{\sum_i \mathrm{N}^{ij}_{\mathrm{eff}}} \tag{5}$$

This factor is applied to all events from sample i_0 within the window j. Conceptually, the effective number of events ensures the weight is not skewed by the difference in the overall normalization of samples, and in each of the mass windows, samples with more concentrated statistics in that window are given a higher weight. In Fig. 4 (a), a clear diagonal pattern can be seen, physically it means that samples with higher true mass are given a higher weight in tail mass windows— consistent with the expected outcome.



(a) (b)

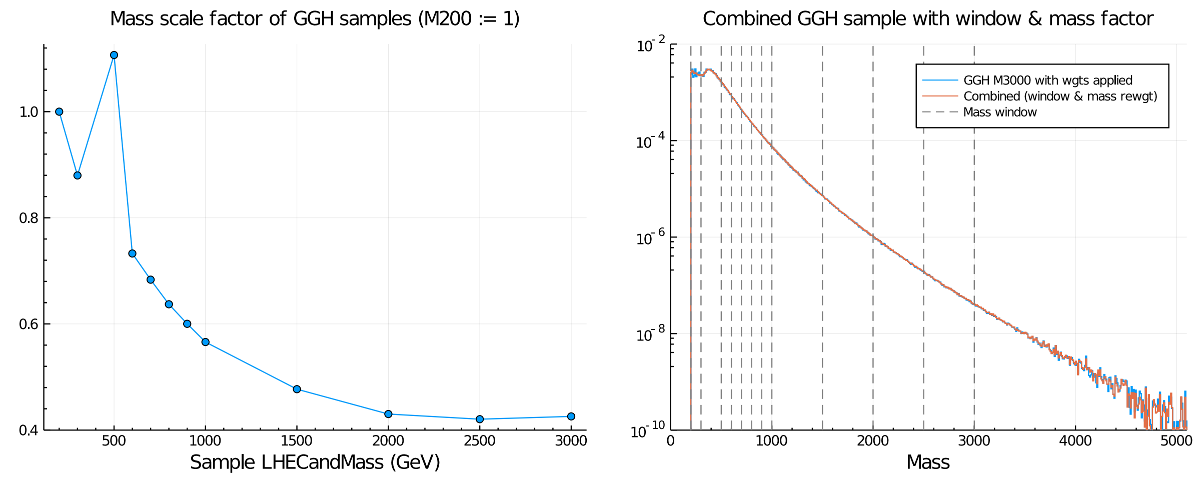
Figure 4: Heatmap of window re-weight factors of different samples and mass windows (left); effects of applying window factors for the combined sample(right)

However, even with the unit normalization, there are still inconsistency in the shape as shown in Fig. 4 (b). This is likely because the finite number of events and non-infinitesimal mass window size used. We introduce another correction factor for this small artifacts. Iteratively going through every sample, between the previous and the next one, derive a sample mass factor based on:

$$\mathrm{wgt}^{i,i+1}_{\mathrm{mass}} = \frac{\sum \mathrm{wgt}_{\mathrm{i}}}{\sum \mathrm{wgt}_{\mathrm{i+1}}}, \text{ for events that has Mass between sample mass of } i \text{ and } i+1 \quad (6)$$

This factor corrects the high variations of overall normalizations between samples, the factors and result are shown in Fig. 11. As expected, high mass samples need a down correction (not by a lot) to eliminate the deviated trend before.

Finally, 1.098946 is multiplied to the weights of all ggZZ processes (all of BKG, SIG,



(a) (b)

Figure 5: Iterative sample mass factors obtained (left) and the final combined sample (right)

BSI of GGH sample) as a K factor for Next-to-next-to-leading-order (NNLO) → Next-to-next-to-next-to-leading-order (N3LO).

Although we used the particular matrix element weights for one of the signal hypothesis, these two correction factors apply too all hypothesis and a plot of them without normalization are shown in Fig. 6. As expected, background exceeds signal by more than 100% which is partially why the constrain is hard to obtain.

2.3 Strategy in variable selection and binning and systematical uncertainties

After preparing the signal samples and decided on the event selection criteria, we move on to decide on the variables and their (2D histogram, as shown in Fig. 7) 'binning' before a combined-limits fit can be applied. As discussed in earlier sections, one of the more inventive variables newly introduced specifically for the analysis is the DJJVBF

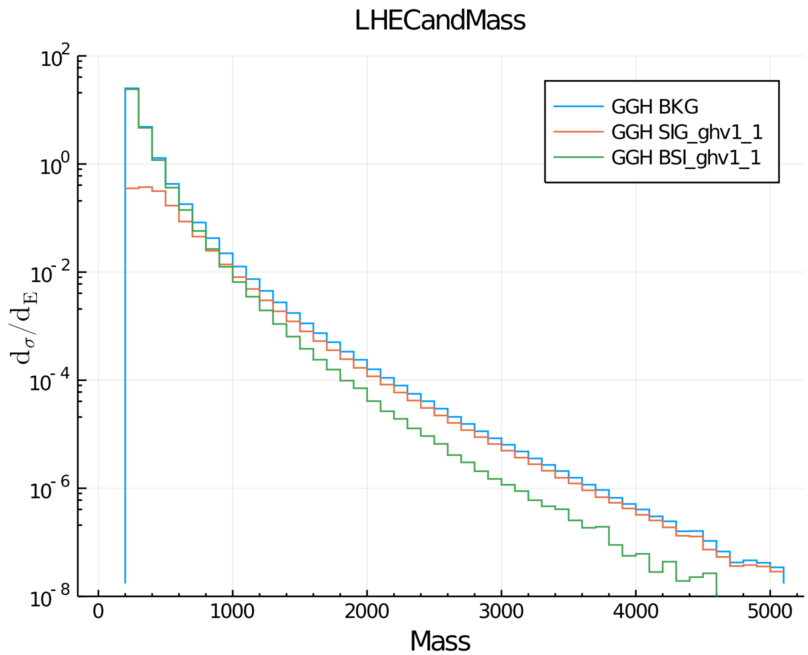


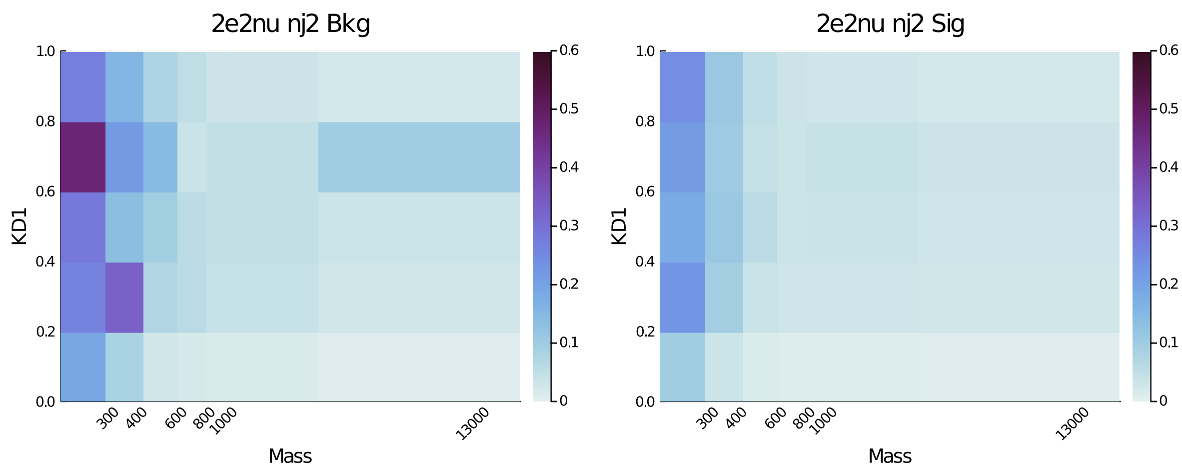
Figure 6: Distributions of background, signal, and background signal interaction

discriminator. However, it is clear that this variable is undefined for events with $N_{\text{jets}} < 2$. To not 'waste' any statistical significance, we use $E_{\text{T}}^{\text{miss}}$ in its place for the $N_{\text{jets}} = 0, 1$ categories. In total, we have 2 (ee or $\mu\mu$) $\times$ 4 ($N_{\text{jets}} = 0, 1, 2, 3+$) = 8 channels to consider when making histogram templates. Bin edges for different categories in number of jets are as the following:

- $N_{\text{jets}} < 2$
  - M_{T}^{ZZ} = 150, 300, 400, 600, 800, 1000, 13000
  - KD1 = DJJVBF = 0, 0.2, 0.4, 0.6, 0.8, 1
- $N_{\text{jets}} >= 2$
  - M_{T}^{ZZ} = 150, 300, 400, 600, 800, 1000, 13000
  - KD1 = $E_{\text{T}}^{\text{miss}}$ = 125, 200, 280, 420, 500, 800, 13000

The higher mass ($M_{\mathrm{T}}^{\mathrm{ZZ}}$) bins are wider because samples have difficulty filling them due to physical reasons (especially for backgrounds) and the cuts being applied. As shown in Fig. 7, the relative error (err/counts) of each bin in the histogram templates are displayed, a balance between significance and uncertainty is obtained. We use the above binning for all samples and 8 channels that are considered in this thesis. See Appendix B.1 for a compilation of template histograms of GGH sample.

For low-yield background samples, due to the nature of NLO samples, bins sometimes would have negative content. To mitigate it's effect on the likelihood fitting (pathological), we replace the bin content by (Integral of the histogram) $\times 10^{-5}$.



(a) (b)

Figure 7: Background (left) and Signal (right) histogram templates for $2e2\nu$ with $N_{\mathrm{jets}} = 2$.

Systematical uncertainties are also included in the fitting:

- Luminosoty: GGH_ZZ, VBF_ZZ, qqZZ, qqWZ
- NRB Estimation: TT
- Branching Ratio of Higgs to ZZ to 4l: GGH_ZZ, VBF_ZZ

- K-fatcor of backgroun gluongluon parameter

Most of them are assigned with a $ln(N)$ uncertainty of 1-σ or 10% except the k-factor which is a parameters directly multiplied with backgrounds Parton Distrubution Function (PDF).

Chapter 3

Results and interpretation

Final results of this thesis is presented. As the result acts as 'expected' limits, the ongoing work and potential interpretation are discussed.

3.1 Limits on Higgs decay width

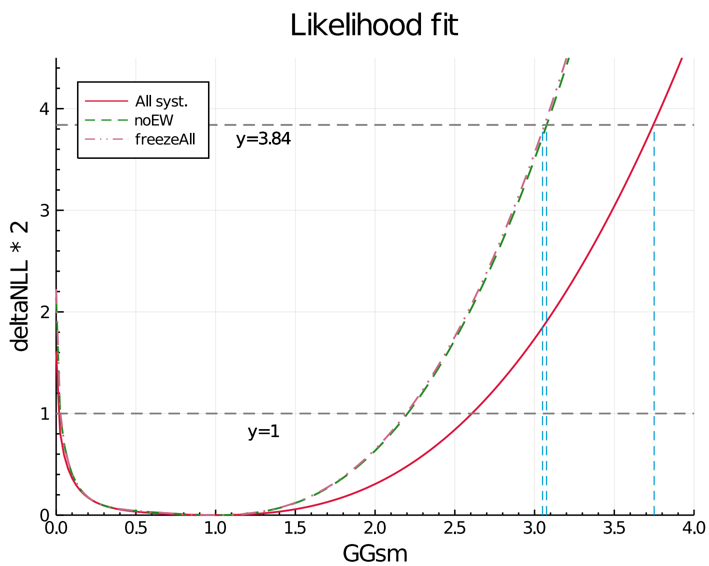


Figure 8: Maximum likelihood fit of $\mu^{\mathrm{H}}_{\mathrm{off-shell}}$ (off-shell rate ratio). For all systematics (red), no Electroweak syst. (green), 0 syst. (orange): y-intersect={1.61, 2.14, 2.22}, 1σ lower limits={0.025, 0.025, 0.025}, 1σ higher limits={2.6, 2.2, 2.2}, 95% CL limits={3.75, 3.08, 3.05}, respectively.

After running through Combined Limited tool for likelihood fitting, we first extract the significance of the off-shell rate. (Fig. 8) The y-axis is understood to be σ^2 in terms of significance, thus the intersection with the y-axis is the signal sensitivity, or in other words, rejection of the 0 width hypothesis (no off-shell), which has a significance of $\sqrt{1.61} \approx 1.26\sigma$ in this fit with all the systematic uncertainties included. As the systematics are turned off', the constraint becomes tighter, producing an error band for the expected final result in the upcoming official anlysis.

Furthuermore, by un-'freezing' the RF=μ_{F} and RV=μ_{V} and adopt the range suggested [15], the constraint on the decay width of Higgs Γ_{H} is shown in Fig. 9. The minimal (max likelihood) falls on 4.07 MeV, consistent with the Standard Model hypothesis being used. Again, 1σ and 95% CL are marked respectively. And a final result of $\Gamma_{\mathrm{H}} < 16.38$ MeV shall be quoted.

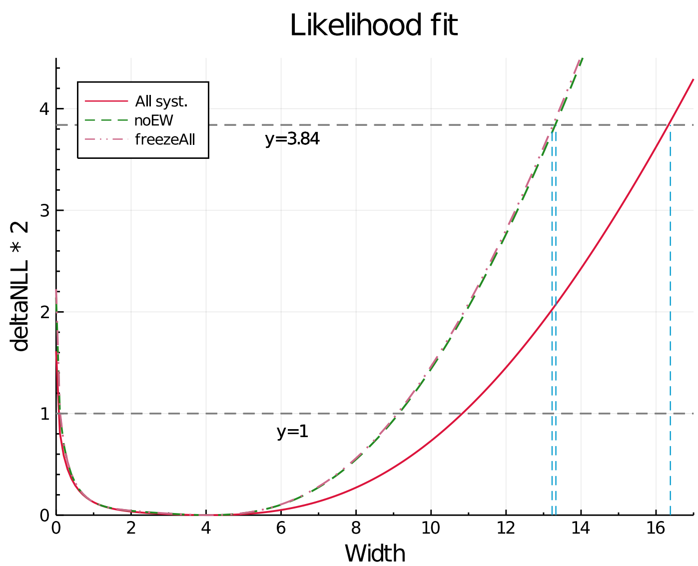


Figure 9: Maximum likelihood fit of Higgs decay width. For all systematics (red), no Electroweak syst. (green), 0 syst. (orange): y-intersect={1.61, 2.14, 2.22}, 1σ lower limits={0.10, 0.10, 0.10} MeV, 1σ higher limits={10.78, 9.16, 9.06} MeV, 95% CL limits={16.38, 13.33, 13.23} MeV, respectively.

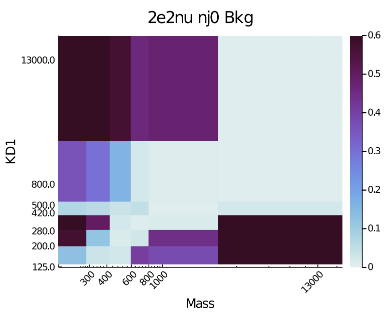
Appendix A

Weights Table for Higgs Sample

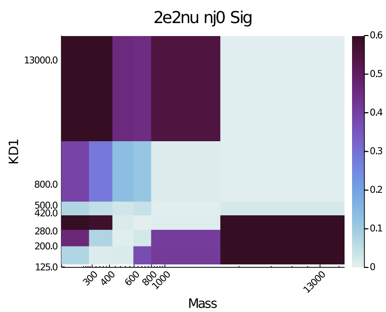
Appendix B

Additional Figures

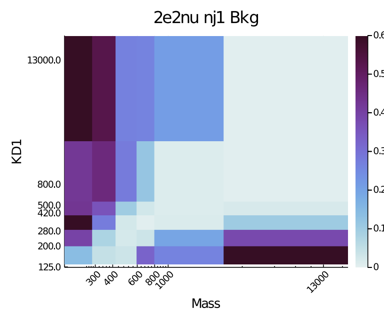
B.1 GGH Sample Fitting Templates of Background and Signal



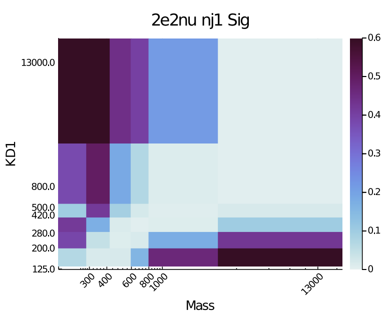
(a)



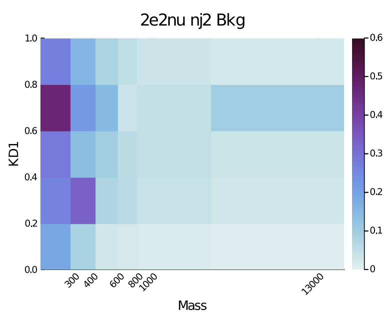
(b)



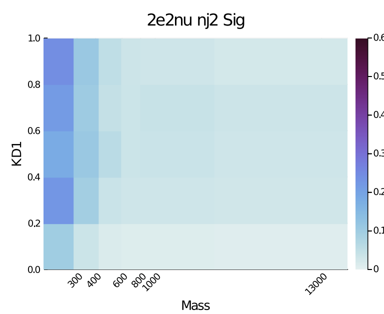
(c)



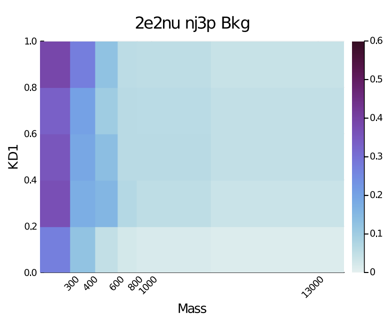
(d)



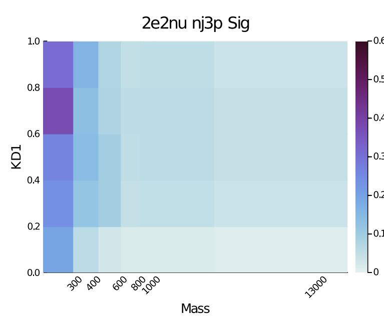
(e)



(f)



(g)



(h)

Figure 10: Iterative sample mass factors obtained (left) and the final combined sample (right)

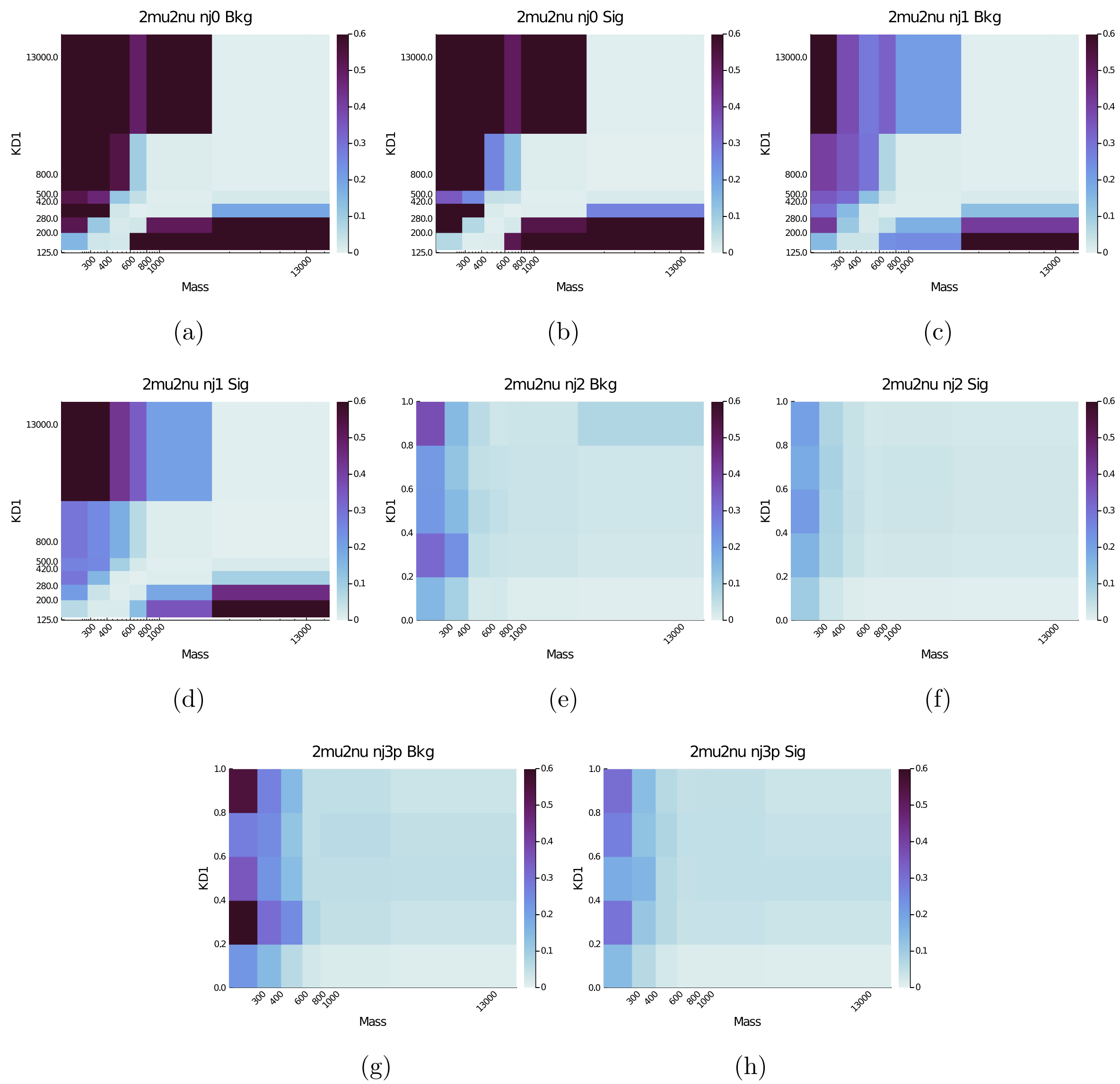


Figure 11: Iterative sample mass factors obtained (left) and the final combined sample (right)

Bibliography

[1] "LumiPublicResults < CMSPublic < TWiki."
`https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[2] Geant4 Collaboration, *Geant4—a simulation toolkit*, .

[3] V. I. Klyukhin, N. Amapane, A. Ball, B. Curé, A. Gaddi, H. Gerwig, M. Mulders, V. Calvelli, A. Hervé, and R. Loveless, *Validation of the CMS magnetic field map*, arXiv:1606.0038.

[4] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, arXiv:1706.0496.

[5] C. Collaboration, *Measurement of the properties of a higgs boson in the four-lepton final state*, arXiv:1312.5353.

[6] N. Kauer and G. Passarino, *Inadequacy of zero-width approximation for a light higgs boson signal*, .

[7] J. M. Campbell, R. K. Ellis, and C. Williams, *Bounding the higgs width at the LHC using full analytic results for* $gg \to e{-}e{+} - +$, .

[8] U. Sarica, *Measurements of Higgs Boson Properties in Proton-Proton Collisions at* $\sqrt{s} = 7$*, 8 and 13 TeV at the CERN Large Hadron Collider*. Springer Theses. Springer International Publishing.

[9] P. Nason and G. Zanderighi, *W+W-, WZ and ZZ production in the POWHEG-BOX-v2*, .

[10] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, .

[11] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the monash 2013 tune*, .

[12] T. N. Collaboration, R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali, *Parton distributions for the LHC run II*, arXiv:1410.8849.

[13] CMS Collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, arXiv:1712.0715.

[14] C. Collaboration, *Measurements of the higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state*, arXiv:1901.0017.

[15] "Measurements of properties of the higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13\mathrm{TeV}$." Library Catalog: cds.cern.ch Number: CMS-PAS-HIG-19-001.