

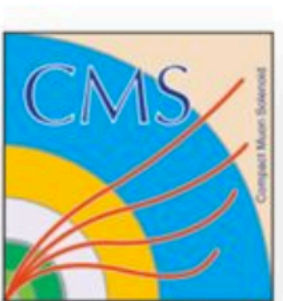
$$h(125) \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$$

Hgg Meeting  
22<sup>nd</sup> October 2018

Tanvi Wamorkar<sup>1</sup>, Toyoko Orimoto<sup>1</sup>, Andrea Massironi<sup>2</sup>

[1] Northeastern University

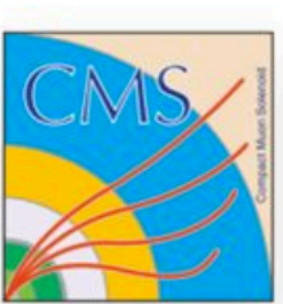
[2] INFN Milano-Bicocca and CERN



# Overview of the updates

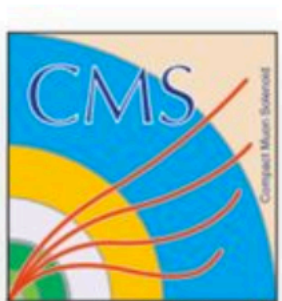
---

- Reminder of the analysis setup [Link to the previous presentation](#)
- Distributions at the pre-selection level & MVA optimization
- Discussion of Di-Photon Pairing
- Background model
- Signal model
- Next steps



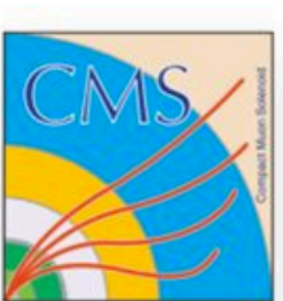
# Analysis Setup

- Analysis framework based on FLASHgg
- Following low mass Hgg analysis as close as possible; same trigger strategy, same processed samples
- Primary dataset: DoubleEG (ReReco) [36/fb with golden json]
- **Signal samples:**
  - /SUSYGluGluToHToAA\_AToGG\_M-XX\_TuneCUETP8M1\_13TeV\_pythia8/ ;  $XX \in [10,60]$  GeV
- **Background samples:**
  - /GJet\_Pt-20to40\_DoubleEMEnriched\_MGG-80toInf\_TuneCUETP8M1\_13TeV\_Pythia8/
  - /GJet\_Pt-20toInf\_DoubleEMEnriched\_MGG-40to80\_TuneCUETP8M1\_13TeV\_Pythia8/
  - /GJet\_Pt-40toInf\_DoubleEMEnriched\_MGG-80toInf\_TuneCUETP8M1\_13TeV\_Pythia8/
  - /QCD\_Pt-30to40\_DoubleEMEnriched\_MGG-80toInf\_TuneCUETP8M1\_13TeV\_Pythia8/
  - /QCD\_Pt-30toInf\_DoubleEMEnriched\_MGG-40to80\_TuneCUETP8M1\_13TeV\_Pythia8/
  - /QCD\_Pt-40toInf\_DoubleEMEnriched\_MGG-80toInf\_TuneCUETP8M1\_13TeV\_Pythia8/
  - DiPhotonJetsBox\_M40\_80-Sherpa
  - DiPhotonJetsBox\_MGG-80toInf\_13TeV-Sherpa



# Online Selection

- Online selection strategy based on low mass Hgg analysis
- [Data] Low mass Diphoton triggers (OR of the two paths)
  - HLT\_Diphoton30PV\_18PV\_R9Id\_AND\_IsoCalold\_AND\_HE\_R9Id\_DoublePixelVeto\_Mass55
  - HLT\_Diphoton30EB\_18EB\_R9Id\_OR\_IsoCalold\_AND\_HE\_R9Id\_DoublePixelVeto\_Mass55
- [Data & MC] Apply low mass pre-selection (tighter than HLT)
  - Removed the following selections from the low mass pre-selection cfi: Scaled  $p_T$  selections, very loose MVA selection
  - Optimize MVA cuts on the  $4\gamma$ 's



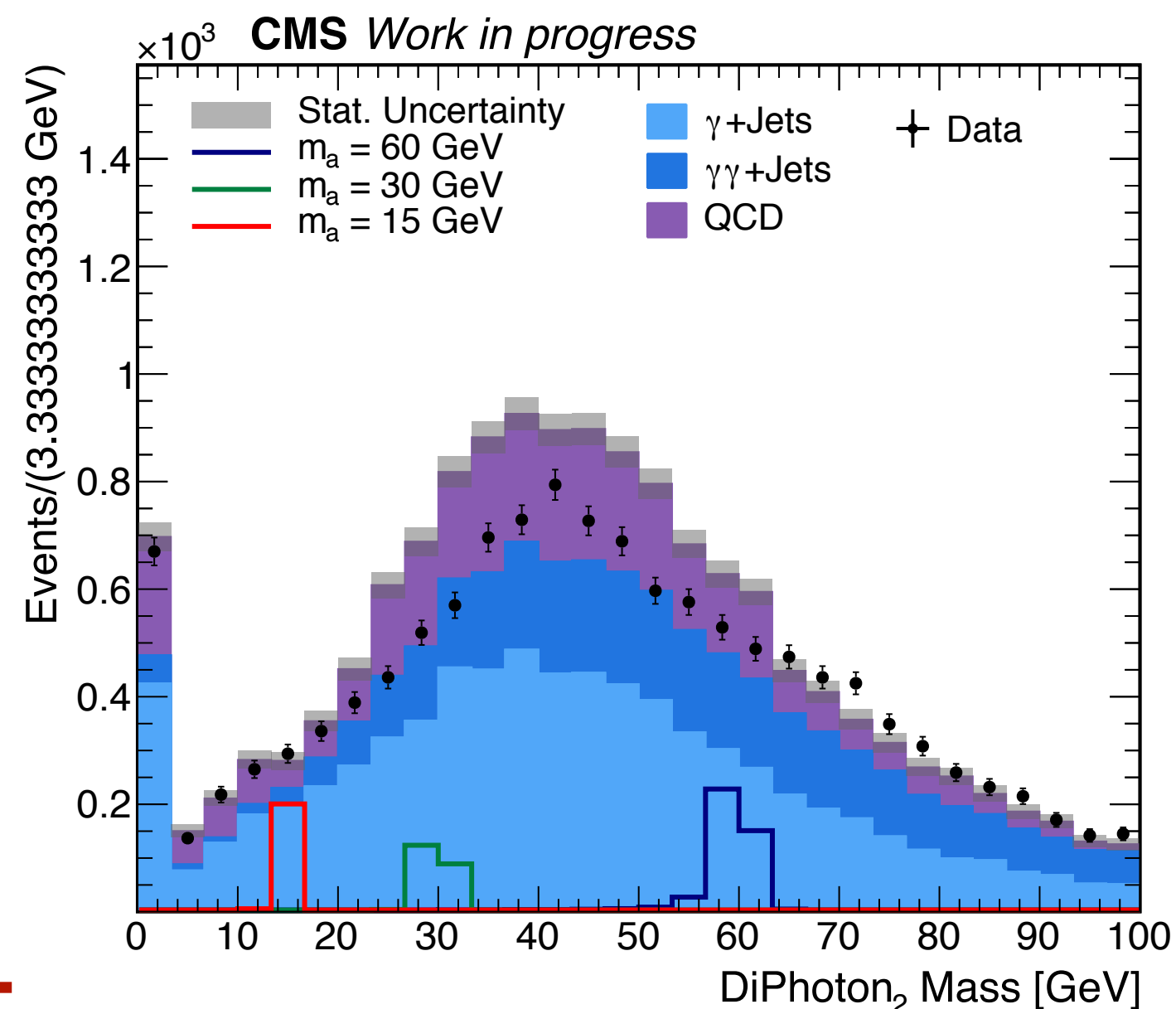
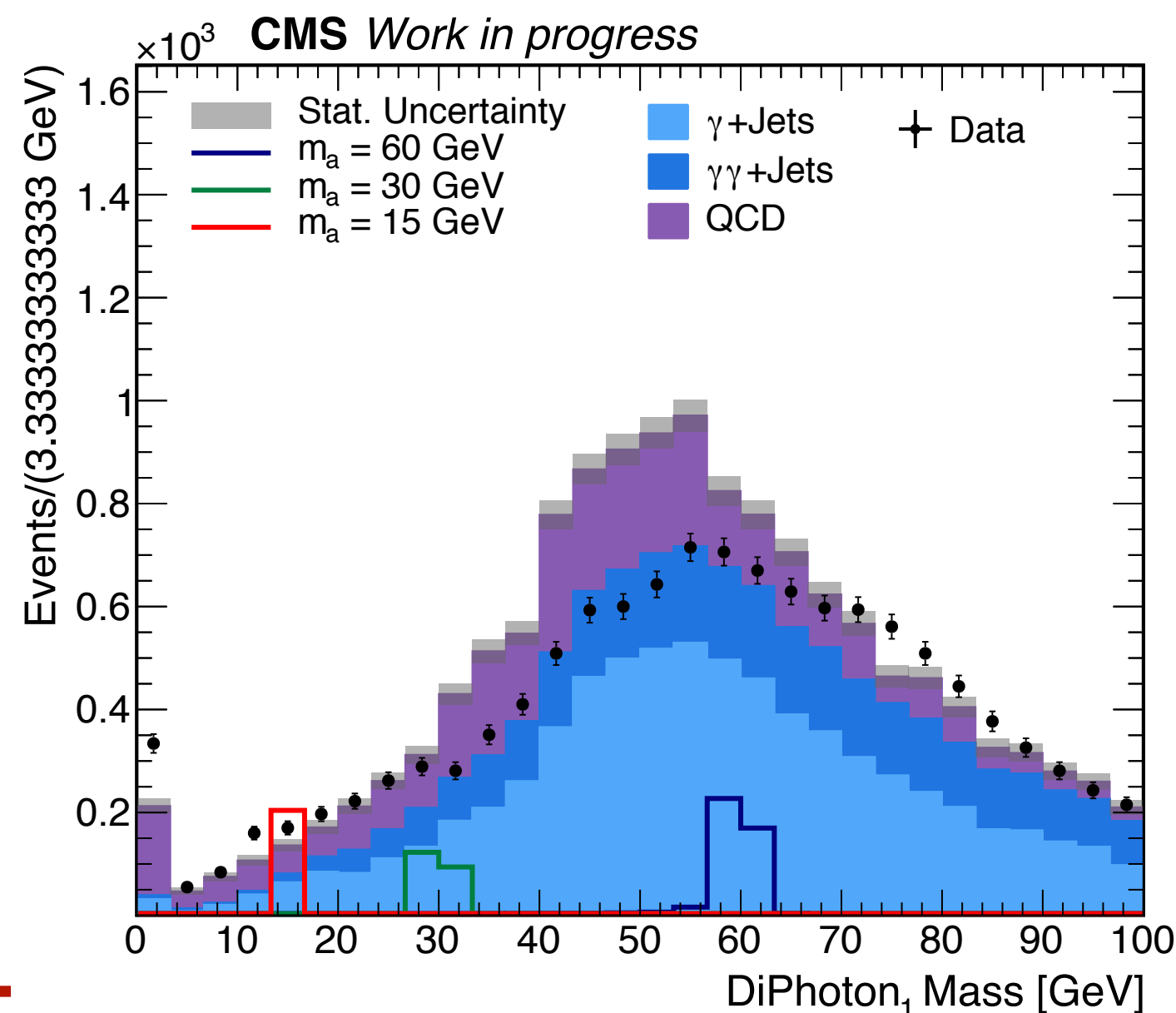
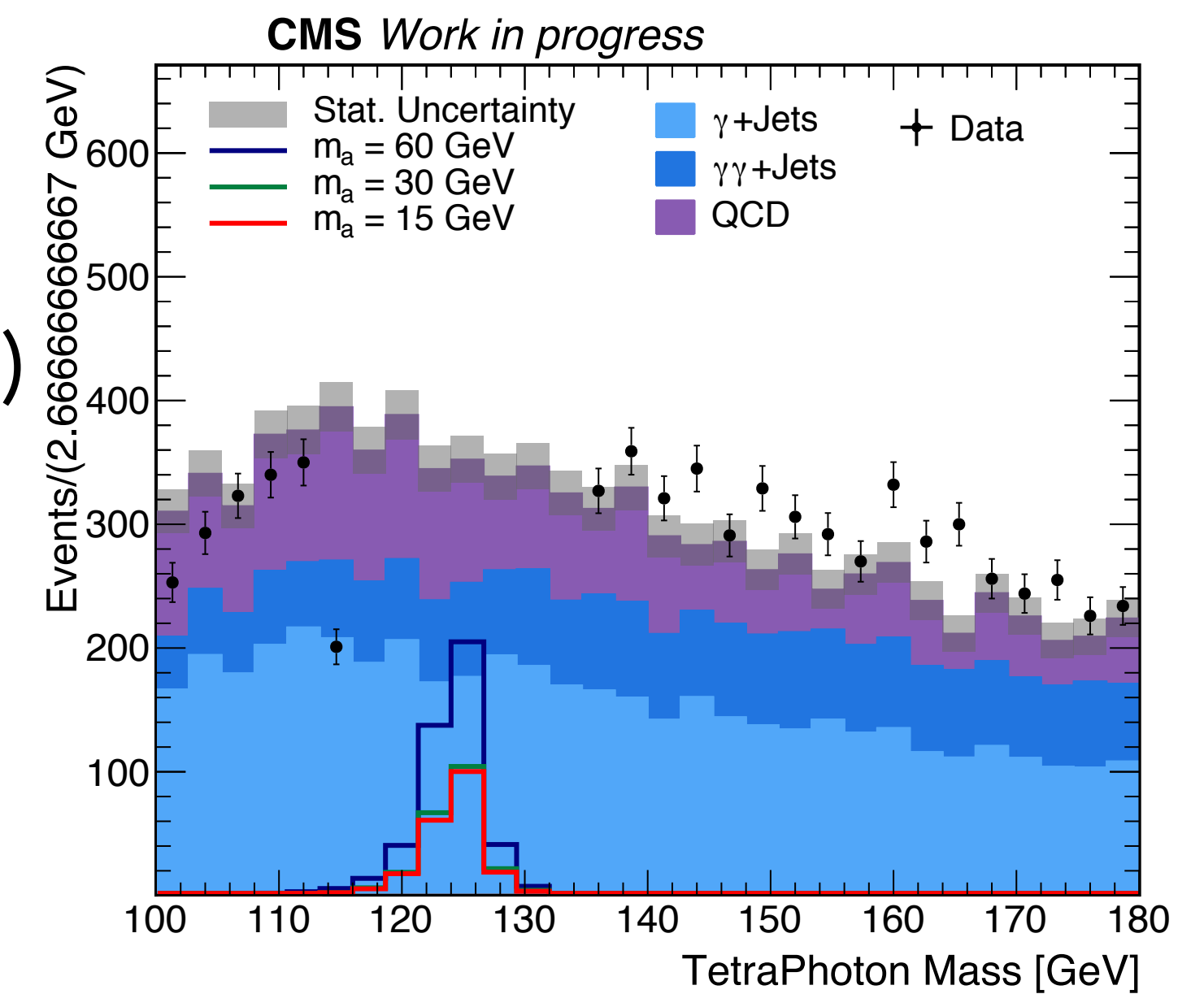
# Object Selection

- Require that at least one di-photon passes the pre-selection
- Photons selection
  - $p_T(\gamma_{1,2}) > 30, 20 \text{ GeV}$  and  $p_T(\gamma_{3,4}) > 10, 10 \text{ GeV}$
  - $|\eta(\gamma's)| < 2.5$  and not within the EB-EE gap
- [Signal MC]:  $\gamma$ 's are gen-matched (geometrical matching based on  $\Delta R$ )
  - In case of events w/  $> 4$   $\gamma$ 's @reco-level, the gen-matched  $\gamma$ 's are selected
- [Data and Background MC]: In case of events w/  $> 4$   $\gamma$ 's, pick the 4 highest  $p_T$   $\gamma$ 's



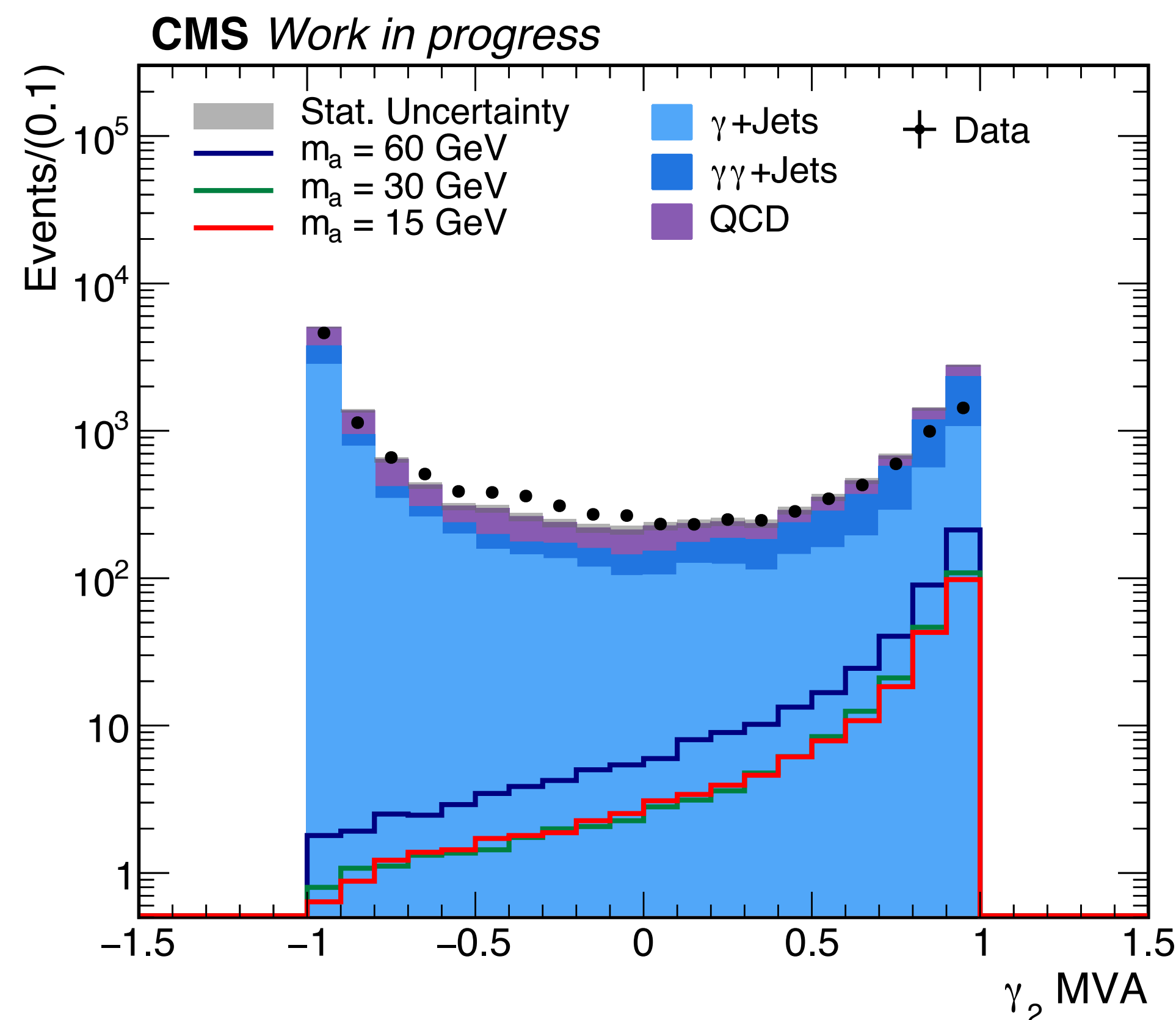
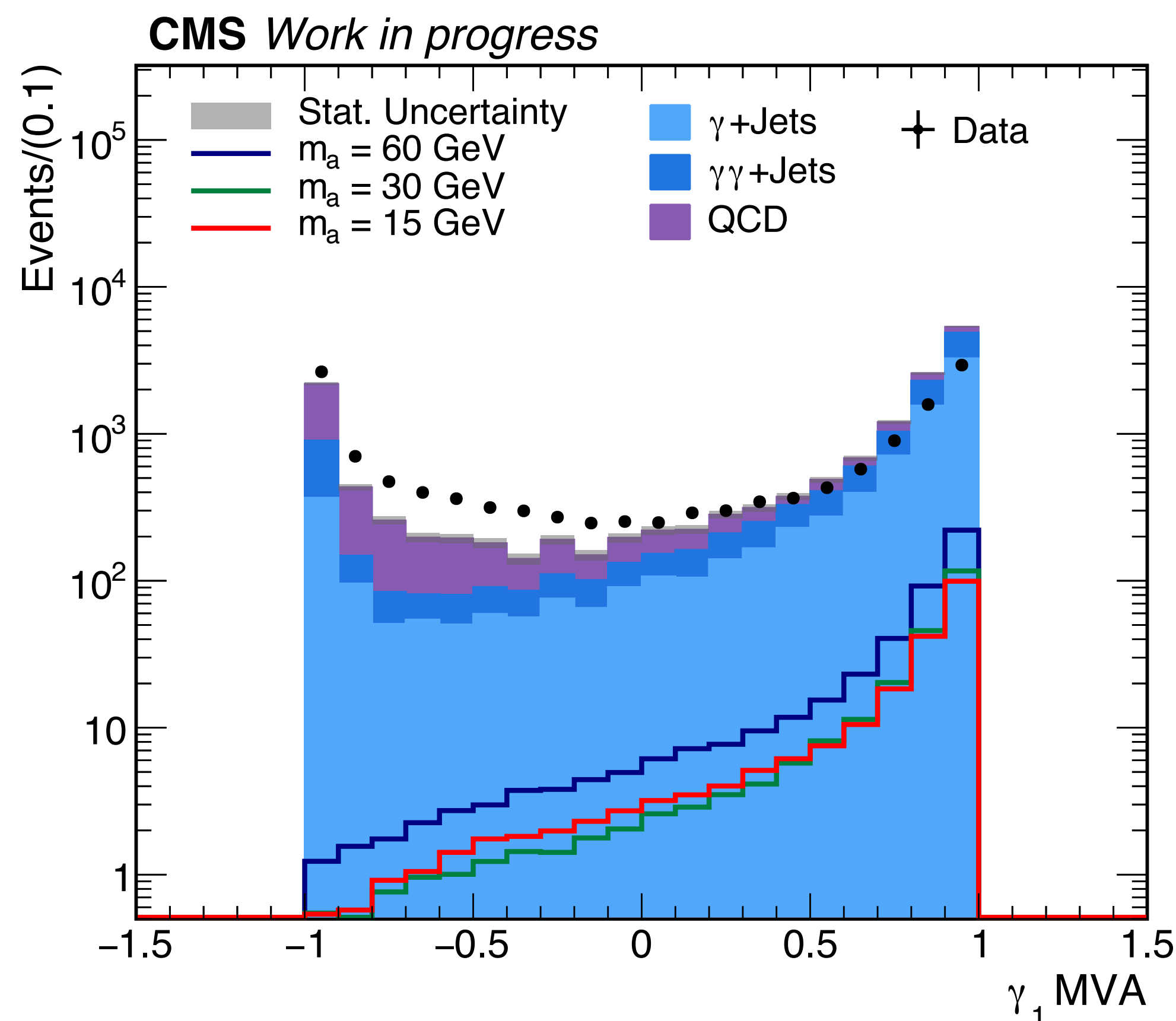
# Preselection Level Plots

- Data/MC comparison with 36/fb of 2016 data
- After applying pre-selection
- Data/MC does not match right now
  - Found a bug in my ntuple producer (wrong Generator level weights)
  - Photon scales and smearing are not applied here
  - Submitted new jobs with fix (will update with corrected plots)
- Blinding region in  $M(\gamma\gamma\gamma\gamma) = [115, 135]$  GeV (removed from the plots)
- Reminder: background MC not used for signal extraction

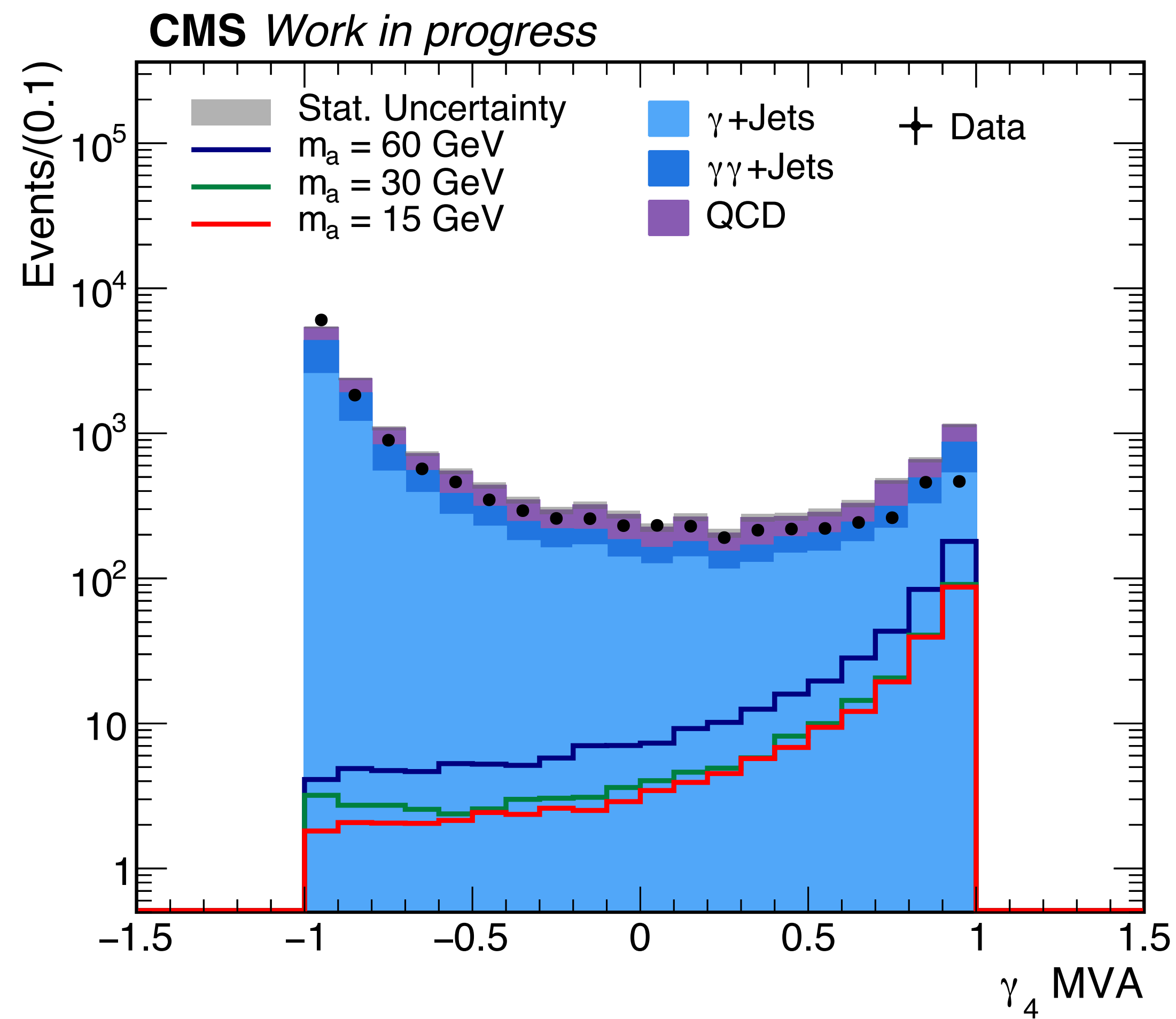
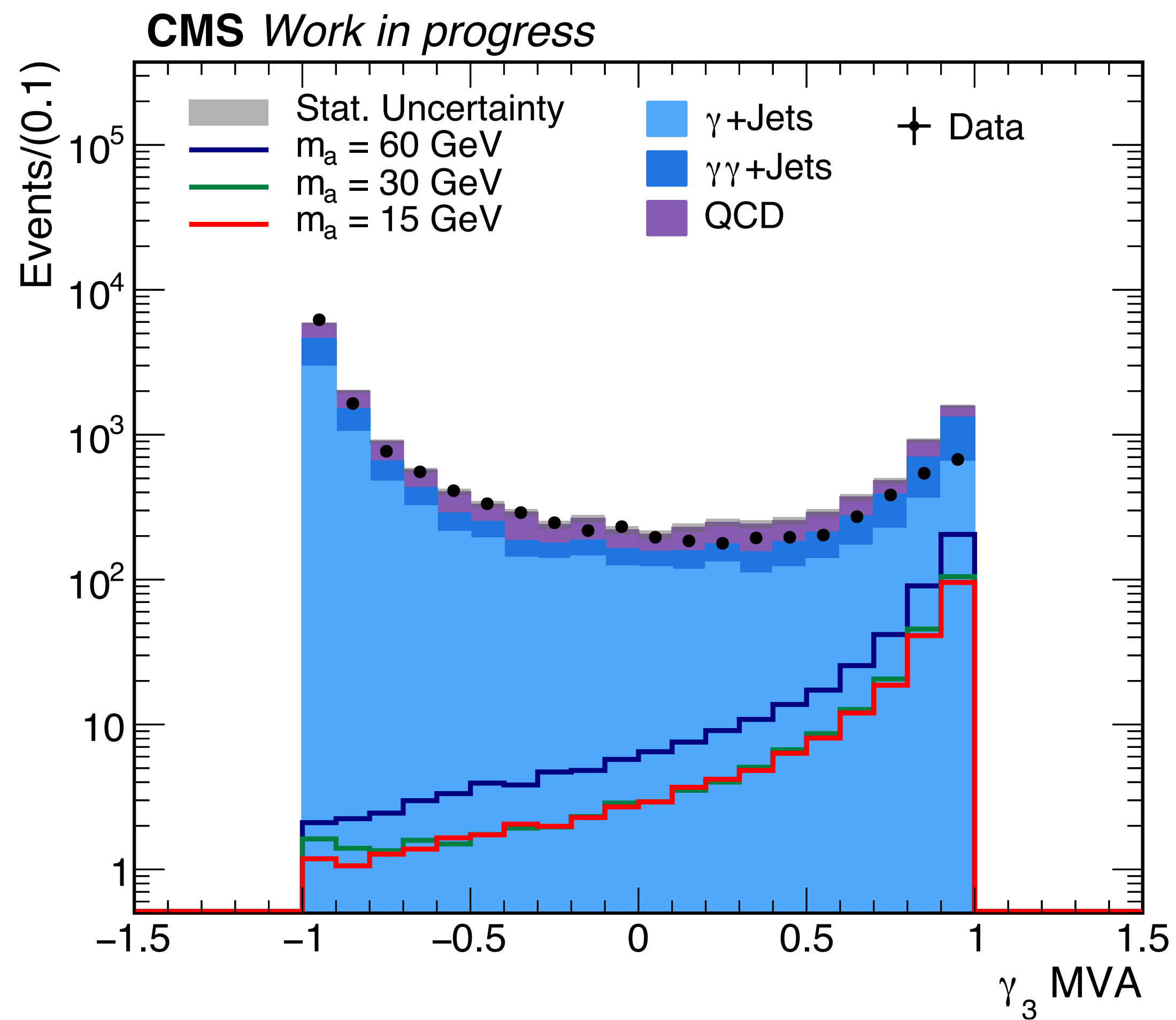


# MVA selection

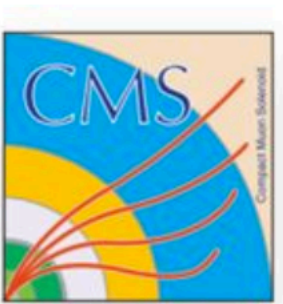
- Removed the loose MVA selection (at  $> -0.9$ ) from the preselection cfi file
- MVA distributions of the  $4\gamma$ 's at the pre-selection level ( $\gamma$ 's are  $p_T$  ordered)
- The idea is to determine a set of selections on the MVA score of the  $4\gamma$ 's s.t we keep a high signal efficiency while rejecting the maximum background, across all signal mass points



# MVA selection







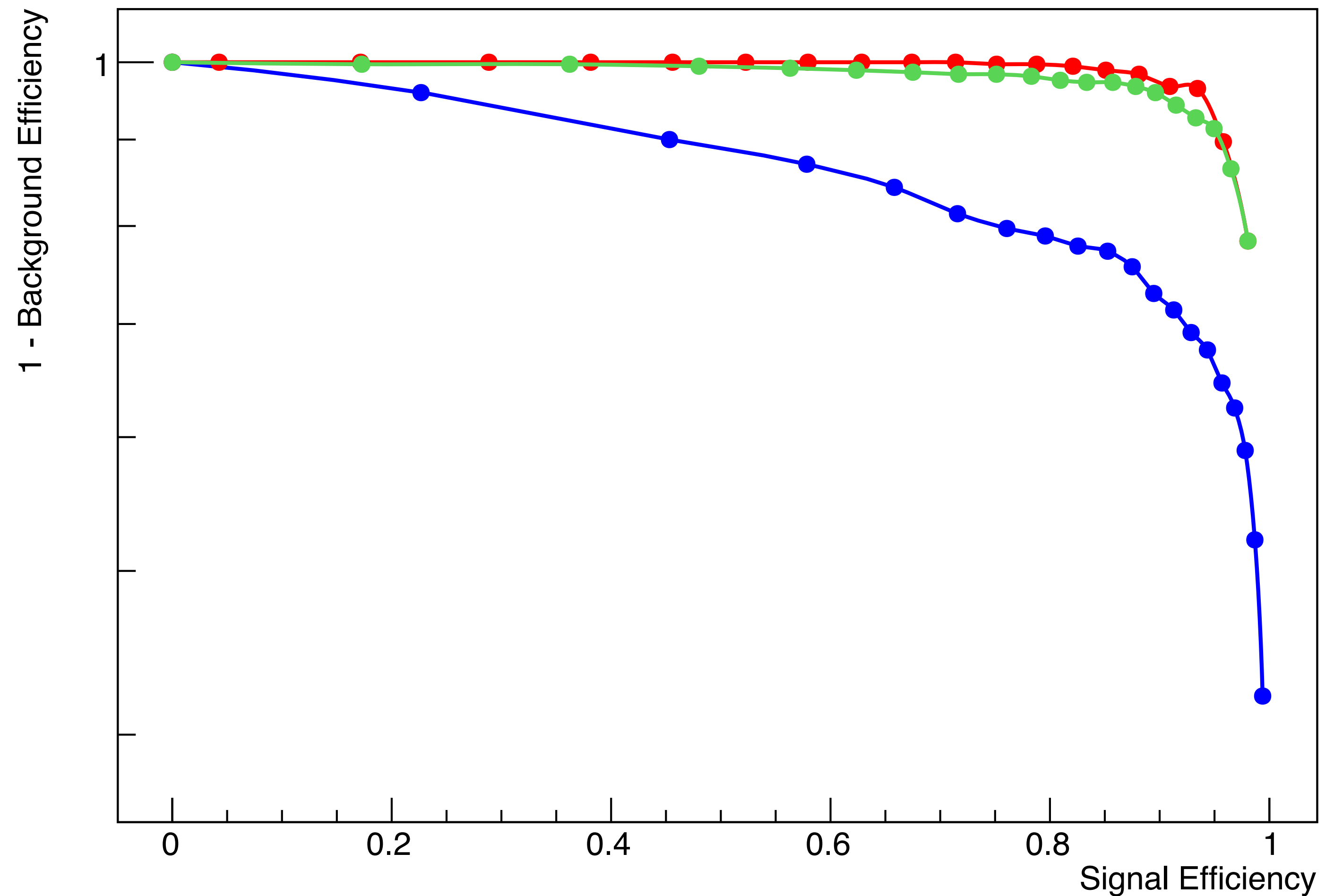
# MVA selection

- Checked three different types of selections
  - **Selection 1**: Cut on the MVA score of the first 2 highest  $p_T \gamma$ 's (starting from cut at -0.9, increment the cut value)
  - **Selection 2**: Cut on MVA score of all 4  $\gamma$ 's
  - **Selection 3**: Keep a loose cut on the first 2 highest  $p_T \gamma$ 's and keep incrementing the cut value for the 3<sup>rd</sup> and 4<sup>th</sup>  $\gamma$
- Check the background rejection, first for all backgrounds separately and then for the combined background

# MVA selection

## QCD

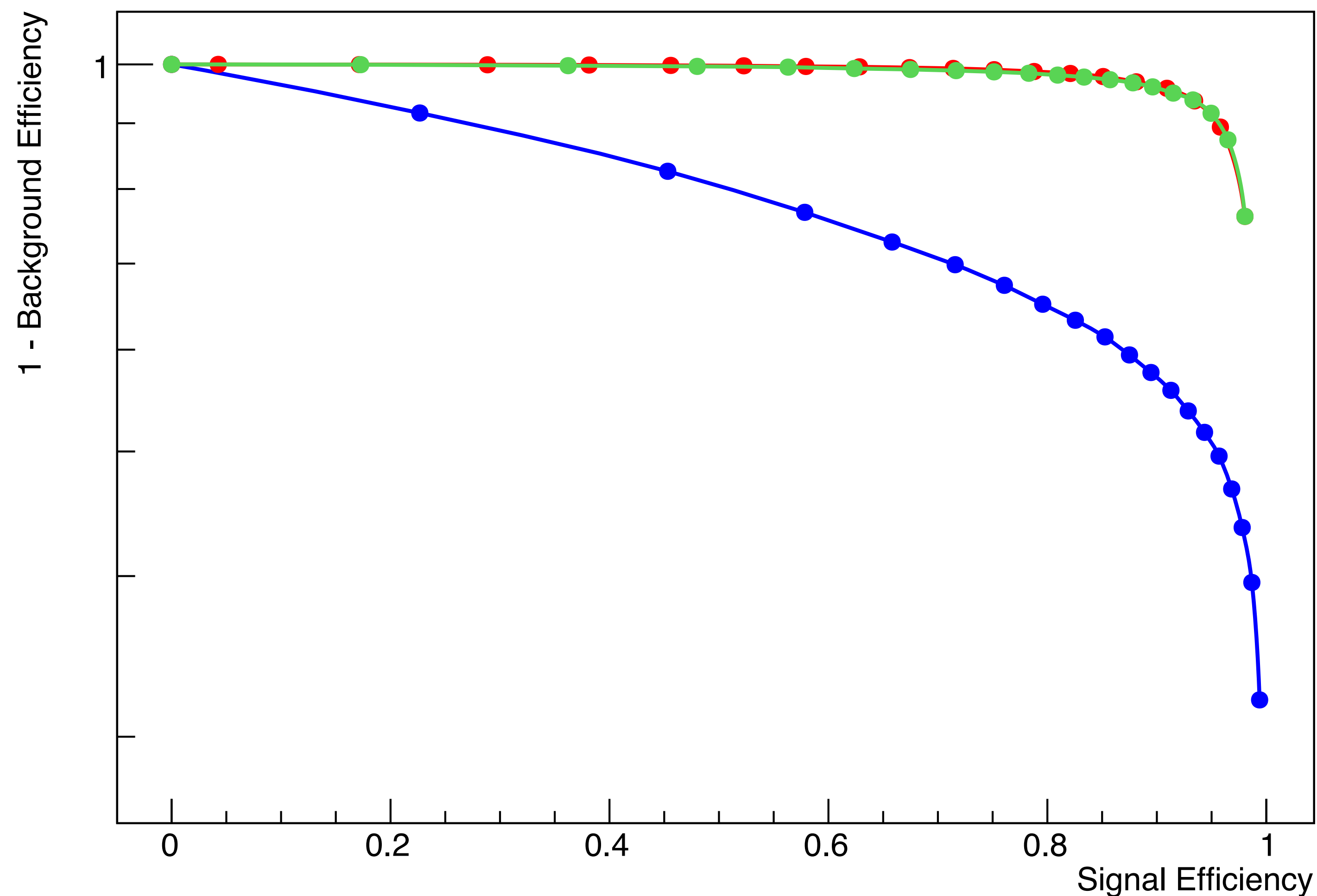
- **Selection 1:** Cut on the MVA score of the first 2 highest  $p_T$   $\gamma$ 's (starting from cut at -0.9, increment the cut value)
- **Selection 2:** Cut on MVA score of all 4  $\gamma$ 's
- **Selection 3:** Keep a loose cut on the first 2 highest  $p_T$   $\gamma$ 's and keep incrementing the cut value for the 3<sup>rd</sup> and 4<sup>th</sup>  $\gamma$



# MVA selection

## $\gamma$ + Jets

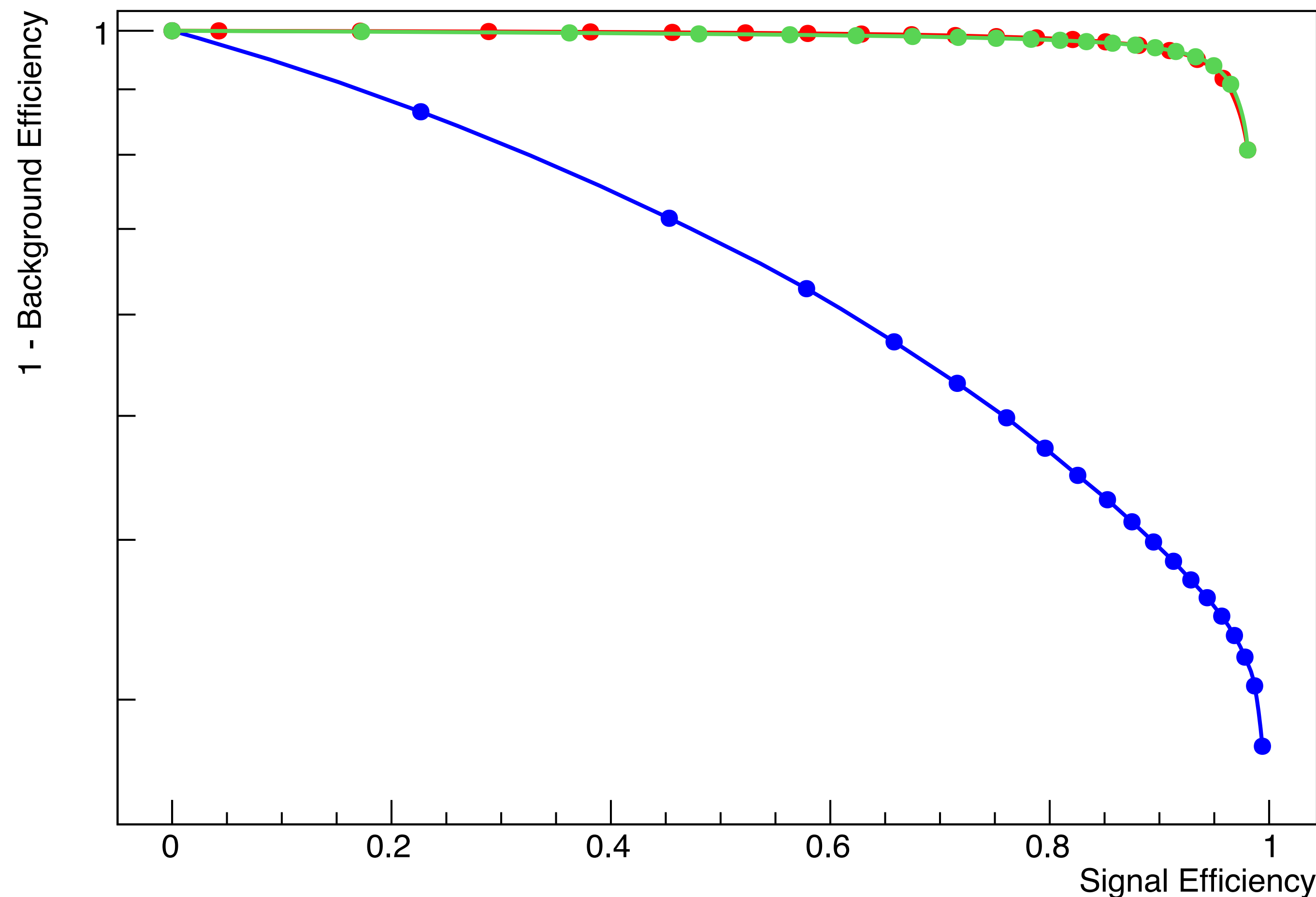
- **Selection 1:** Cut on the MVA score of the first 2 highest  $p_T$   $\gamma$ 's (starting from cut at -0.9, increment the cut value)
- **Selection 2:** Cut on MVA score of all 4  $\gamma$ 's
- **Selection 3:** Keep a loose cut on the first 2 highest  $p_T$   $\gamma$ 's and keep incrementing the cut value for the 3<sup>rd</sup> and 4<sup>th</sup>  $\gamma$



# MVA selection

## Di- $\gamma$ + Jets

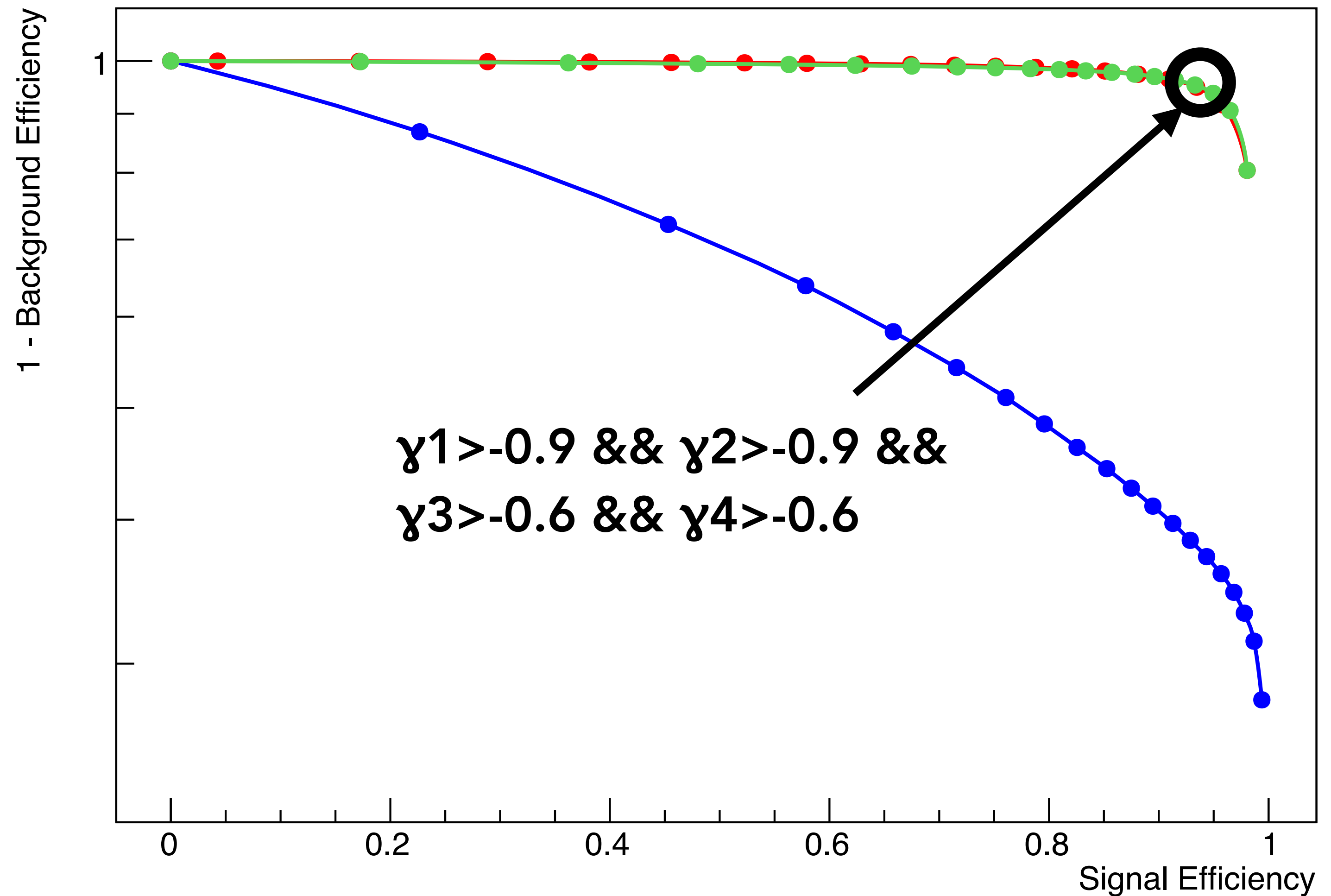
- **Selection 1:** Cut on the MVA score of the first 2 highest  $p_T$   $\gamma$ 's (starting from cut at -0.9, increment the cut value)
- **Selection 2:** Cut on MVA score of all 4  $\gamma$ 's
- **Selection 3:** Keep a loose cut on the first 2 highest  $p_T$   $\gamma$ 's and keep incrementing the cut value for the 3<sup>rd</sup> and 4<sup>th</sup>  $\gamma$



# MVA selection

## All background combined

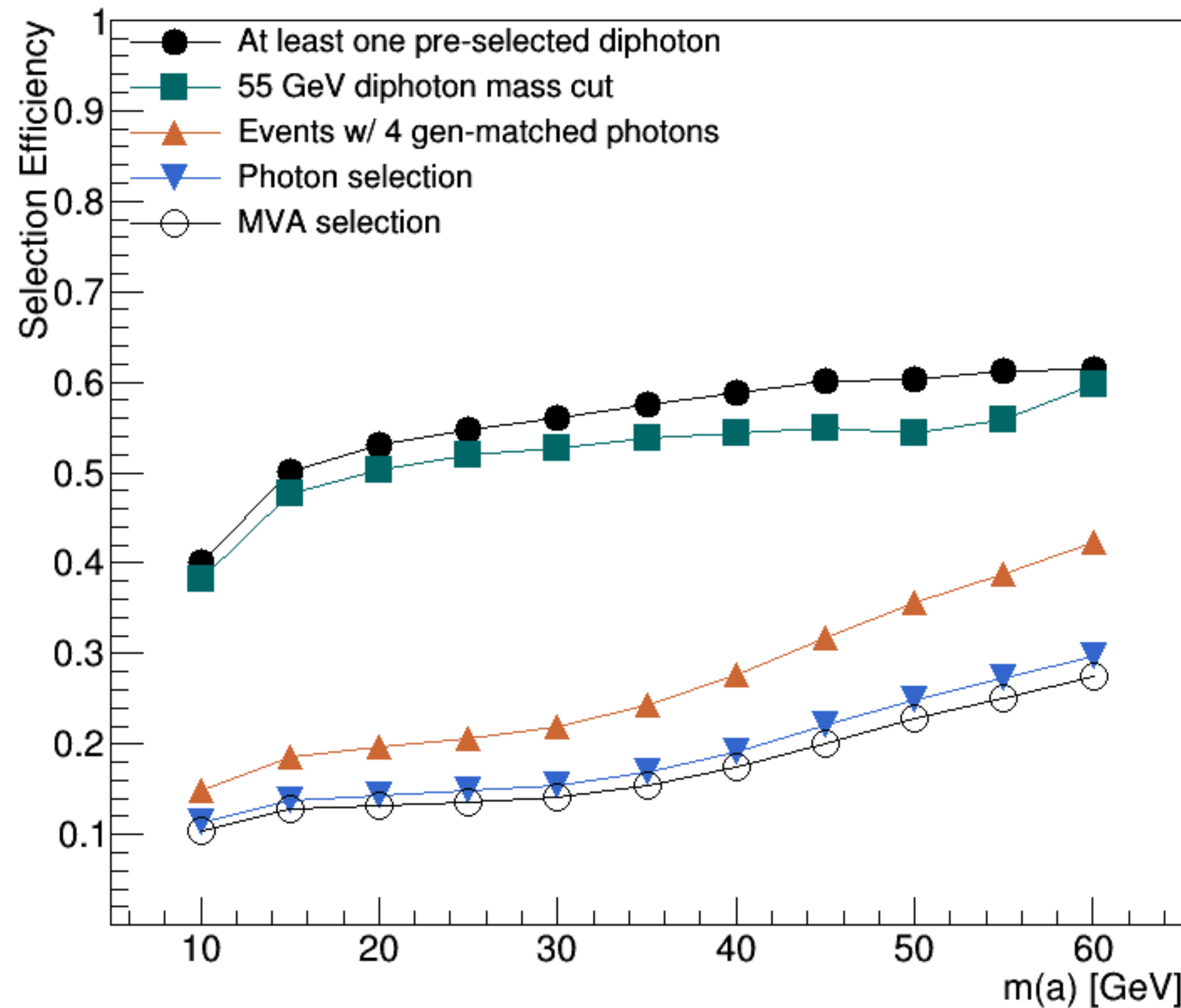
- Applying this selection keeps 93% of the signal efficiency while rejecting 92% of the background
- Question: Are there scale factors available for a different value of the selection?





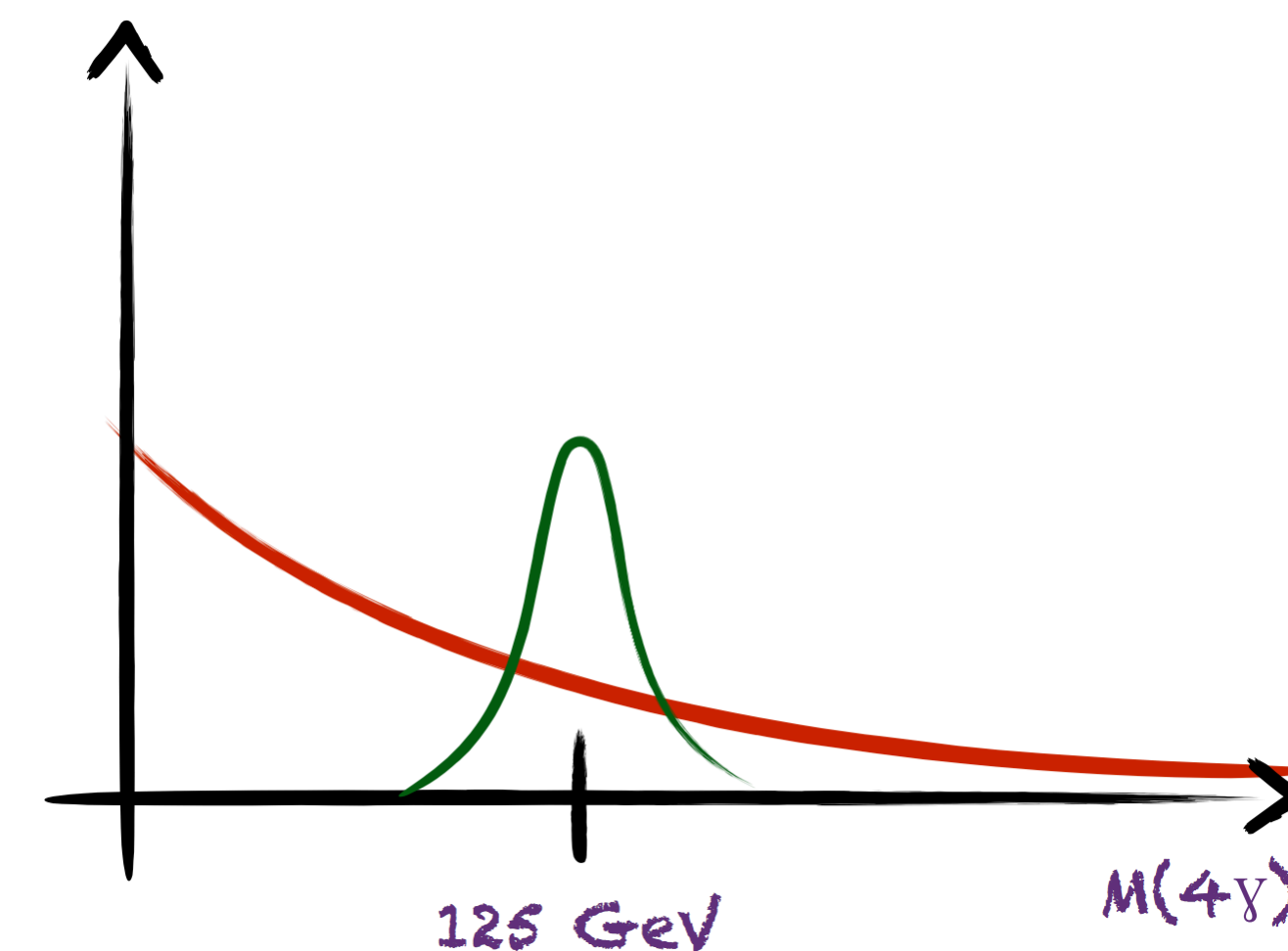
# Signal Efficiency

- Signal efficiency after the different selection steps



# Analysis Strategy

- Can explore 2 possibilities:
- Strategy 1:
  - Signal extraction by parametric fit in  $M(4\gamma)$
  - According to signal mass hypothesis, select mass window in  $M(2\gamma)$  and fit the  $M(4\gamma)$  distribution
- Strategy 2:
  - From the data sidebands of the  $M(4\gamma)$  distribution, perform a signal extraction by parametric fit in  $M(2\gamma)$
  - In this case, it is important to ensure that the  $M(2\gamma)$  signal distribution has a narrow peak and few events in the tail (otherwise, the pdf is not gaussian)
  - Important to optimize diphoton pairing in this case



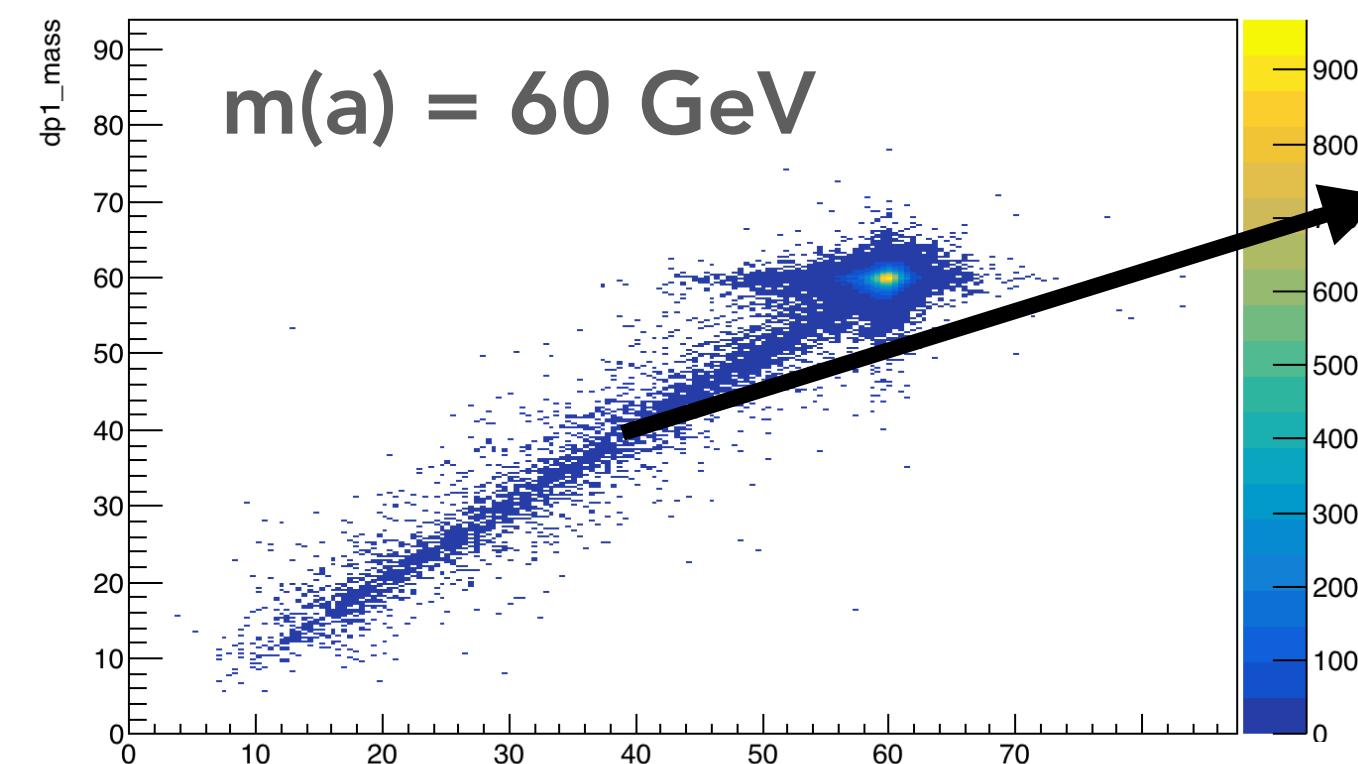
# DiPhoton Pairing

- Once an event passes the preselection requirements, diphoton pairs are created from the 4  $\gamma$ 's in the event ( $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ )
- Three unique combinations of diphoton pairs ( $\gamma_{12}, \gamma_{34}$ ), ( $\gamma_{13}, \gamma_{24}$ ), ( $\gamma_{14}, \gamma_{23}$ ); out of these combinations at most one pairing correctly maps the true pseudo scalar "a" decay products
- Currently, pairing is done such that the diphoton mass difference is minimum  

$$\Delta M = |m_{\text{diphoton1}} - m_{\text{diphoton2}}|$$

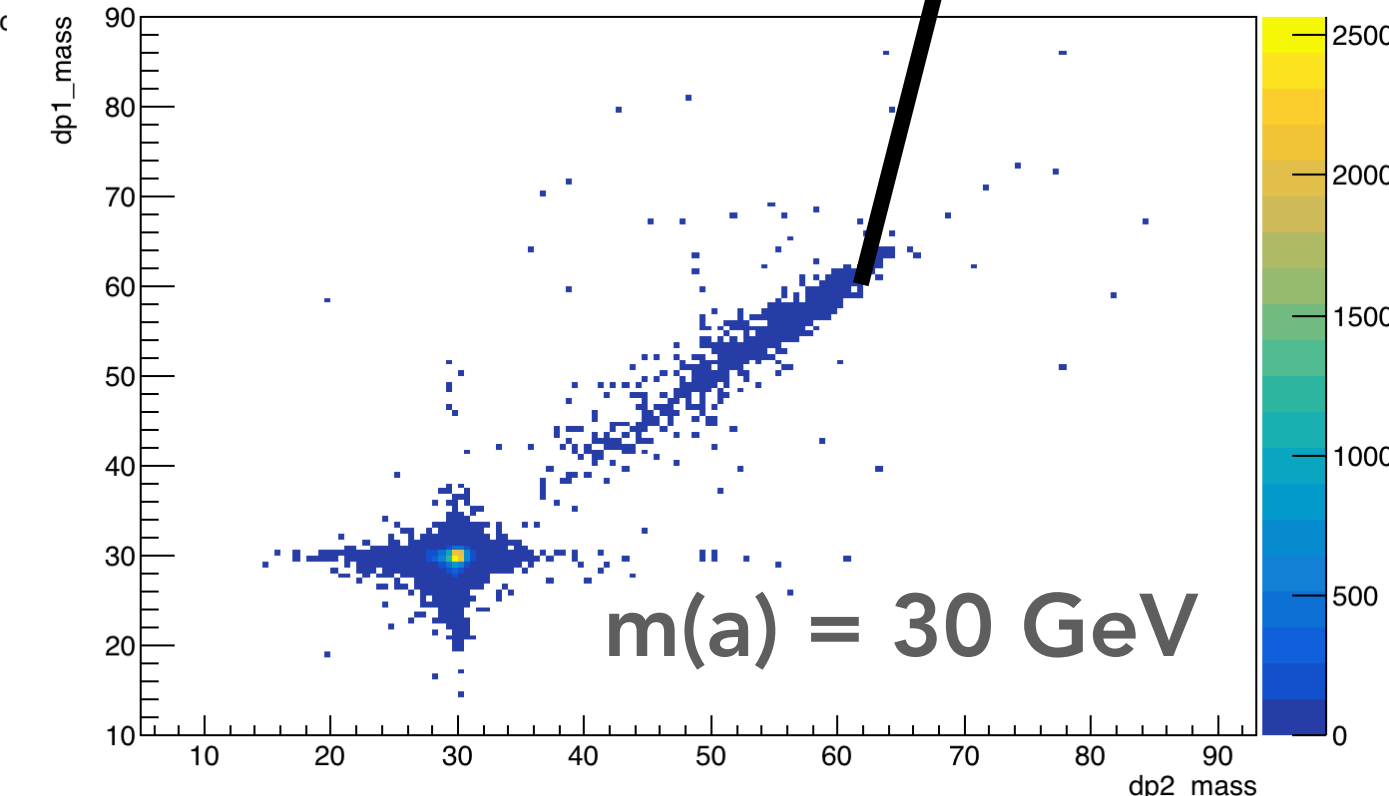
```
## Analyzing event 79
0 1 Pair1 mass 60.7878757267 2 3 Pair2 mass 27.1668615445
0 2 Pair1 mass 44.2490779655 1 3 Pair2 mass 45.9583165151
0 3 Pair1 mass 61.281119926 1 2 Pair2 mass 59.5444302031
Chosen pair: Diphoton1 0 2 Diphoton2 1 3
```

- For e.g : Even though the 3rd pair is the correct one, the pair chosen in this event was the 2nd one (because of minimum  $\Delta M$ )

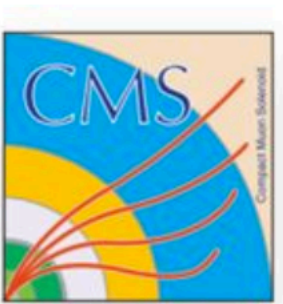


**Events in the tail**  
 (~10% of the total events across  $m(a)$ )

**Diphoton<sub>1</sub> Vs Diphoton<sub>2</sub> Mass distribution**

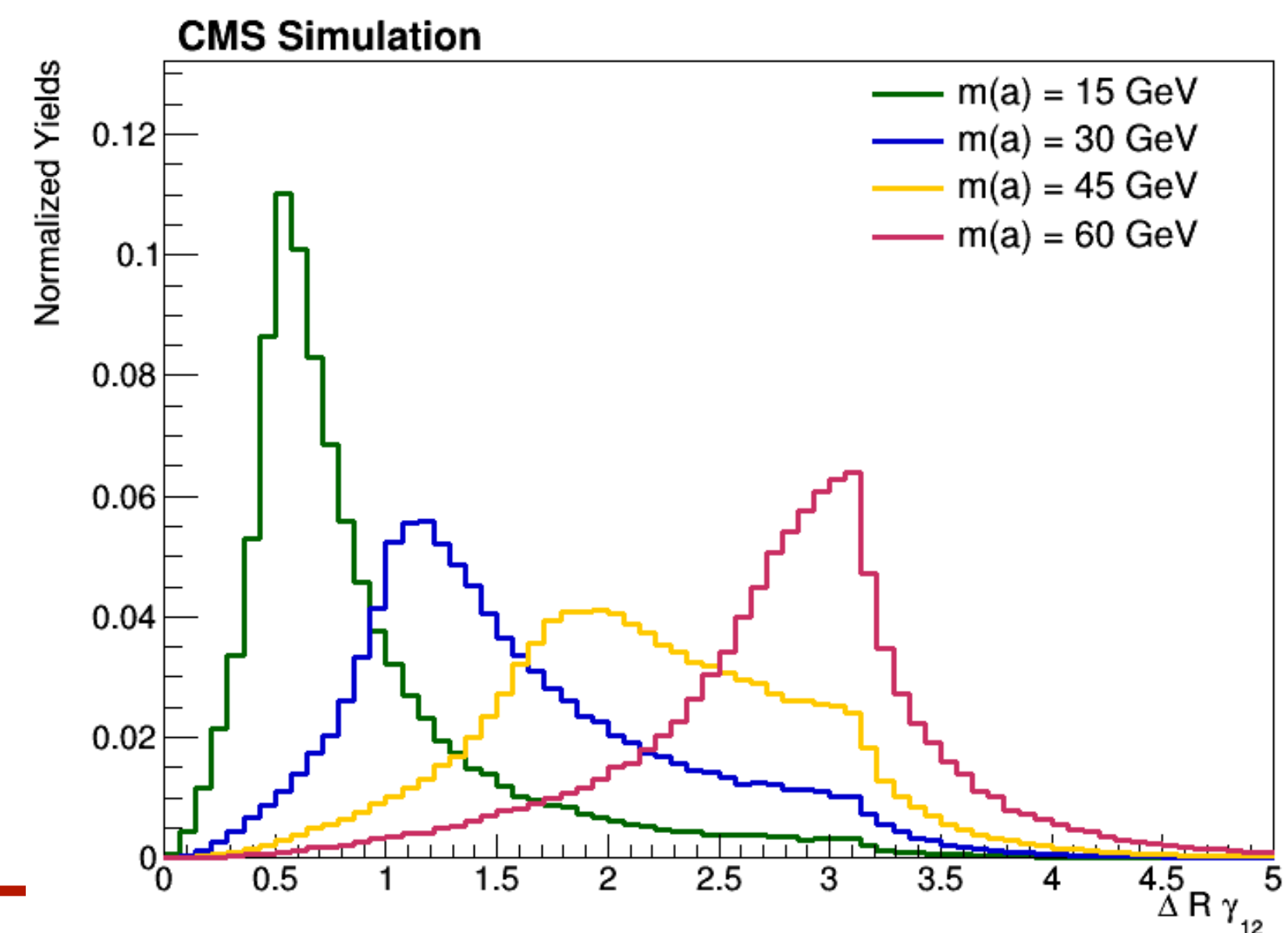
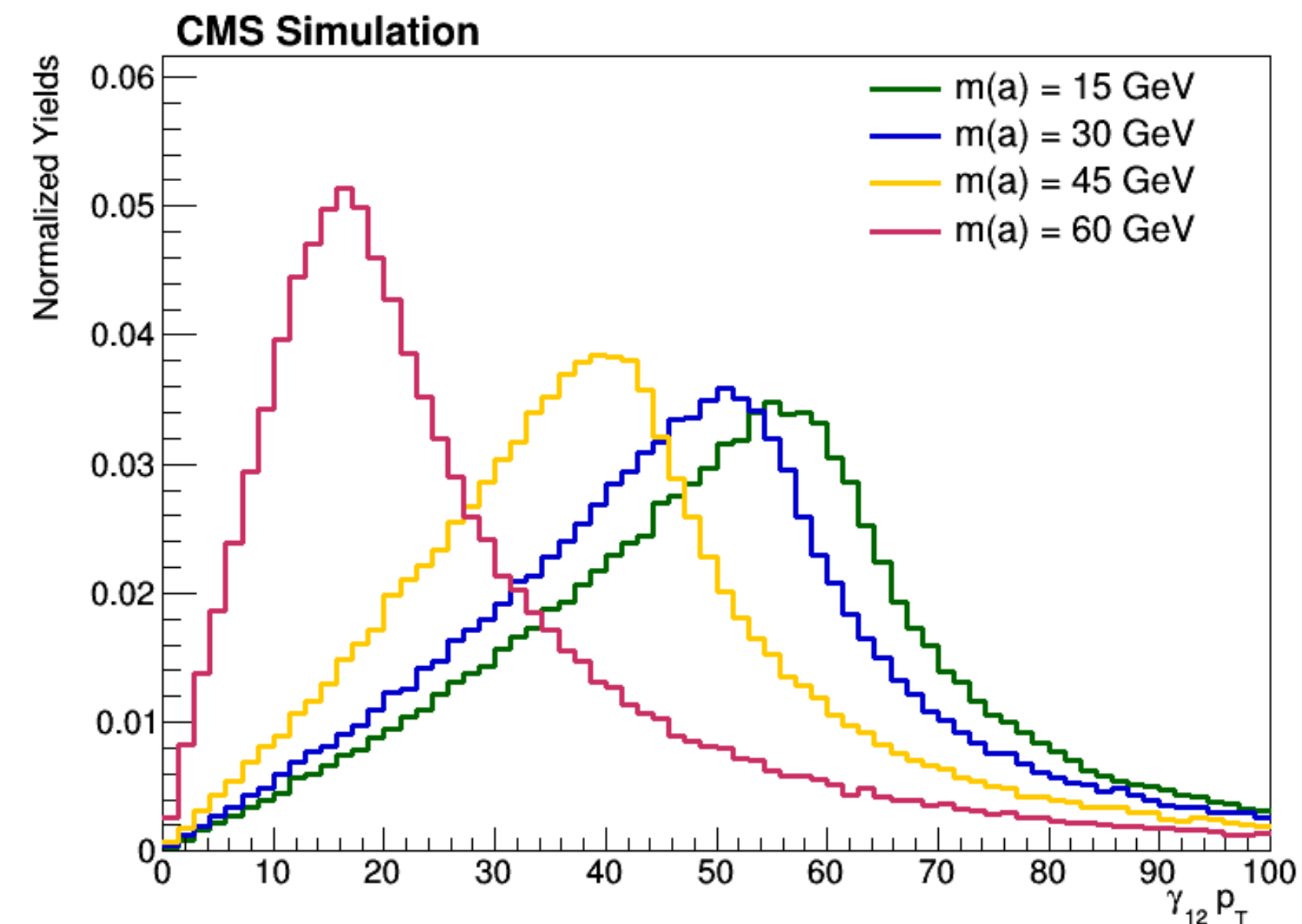






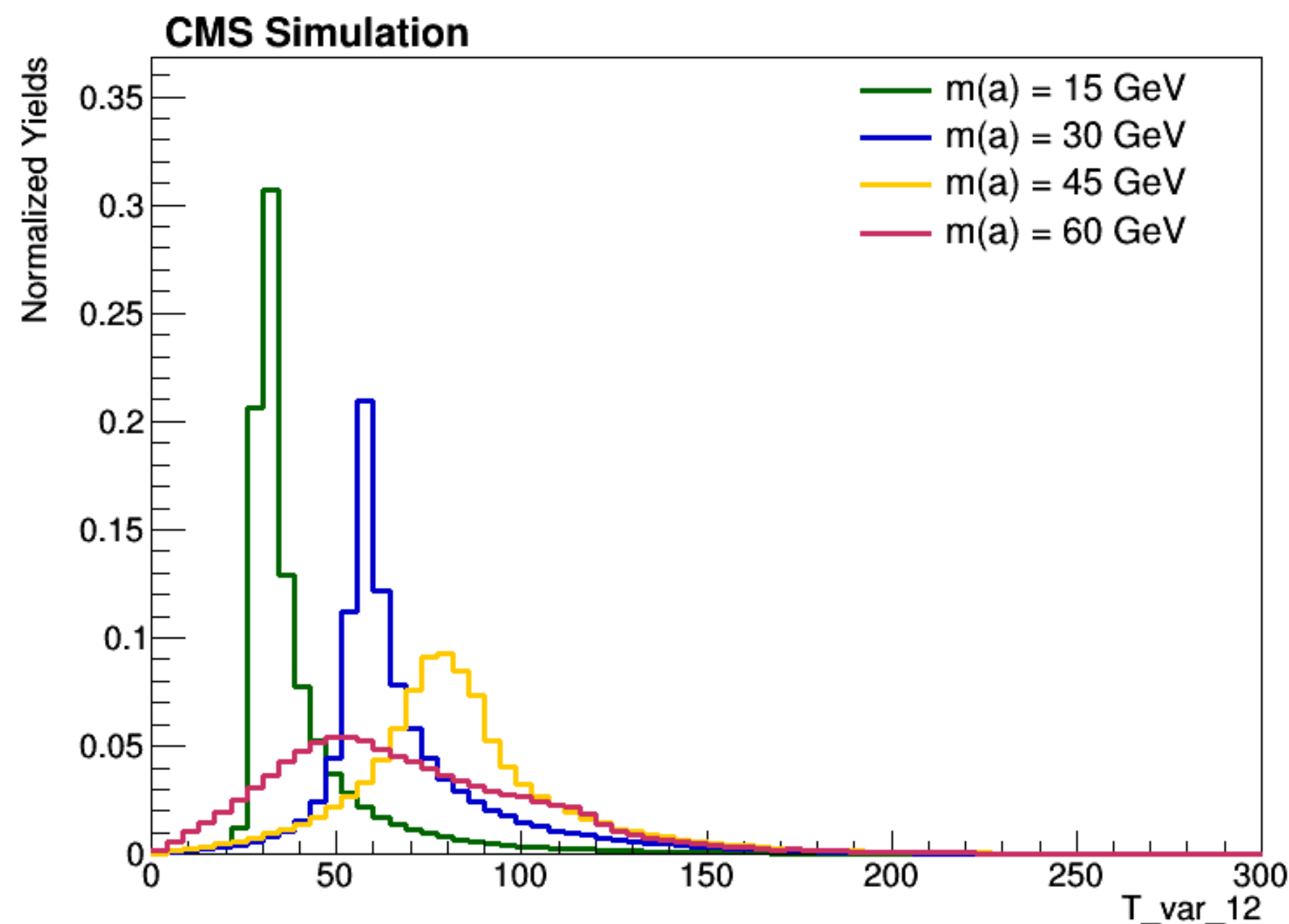
# DiPhoton Pairing

- Want to minimize the number of events in the tail
- Alternative pairing method?
- Look at Gen-Level distribution
- Top:  $p_T$  distribution of "a"
- Bottom:  $\Delta R$  distribution of  $\gamma$ 's coming from the same "a"
- For  $m(a) = 60$  GeV, the "a's" will be almost back-to-back, and its product  $\gamma$ 's will be well separated in  $\Delta R$ ; as we go down in  $m(a)$ , the "a" is more boosted and its product  $\gamma$ 's are geometrically closer to each other
- Define a variable:
  - $\Delta = (p_T^a) \times (\Delta R_{\gamma 1, \gamma 2})$

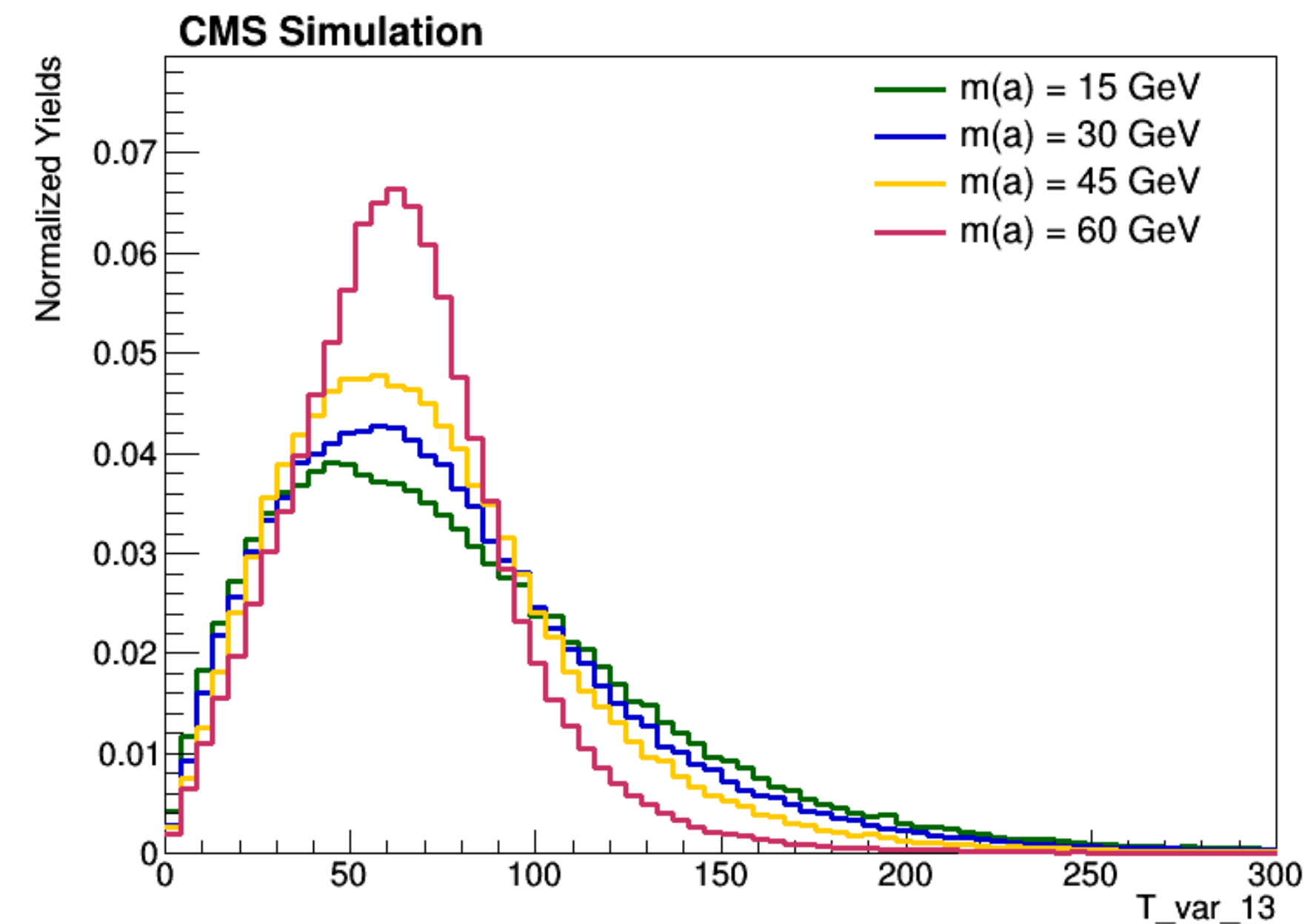


# DiPhoton Pairing

- Define a variable:
  - $\Delta = (p_T^a) \times (\Delta R_{\gamma 1, \gamma 2})$
- Try to minimize "Delta" to pick correct diphoton pair
- But, across the masses, this variable does not have the best discriminating power
- Trying different pairing methods:
  - Minimize  $\frac{\Delta m}{\Delta R_{\gamma 12} + \Delta R_{\gamma 34}}$
  - Minimize  $\frac{\Delta m}{p_T(a1) + p_T(a2)}$



Delta distribution for correct pair



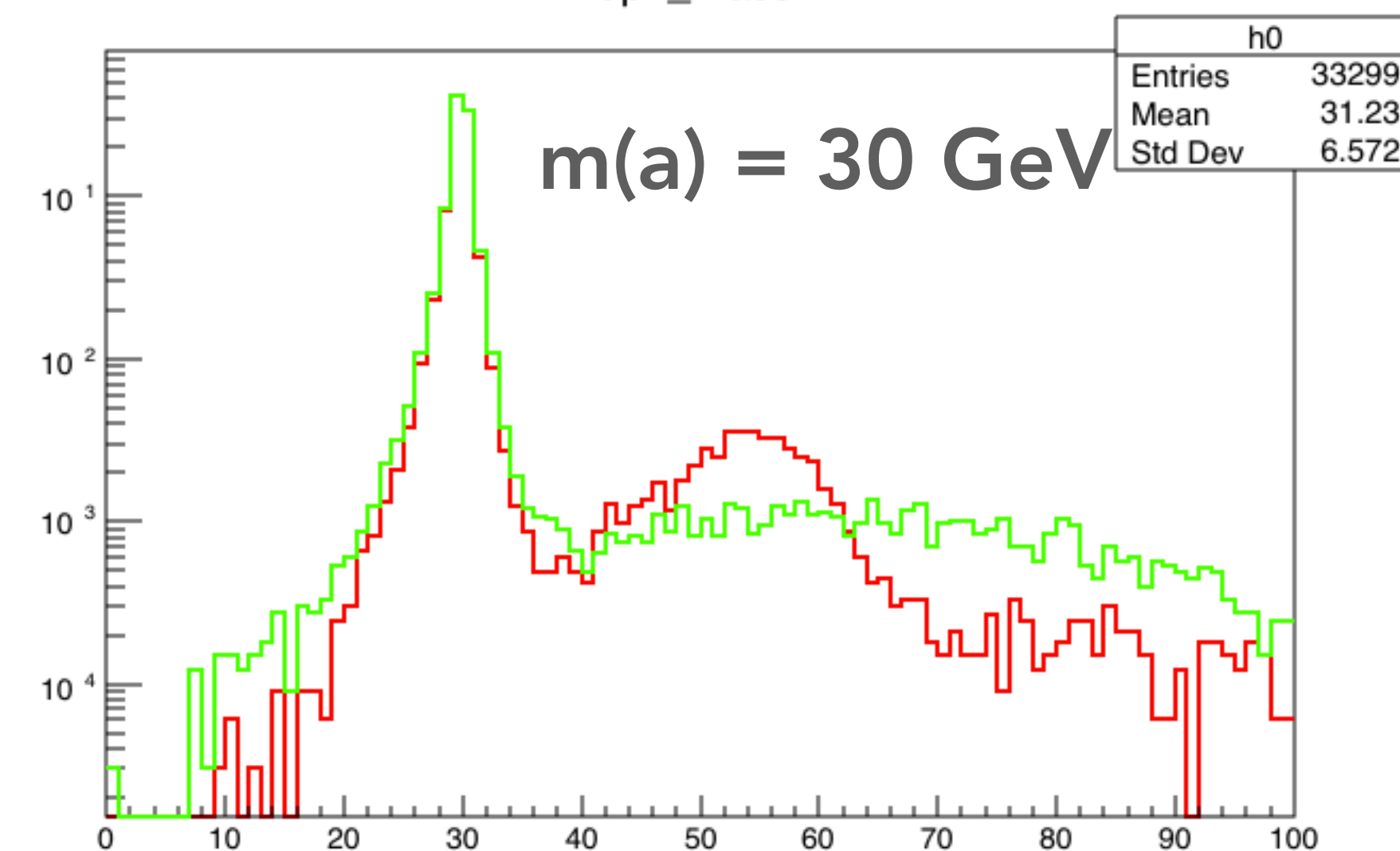
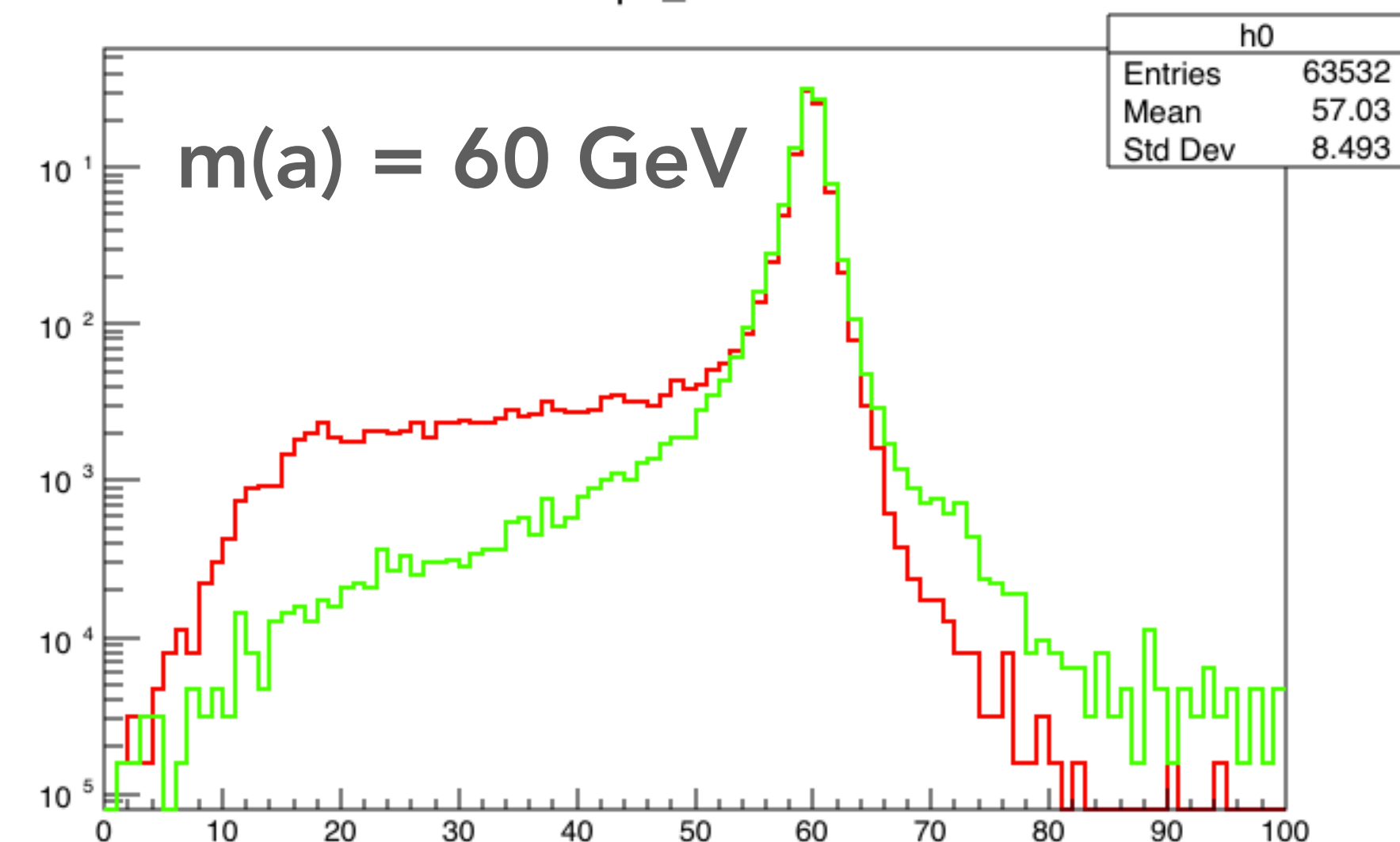
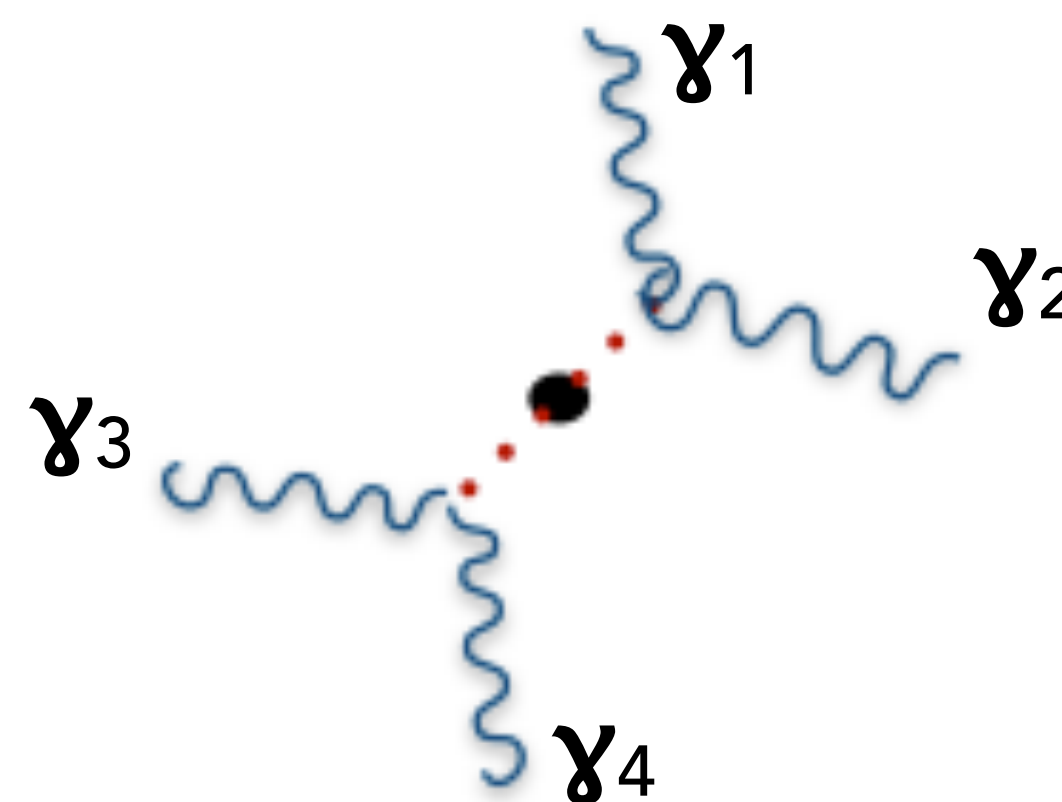
Delta distribution for wrong pair

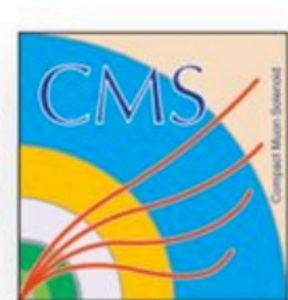
- So far, the  $\Delta M$  minimization pairing has the best performance across the entire  $m(a)$  range



# DiPhoton Pairing

- More checks about the pairing
- **Red:** Pairing according to minimization of  $\Delta M$
- **Green:** Pairs made from gen-matched  $\gamma$ 's
- The gen-matched pairing does not show improvement
- How is Gen-Reco pairing done?
  - For every Gen-level  $\gamma$ , look for the Reco-level  $\gamma$  closest in  $\Delta R$
- Hypothesis: For the case, where all  $\gamma$ 's are resolved, the min.  $\Delta R$  method could be picking up  $\gamma$  from the wrong "a"
- Possible fix: For pairing, for the gen  $\gamma$ , look for reco  $\gamma$ 's within a cone of  $\Delta R = 0.15$ 
  - If more than one reco  $\gamma$  is found within the cone, then pair by minimization of  $\Delta p_T$
  - Testing this hypothesis now



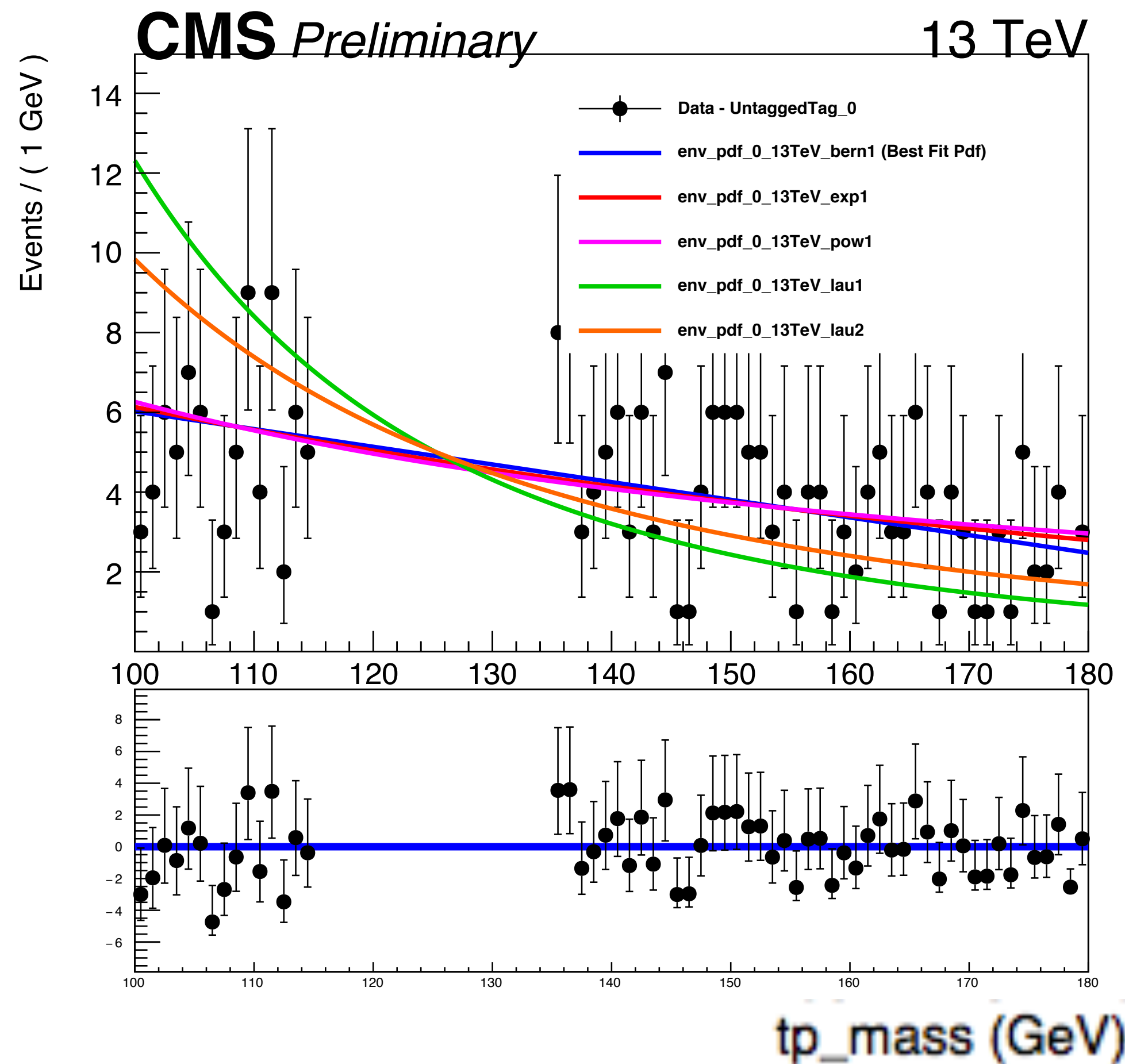


# Moving forward..

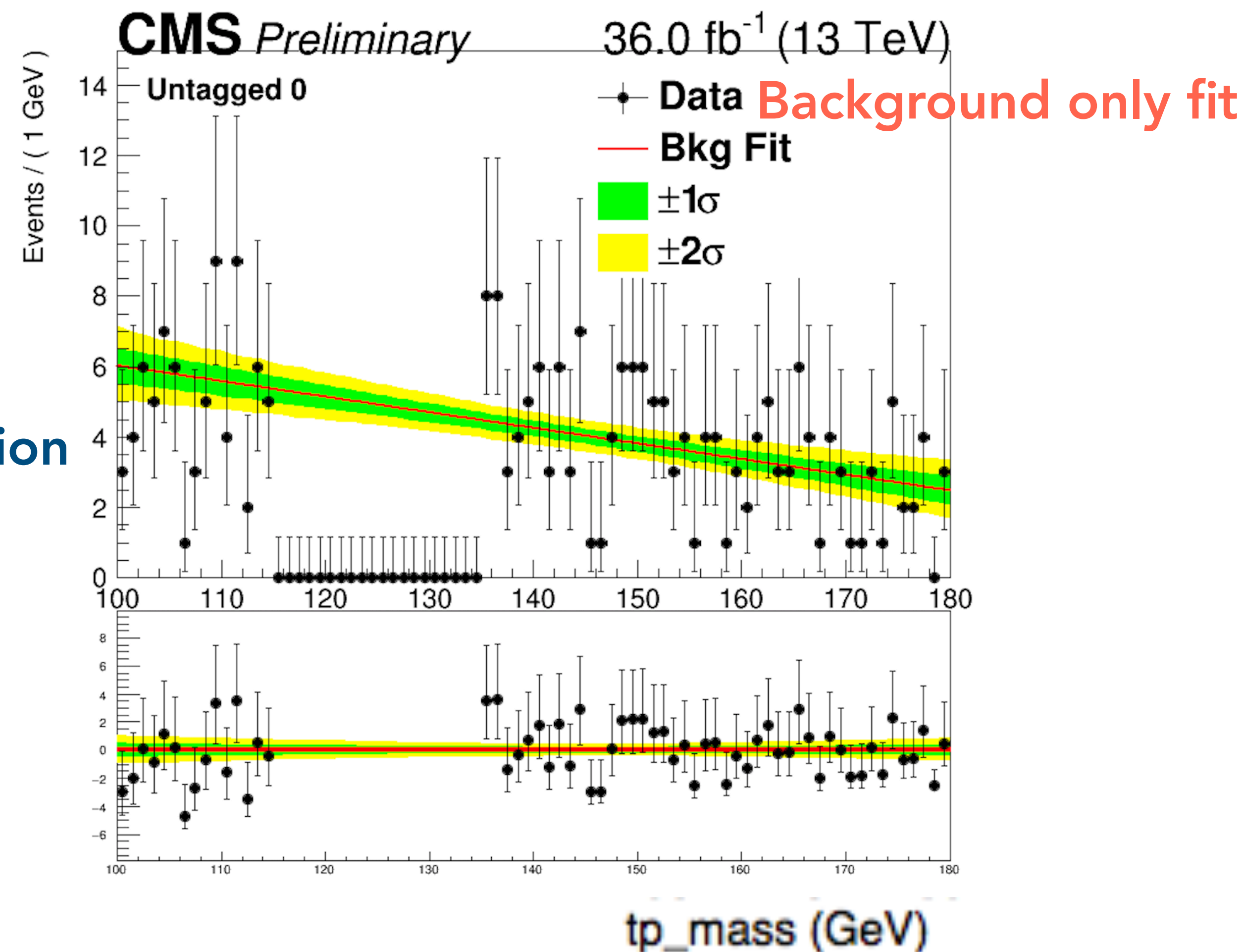
---

- It's important to understand the di-photon pairing issue, to be able to scan  $M(2\gamma)$
- On the following slides, for the background and signal model fitting, I have not used any selection on the di-photon object (the inclusive  $M(4\gamma)$  distribution i.e, without using any di-photon mass window)
- Thanks to Ed and Yacine for their help with flashggFinalFit machinery!

# Background Model



$M(4\gamma)$   
distribution

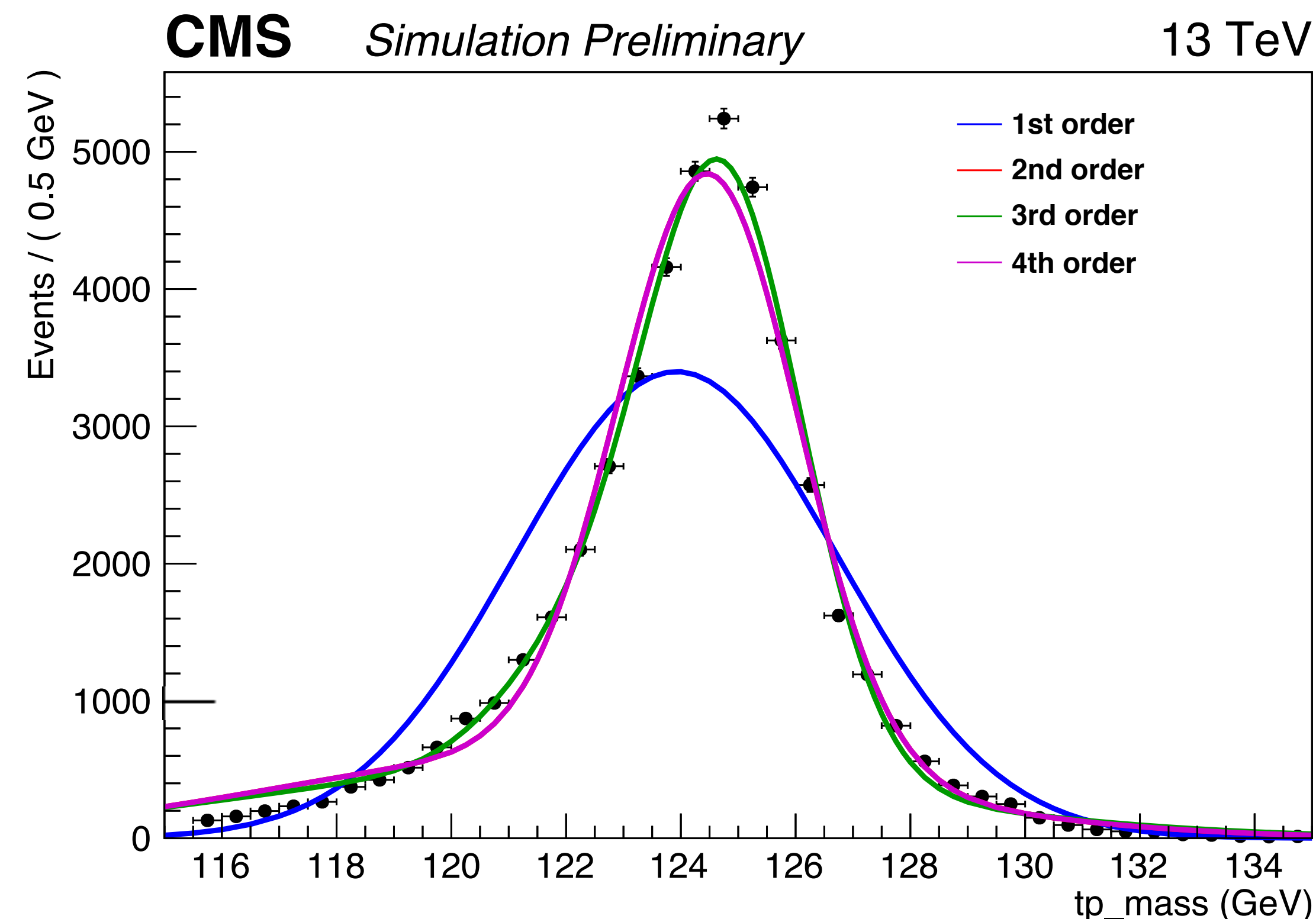


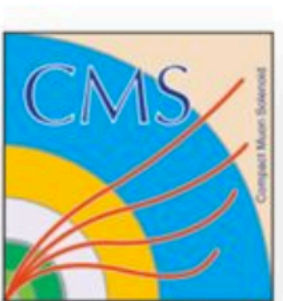
- Without any categorization or any mass window cut on  $M(2\gamma)$ , inclusive  $M(4\gamma)$  distribution
- Order 1 Bernstein polynomial chosen as the best fit PDF



# Signal Model

- For the signal model, the first step is to determine how many gaussians should be used to fit the  $M(4\gamma)$  distribution
- The fit did not fail here, but the low end tail of the distribution is not being modeled properly
  - Check the fit with a Gaussian core + DCB (for the tail)
- Working on adding the photon systematics and will update on this soon
- After this, can proceed to the next steps of signal model





# Summary and Next steps

---

- Some issues to be resolved, before moving on to the statistical side of the analysis
  - Fix the generator level weights and produce Data/MC comparison plots with the fix
  - Di-Photon pairing:
    - Check the alternate Gen-Reco pairing
    - Other suggestions are welcome
- Working on these checks
- Next step would be to follow through with the entire flashggFinalFit machinery, and compare where we stand wrt the ATLAS result
  - [ATLAS Paper link](#)