Di-photon background studies for HH→ bbγγ analysis in CMS

Soumya Mukherjee

8th December 2021, 23rd MCnet Meeting, University of Manchester
About myself

I am a PhD student in **Tata Institute of Fundamental Research, Mumbai** from 2018. Working in CMS experiment with **Run-2 data**

**Thesis topic:**

(i) Non resonant di-Higgs production in $bb\gamma\gamma$ final state.  \[\text{JHEP 03 (2021) 257}\]
(ii) Measurement of Higgs coupling with b-quarks in Vector Boson Fusion production mode.
(iii) Missing transverse energy (MET) object reconstruction in CMS detector

Currently I am in **University of Manchester** working with **Prof. Mike Seymour** for MCnet project

→ Study of diphoton production in LHC with Herwig
**Introduction to non-resonant HH process**

- Non-resonant HH process → able to directly probe **Higgs self coupling** ($\lambda_{HHH}$) at the LHC
- $\lambda_{HHH} = \lambda$ is one of the most important characteristics of the Higgs boson, but yet to be measured
- Major production mode is: Gluon-gluon fusion
- Cross section is small due to the negative interference between the two diagrams

**Gluon-Gluon Fusion (ggHH)**

- x-sec at N2LO QCD accuracy:
  - @ 13 TeV 31.05 $\pm$2.2% / -5.1% fb
  - @ 14 TeV 36.69 $\pm$2.1% / -4.9% fb

- Coupling parameters are measured in terms of coupling modifiers wrt SM in $\kappa$ - framework

\[
\kappa_t = \frac{y_t^{\text{obs}}}{y_t^{\text{SM}}} \quad \kappa_{\lambda} = \frac{\lambda^{\text{obs}}}{\lambda^{\text{SM}}}
\]
HH search in LHC

- HH decay modes being explored using full Run2 (137 fb\(^{-1}\)) data both by CMS & ATLAS:
- Modes with large branching ratios (BR) utilized for atleast one of the H decays:
  - → bb (58%) and WW*(21%)
- HH→ 4b , bb\(\tau\)\(\tau\), bbWW, bbZZ, 4W, WW\(\tau\)\(\tau\), 4\(\tau\), WW\(\gamma\)\(\gamma\)

- Despite of small branching ratio of H→\(\gamma\)\(\gamma\) , due to good ecal resolution and comparatively less background, HH → bb\(\gamma\)\(\gamma\) is the on of the best sensitive channel.
Important features

- **Event selections**
  - Data and MC photons, b-jet selection

- **Signal Vs Background separation**

- **Construction of analysis categories based on purity**

- **Fitting the signal and background \( m_{\gamma\gamma} \) and \( m_{bb} \)**

- **Extraction of results**

---

**Why MC is important in this analysis**

Although the final results derived by the direct **fitting of data** assuming the **full background**

But for the training and validation of any multivariate training (BDT or DNN), we have to rely on MC background instead of data to get rid of the any kind of bias.

---

**Current scenario**

The major background in this analysis is \( \gamma\gamma/\gamma + \text{jets} \).

The current available MC samples for \( \gamma\gamma/\gamma + \text{jets} \) are not enough to follow the data, we have large deficiency of MC simulate events comparing to Data (factor 2.4).

---

The same approach is being applicable for all \( H \rightarrow \gamma\gamma \), although in single \( H \rightarrow \gamma\gamma \) phase space the MC has better agreement with data.
MC description of currently used $\gamma\gamma$ samples

→ Prompt di-photon production along with some jets (upto 3 jets)
  Generator : Sherpa
  Additional : Invariant mass of two photons $m_{\gamma\gamma} > 80$ GeV

→ Fake - prompt production
  Generator : Pythia
  Additional : EM-Enriched filter applied to enhance the fake photon contribution
    Invariant mass of two photons $m_{\gamma\gamma} > 80$ GeV

→ QCD fake - fake production
  Generator : Pythia
  Additional : Double EM Enriched filter applied to have two isolated photons
    Invariant mass of two photons $m_{\gamma\gamma} > 80$ GeV
  Problem: Very low filter efficiency, high weights due to large cross section, small number of events selected
  Current solutions: Data-Driven method from a control region orthogonal to signal or (analysis region)

→ Prompt di-photon production along with $1/2$ b-jets (upto total 3 jets)
  Generator : Sherpa
  Additional : Invariant mass of two photons $m_{\gamma\gamma} > 80$ GeV

Important for HH→bb$\gamma\gamma$
Diphoton variable modelling in Single - $H \rightarrow \gamma \gamma$ analysis

For the GGH $\rightarrow \gamma \gamma$ process where no requirements on jets, MC prediction agrees with data well

Problem arises when there is some jet requirements, overestimation of fake - photon contributions
Diphoton variable modelling in HH → bbγγ analysis

→ Overall normalization factors has been applied to the γγ + jets contributions for HH→ bbγγ is 2.4 and for VBFHH → bbγγ is 4.4
→ The normalization factors are basically high compared to the usual k-factors for different processes.
→ The main reason for this is mis-modelling of γγ + jets monte carlo which can’t follow the data
The signal has been modelled from MC simulation using multi gaussian fits. The background contributions have been estimated directly from data using exponential function.
Results from Run-2 data from LHC (CMS only)

Allowed range @ 95% CL

$\kappa_\lambda = \frac{\lambda_{HHH}}{\lambda_{SM}^{HHH}}$

$\kappa_\lambda$ = 1

$\sigma_{HH}B(HH \rightarrow \gamma\gamma b\bar{b})$ (fb)

$C_{2V}$ = 1

Observed: $-3.3 < \kappa_\lambda < 8.5$

Expected: $-2.5 < \kappa_\lambda < 8.2$

95% CL Upper Limit on (inclusive cross section*BR)

<table>
<thead>
<tr>
<th>Inclusive HH (* SM)</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.7</td>
<td>5.2</td>
</tr>
</tbody>
</table>

JHEP 03 (2021) 257
Current studies with Herwig interface

→ I have started with in-build $\gamma\gamma$ production module: MEGammaGamma

→ ATLAS 2011_S9120807 @CoM 7 TeV used as reference
Plan

- We have seen to analyse Run-2 data for $HH \rightarrow b \gamma \gamma$ that the available set of MC samples are not sufficient.
- In CMS currently no Herwig diphoton background sample is available so far.
- I have currently started diphoton production in LHC through standalone Herwig 7 (not included in CMS environment), ATLAS 7 TeV data used as reference.
- Shape of MC distributions follows data but disagreement founds in event yields, currently working on it.
- Once diphoton background production will be done with good agreement with data-mc, I will go specifically for the production of $\gamma \gamma$ along with 2-bjets in Herwig.
- Final target is to include the model in Run-3 CMS analysis.

thank you