Search for SM Higgs Boson decaying to two photons in Run 2 LHC collisions at CMS

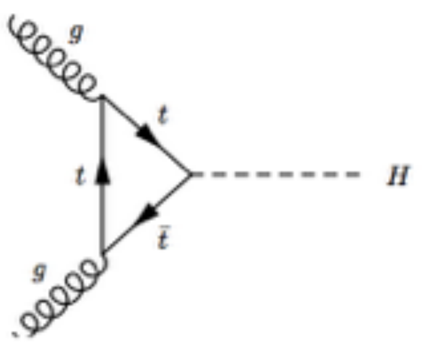
Louie Corpe (Imperial College)
on behalf of the CMS Collaboration.

Joint Annual HEPP and APP Conference 2016
Parallel Session 2A, 22.03.16

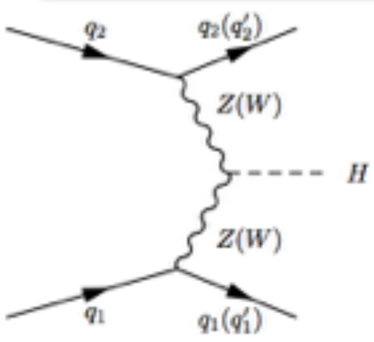
Introduction

- H→γγ played key role in Higgs boson discovery. Low BR (<1%) but clean signature: **sensitive decay channel.**
- **First 13 TeV CMS H→γγ analysis** based on **2015 dataset (2.7 fb^{-1}).** Follows Run 1 (7+8 TeV) analysis structure.
- We consider a signal region of 115 < $m_{\gamma\gamma}$ < 135 GeV and background regions extending to 100 < $m_{\gamma\gamma}$ < 180 GeV.

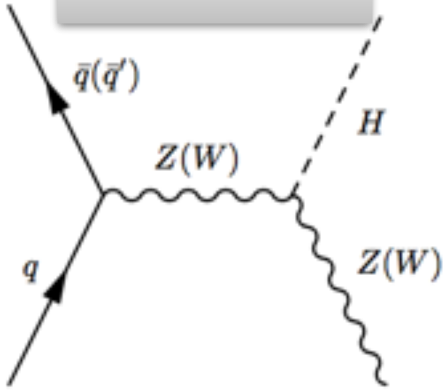
ggH ~87%



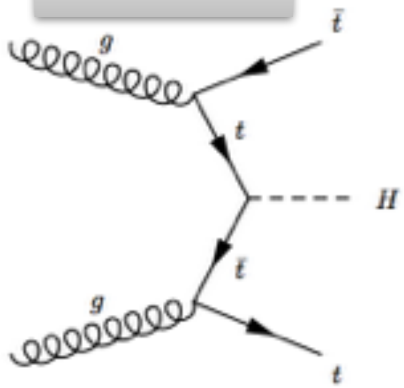
VBF ~ 7.5%



VH ~4.5%



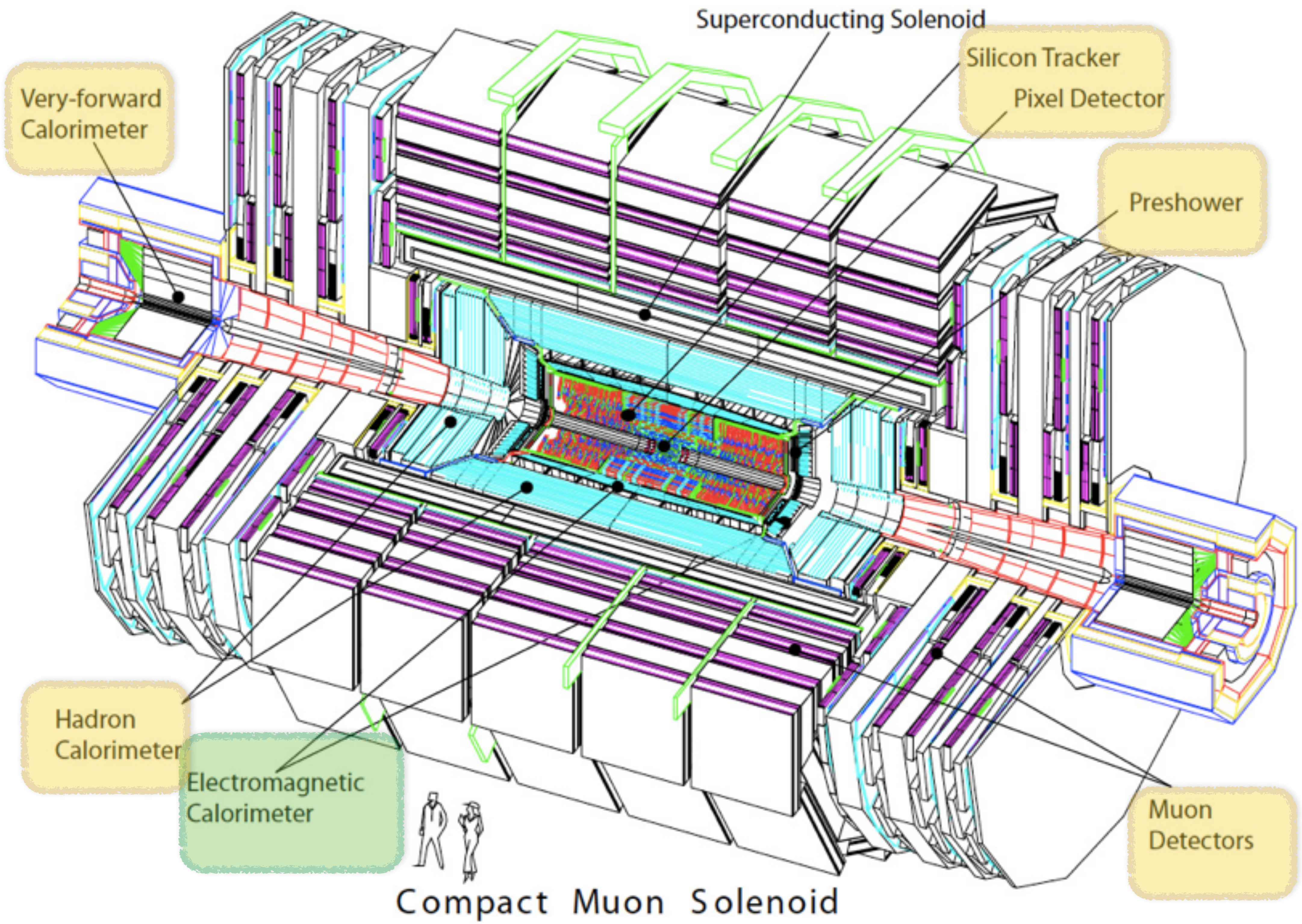
ttH ~1%



- **Diphotons split into 8 categories to maximise sensitivity:**
  - **ttH categories**: Target **leptonic and hadronic** decays separately.
    Particularly statistics-limited, but large increase in cross-section wrt 8TeV expected (factor ~4).
  - **VBF categories**: use MVAs to identify **events with dijets**. Classify by sensitivity into **2 subcategories.**
  - **Untagged categories**: mostly populated by ggH (and VH), bring the largest contribution to our sensitivity.
    Classify by sensitivity into **4 subcategories.**

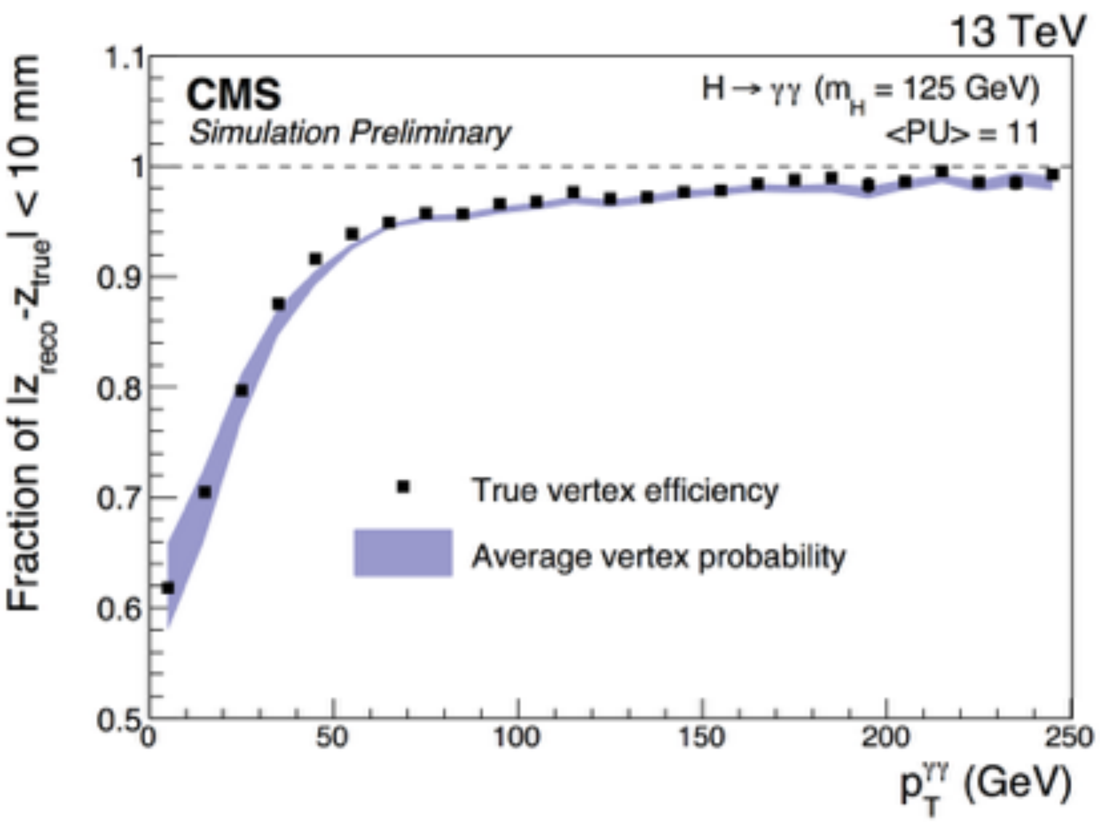
ttHLeptonic → ttHHadronic → VBF0-1 → Untagged0-3

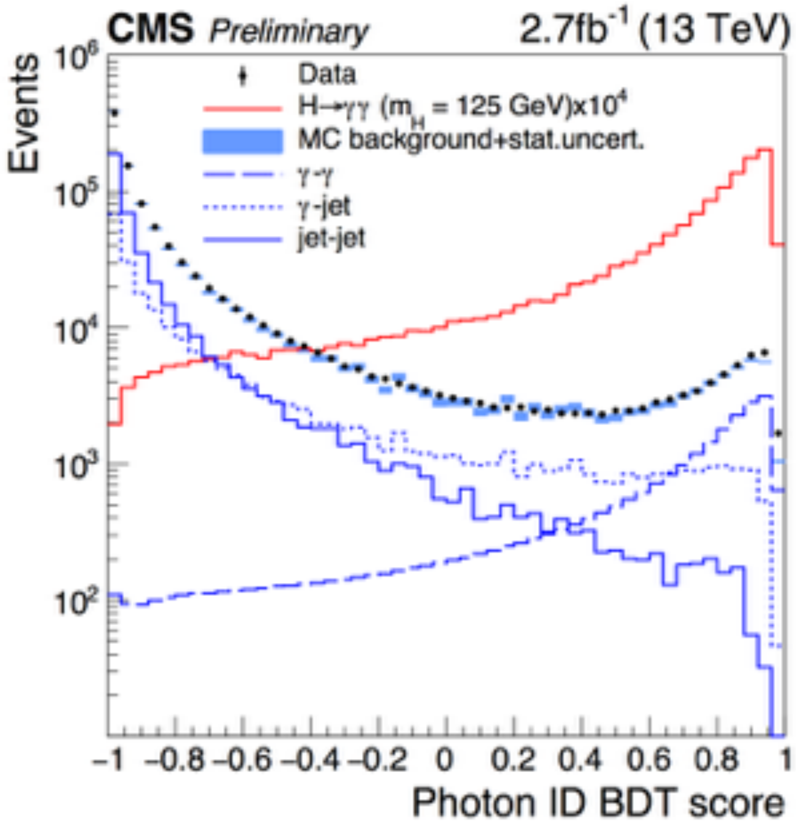
The CMS detector



Imperial College London

Vertex ID and photon/diphoton ID





Vertex ID

$$m_H = m_{\gamma\gamma} = \sqrt{2E_{\gamma 1}E_{\gamma 2}(1-\cos\alpha)}$$

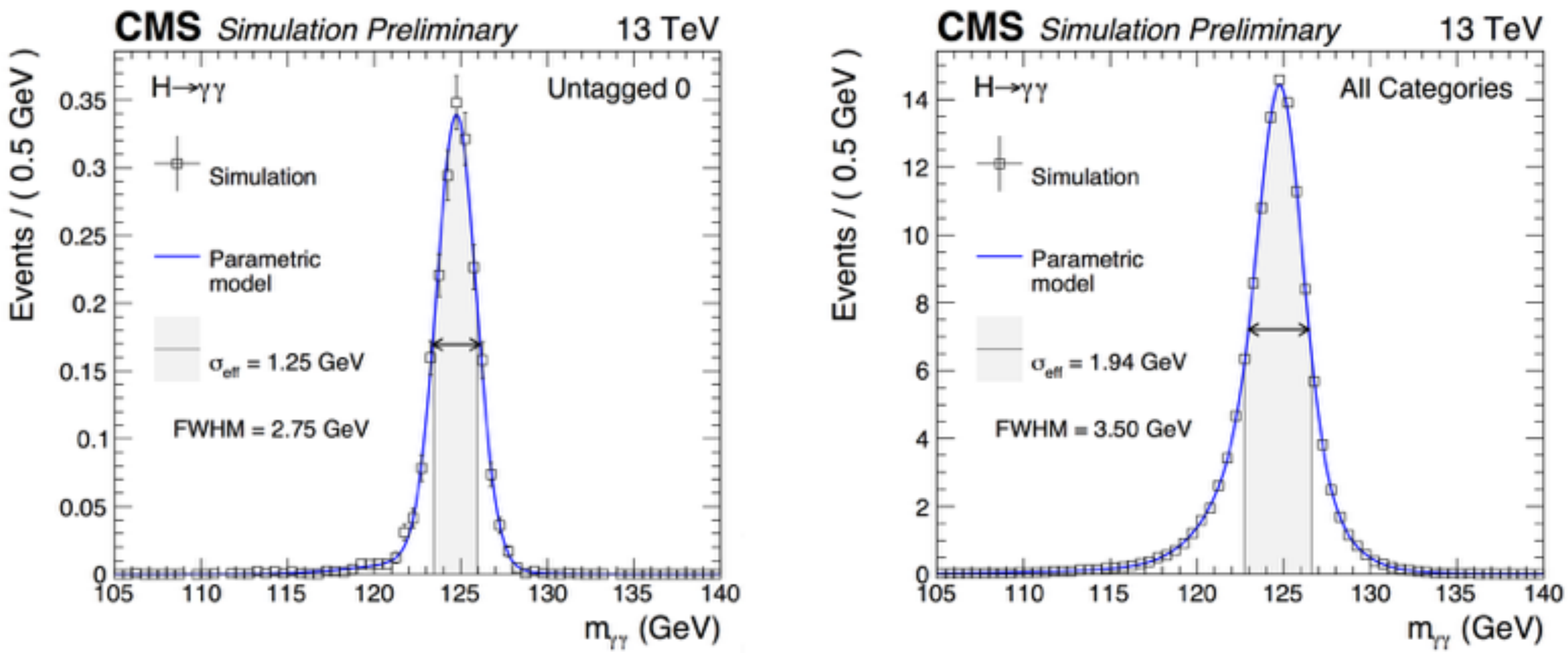
- **Correct vertex assignment is crucial to diphoton mass resolution.**
- $|z_{reco} - z_{true}| < 1$cm is found to keep angular contributions negligible wrt energy resolution
- **Vertex identification MVA** exploits recoiling tracks against diphoton system and conversion tracks (when present).
- Also extract an estimate of the probability to choose the right vertex, which is used as input to the Diphoton BDT later.

Photon ID and diphoton pairs

- Photon identification aims at selecting **prompt photons against $\pi^0/\eta \rightarrow \gamma\gamma$ and electrons**
- **Multivariate approach** with BDT, trained on γ+jet samples, combining shower shape and isolation variables
- A further **multivariate discriminator is used to identify signal-like diphoton pairs:** kinematics, high photon ID scores and good mass resolution from background

Signal and Background Modelling

- **Signal Model:** For each category/process, **fit $m_{\gamma\gamma}$ distribution to a sum of N≤5 Gaussians**, separately for:
  - **Vertex correctly identified** ($|Z_{true} - Z_{chosen}| \leq 1cm$) - mass resolution dominated by energy resolution.
  - **Vertex incorrectly identified** ($|Z_{true} - Z_{chosen}| > 1cm$) - mass resolution dominated by uncert. on vertex position.
- **Each Gaussian parameter is interpolated between mass points at 120, 125 and 130 GeV,** → yields a **smooth parametric model** in each process/category.
- Normalisation obtained as a function of efficiency*acceptance.



- **Background Model:** produced using **data-driven method as in Run 1 analysis**

Systematics

- Key systematics fall into 3 classes of uncertainties:

Affecting shape of signal mass distribution (parametric nuisance built into model)

Photon energy scale and resolution,
Corrections for Material,
Scintillation efficiency,
ECAL Light Yield,
Detector simulation,
Vertex efficiency

Not affecting shape of signal mass distribution (uncorrelated log-normal uncertainty),

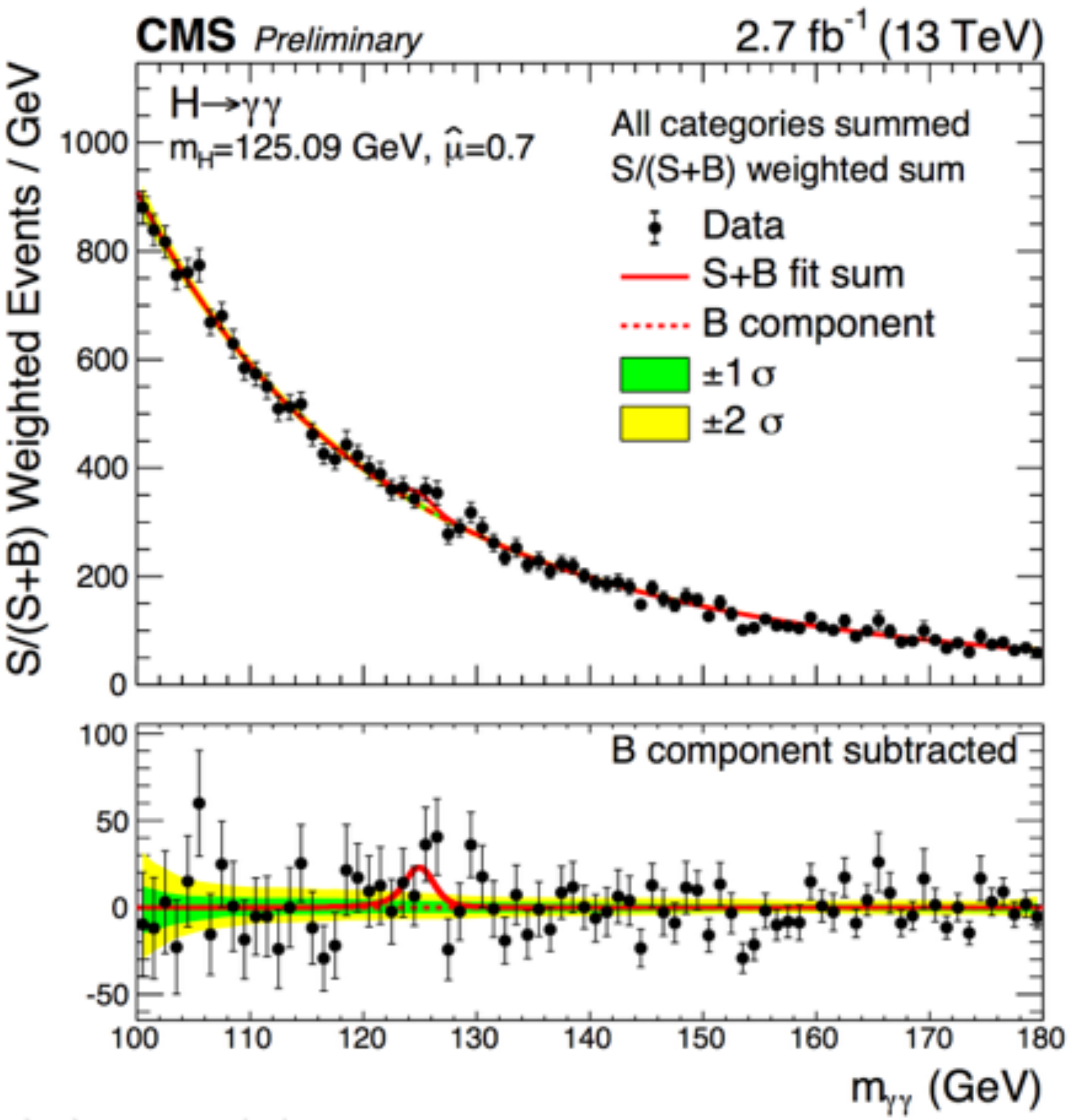
Theory uncertainties (PDFs,α_S, QCD scale,$H\rightarrow\gamma\gamma$ BR)

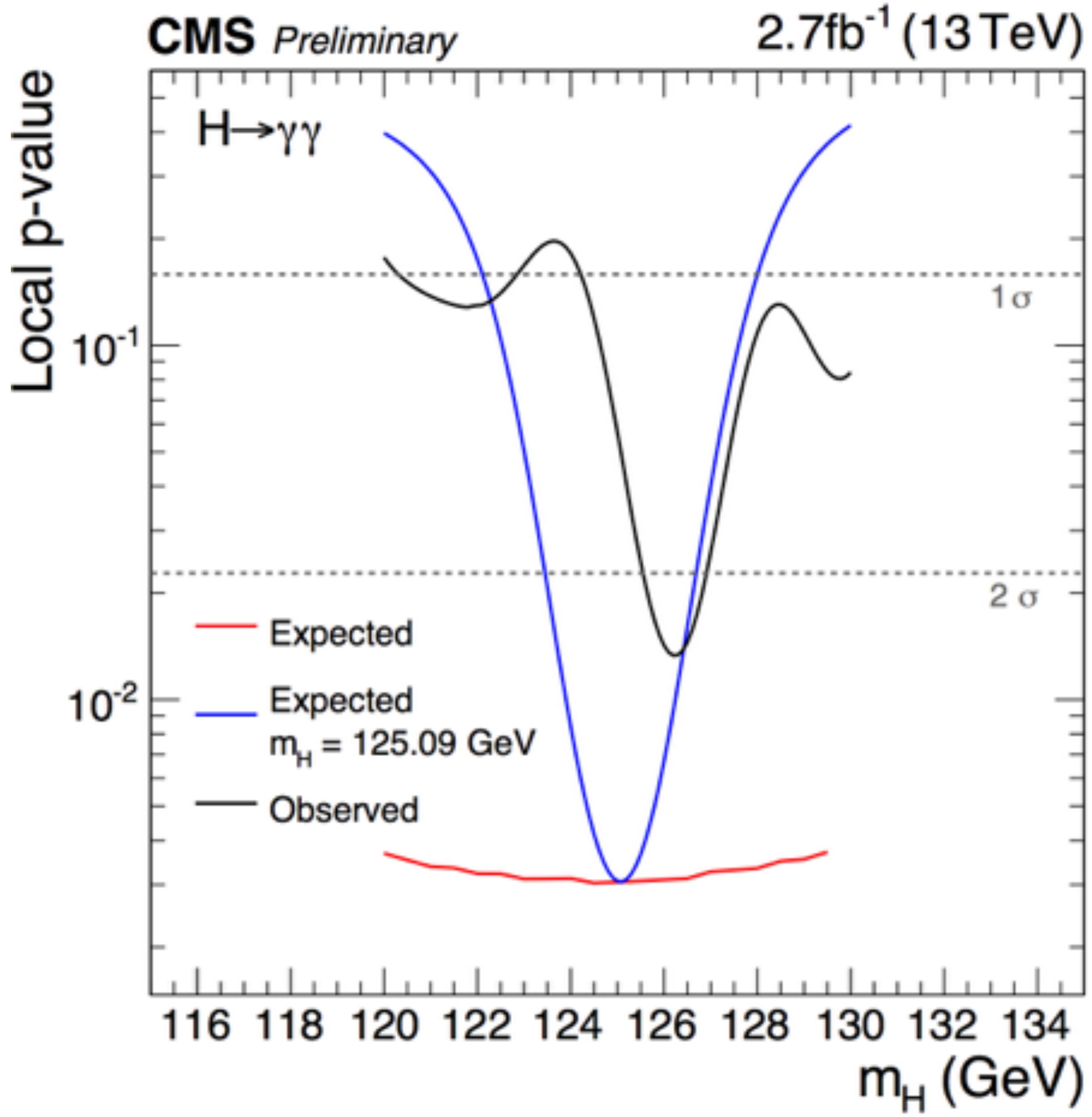
Trigger efficiency,
Integrated luminosity,
Vertex efficiency,
Preselection,
B-tagging efficiency,
Gluon-splitting fraction,
ID efficiency for e and μ

Affecting classification and categorisation (category migration)

Photon ID MVA uncertainty
Jet energy scale,
Jet energy smearings,
Jet Veto,
Underlying event/parton shower

Unblinded $m_{\gamma\gamma}$ and p-value plots



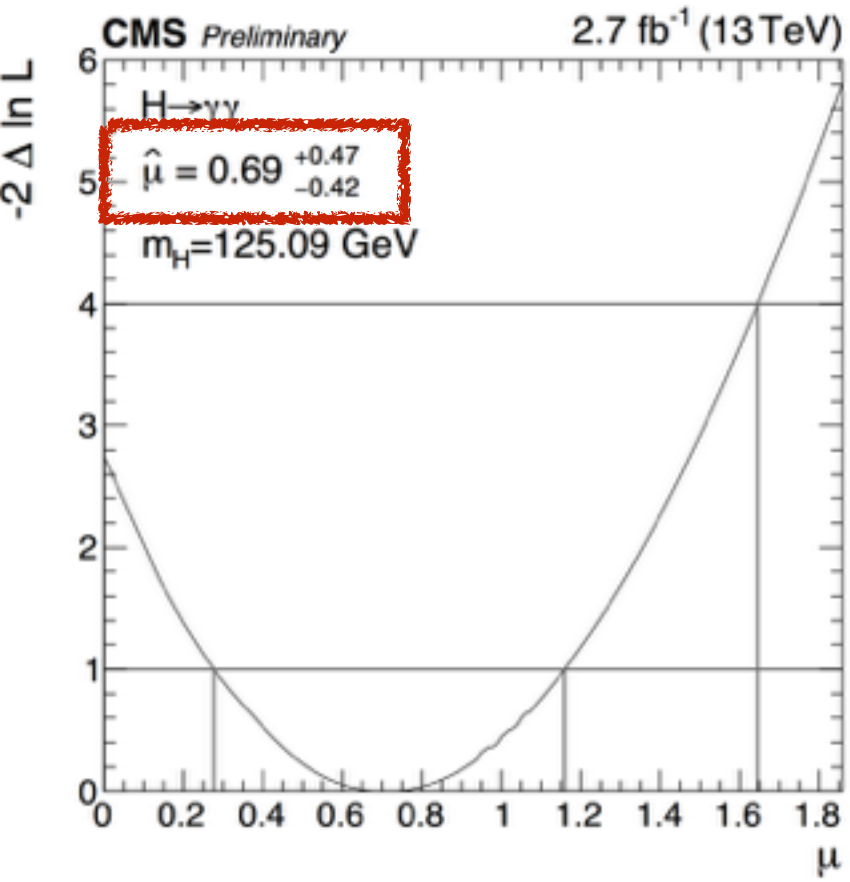


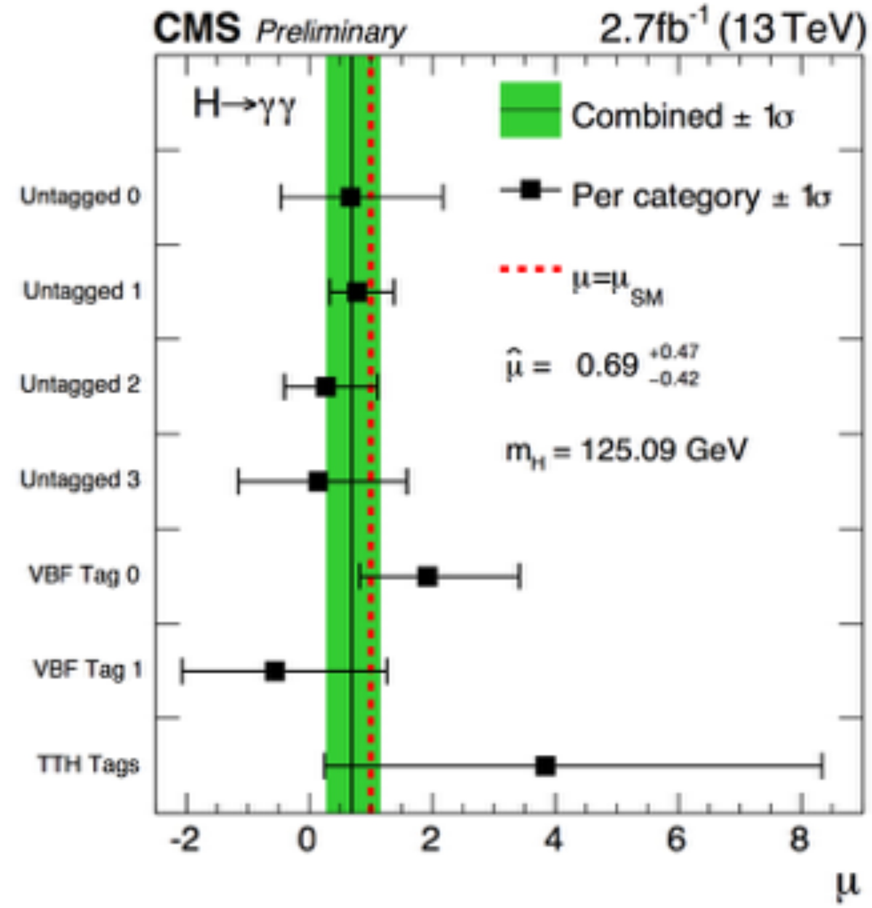
- Data and Signal + Background Fits shown, in the scenario where **m_H is fixed to the Run 1 best fit value of 125.09 GeV**
- **Weighted by S/(S+B)**

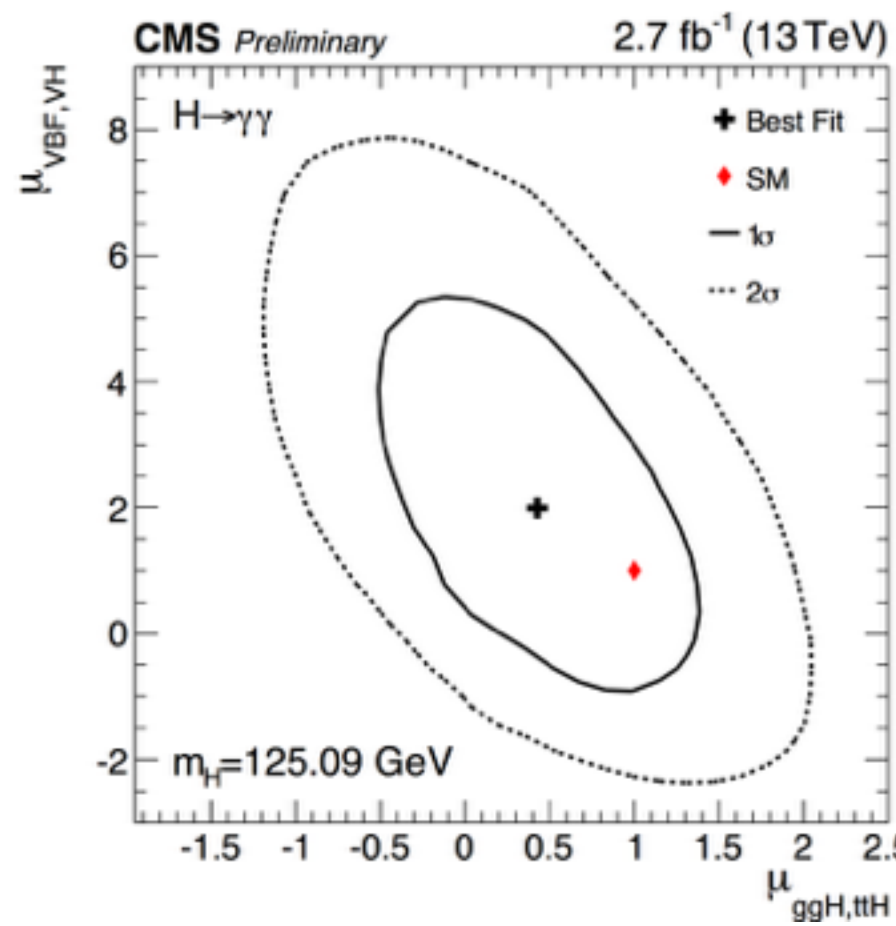
- Expected significance at M_H=125.09 GeV was **2.74 σ**
- Observed significance at M_H=125.09 GeV is **1.67 σ**

Production rate modifier

$\mu := \sigma / \sigma_{SM}$







- Perform a likelihood scan versus μ for all categories together to determine the best fit assuming a fixed m_H=125.09 GeV.
- Also perform similar scan for each category individually.
- We also allow the μ for fermionic and bosonic Higgs production processes to float separately in the 2D scan.

Best fits

$\mu_{ggh,tth}$: $0.43^{+0.59}_{-0.63}$

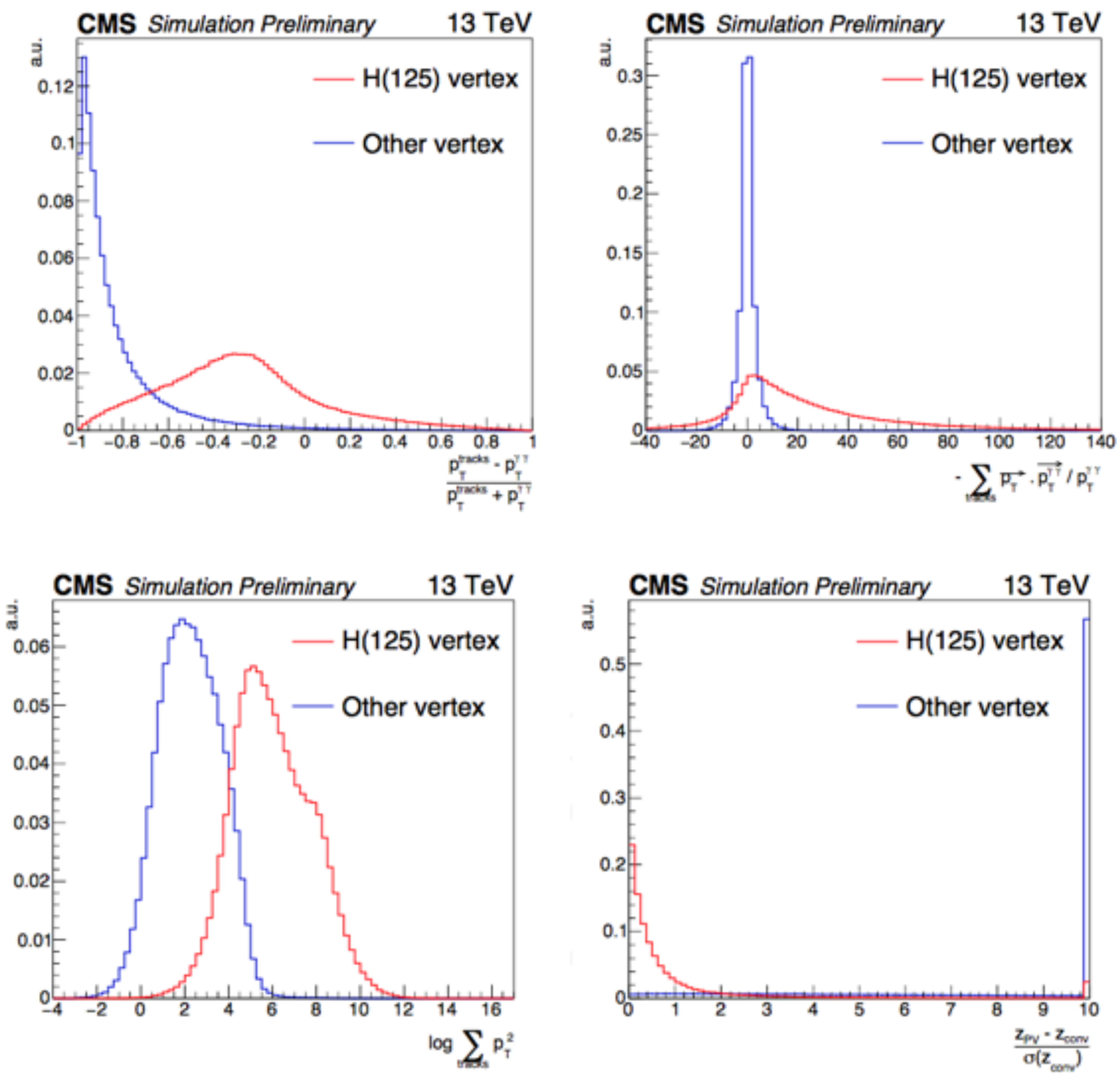
$\mu_{VBFT,TH}$: $1.99^{+2.14}_{-1.98}$

Summary and Conclusions

- The CMS $H\rightarrow\gamma\gamma$ working group has prepared its **first results using the 2015 13 TeV** dataset with **2.7 fb^{-1}.**
- Demonstrated **first steps towards rediscovery of Higgs boson** in diphoton decay channel.
- Significance : **Expected 2.74σ and observed 1.64σ** at Run 1 best fit value of m_H =125.09 GeV
- Fixing m_H =125.09 GeV, best fit **μ=0.69, but with large uncertainties of +0.47 - 0.42**
- Results are **compatible with the SM** within 1σ.
- Ready for more data as Run 2 continues in 2016!

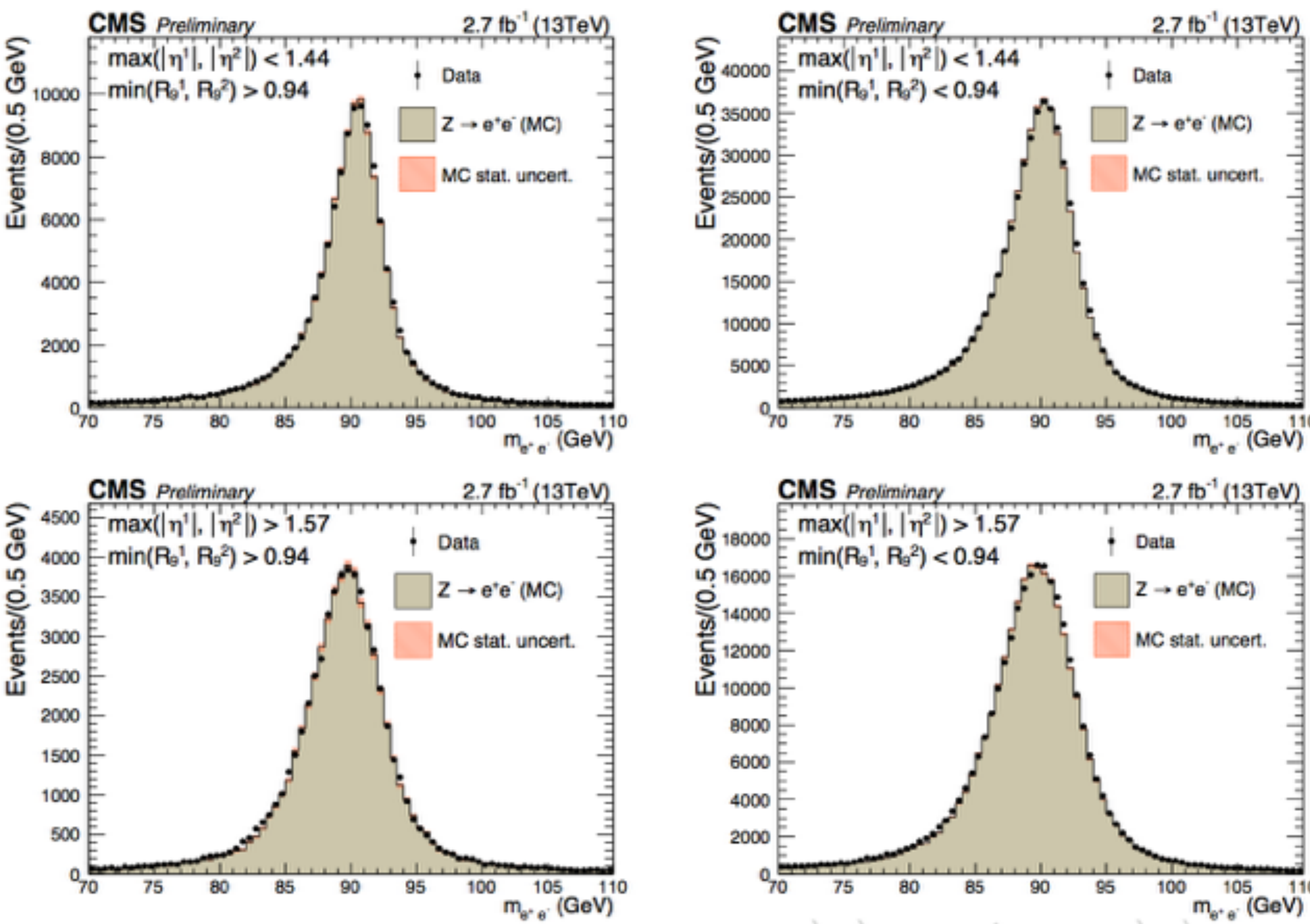
BACKUP

Vertex ID



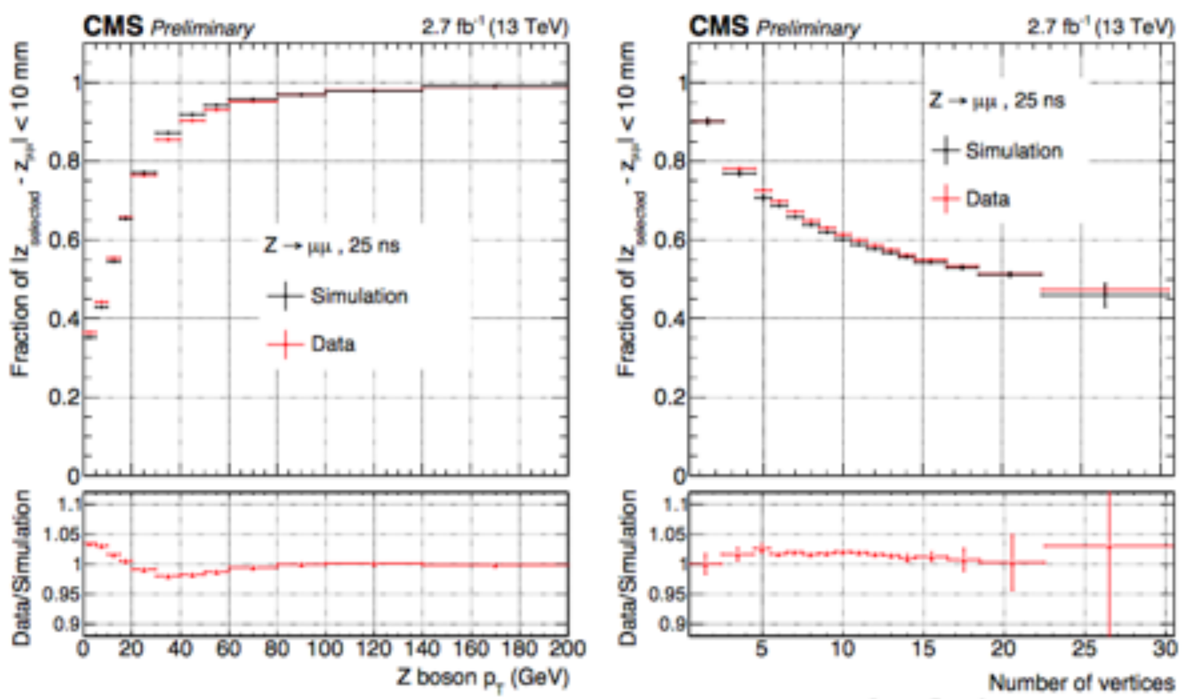
- Inputs to the vertex ID BDT.
  - ptasym
  - ptbal
  - sumpt2
  - pull
- correct vertex (dZ<1cm)
- incorrect vertex (dZ >1cm)

Energy Reconstruction

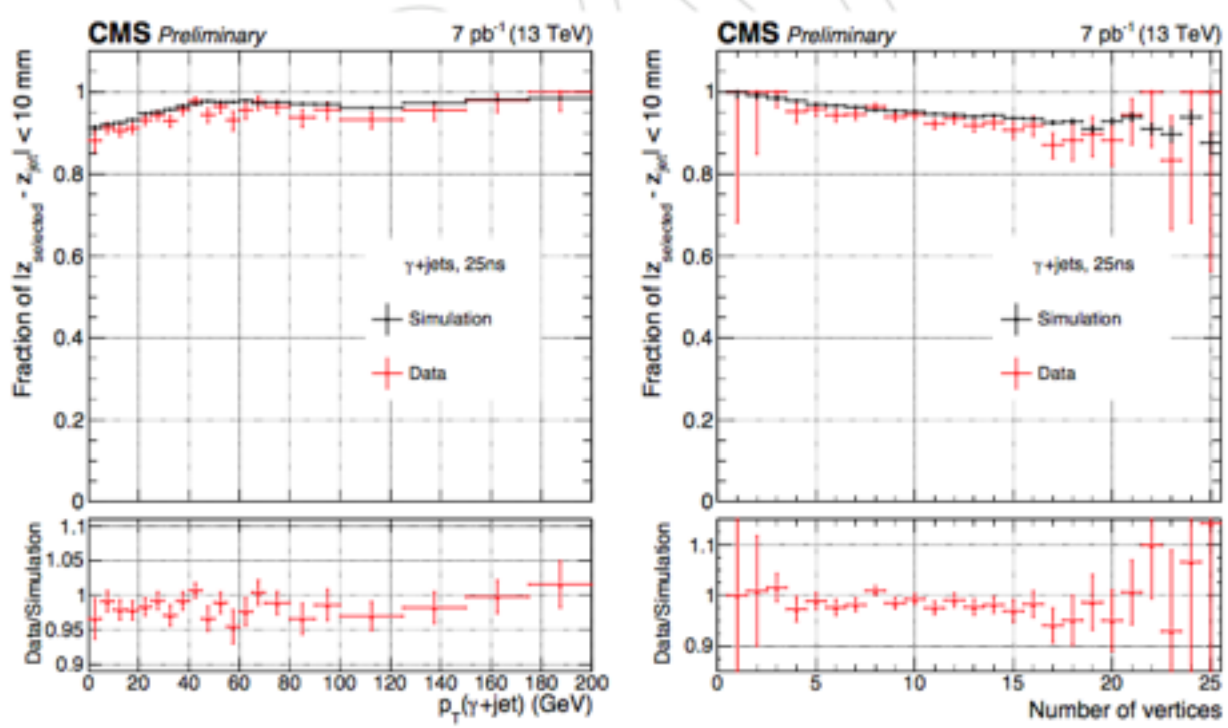


- data vs MC comparison for invariant mass of Z→ee electrons reconstructed as photons.
- Events split into |η| and R_9 categories. (R9 is the ratio of the amount of energy in a 3x3 array of ECAL crystals around the seed divided by the total energy in the supercluster. High R_9 (>0.94) photons are likely to be unconverted, while Low R_9 photons are likely to have undergone $\gamma \rightarrow e^+e^-$).

Vertex Efficiency Checks

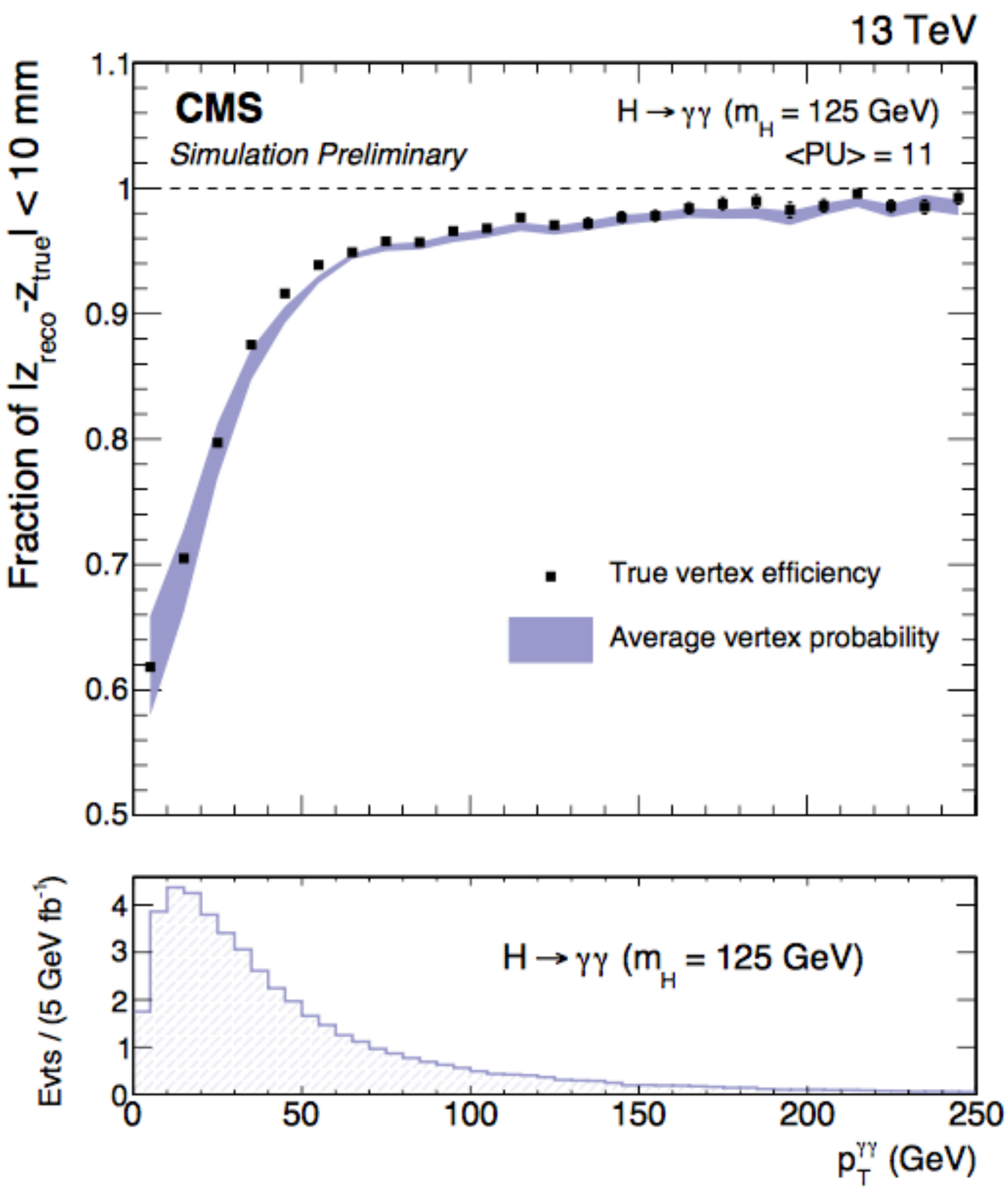


- Vertex finding efficiency as a function of pT and number of vertices in Z→μμ events where the muon tracks have been removed.



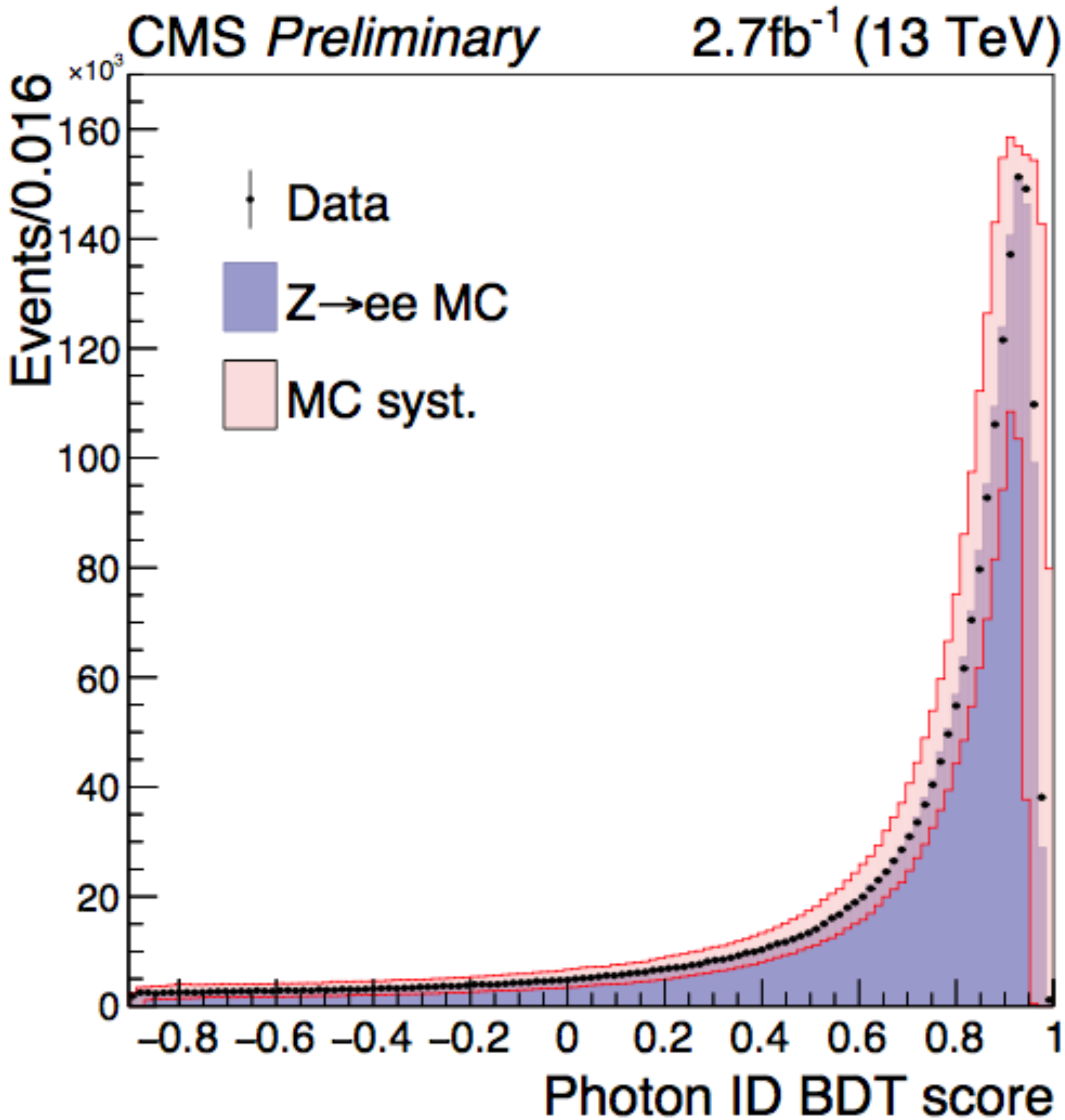
- Vertex finding efficiency as a function of pT and number of vertices in ɣ+Jets events, with converted photons.

Vertex ID extra material



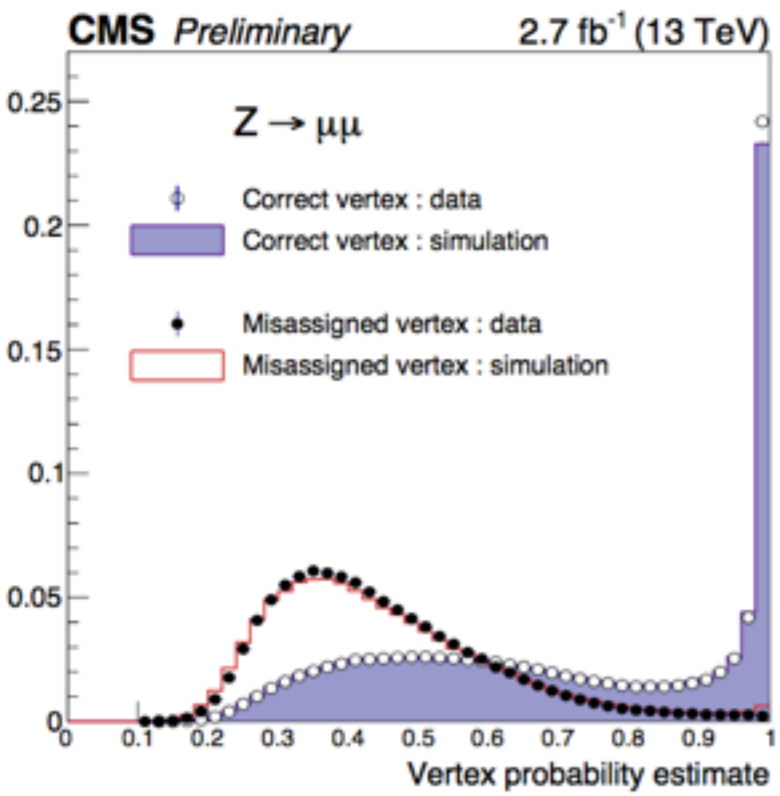
- true vertex efficiency vs average estimated vertex probability as a function of the pT of the diphoton system.
- Reweighting done according to production modes and to average PU.
- diphoton system PT distrubution is also shown.

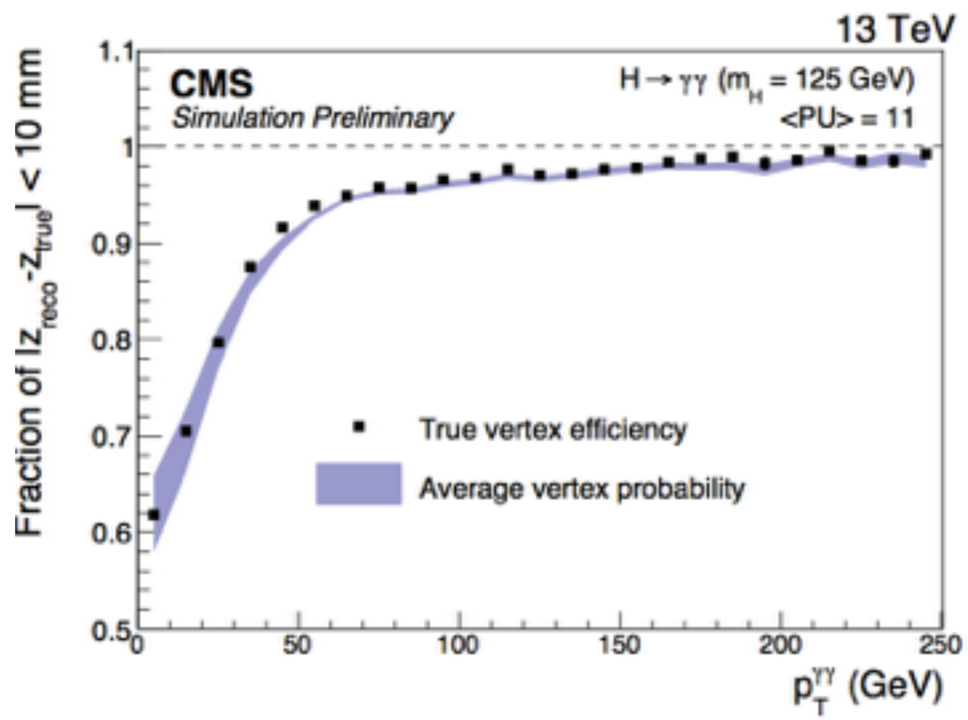
Photon ID extra material

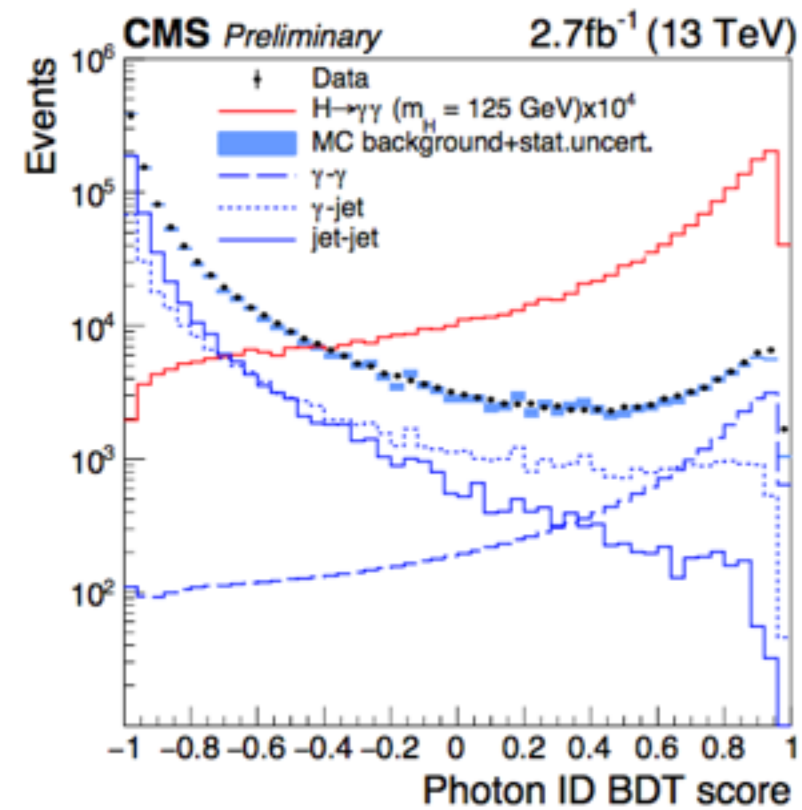


- Check of photon ID BDT on Z→ee events where the electrons are reconstructed as photons.

Vertex ID and photon ID







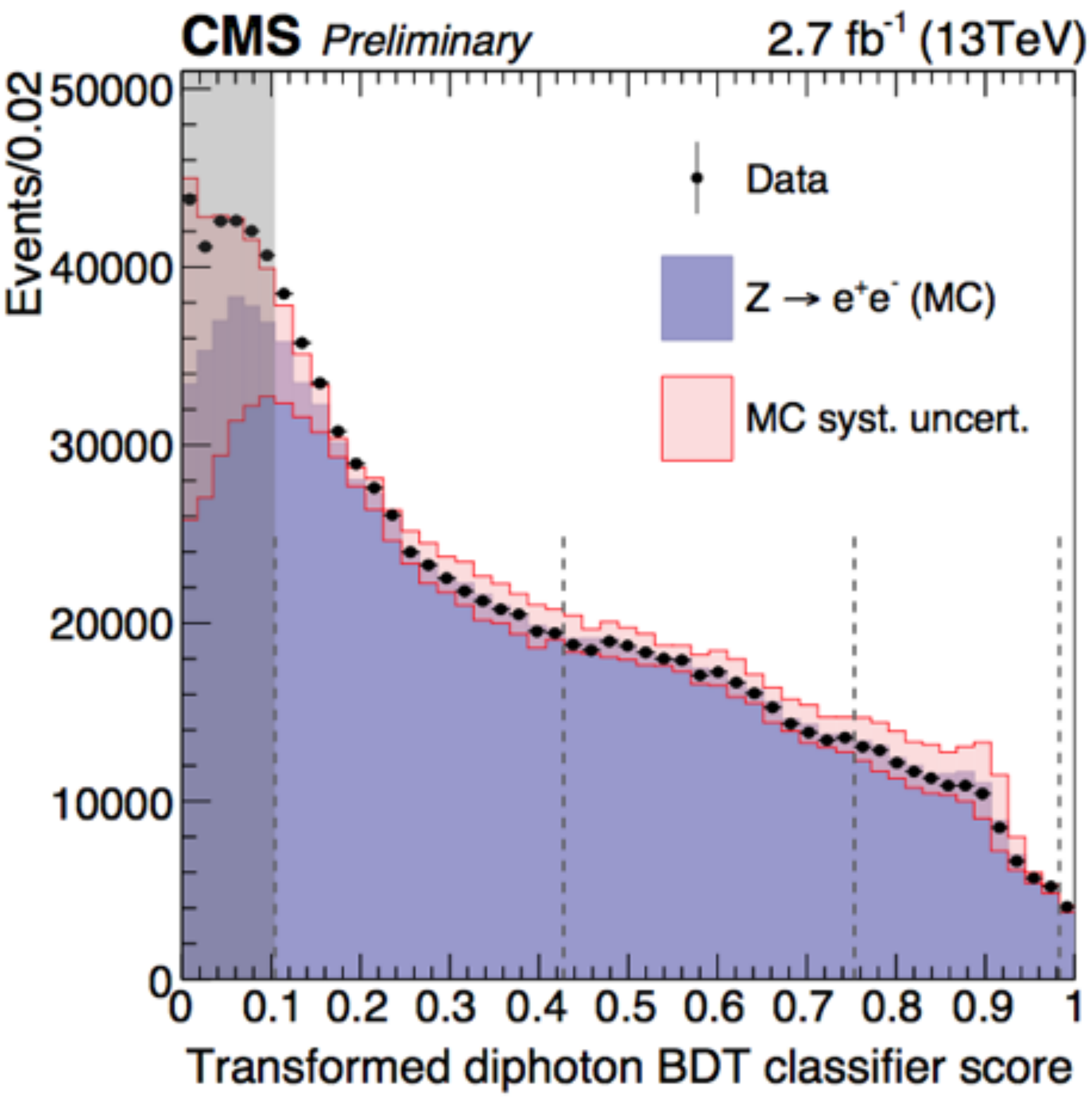
Vertex ID

- **The correct vertex assignment is crucial to diphoton mass resolution.**
- $|z_{reco} - z_{true}| < 1$cm is found to keep angular contributions negligible wrt energy resolution
- **Vertex identification MVA** exploits recoiling tracks against diphoton system and conversion tracks (when present), trained on Hgg samples for m_H=125GeV

Photon ID

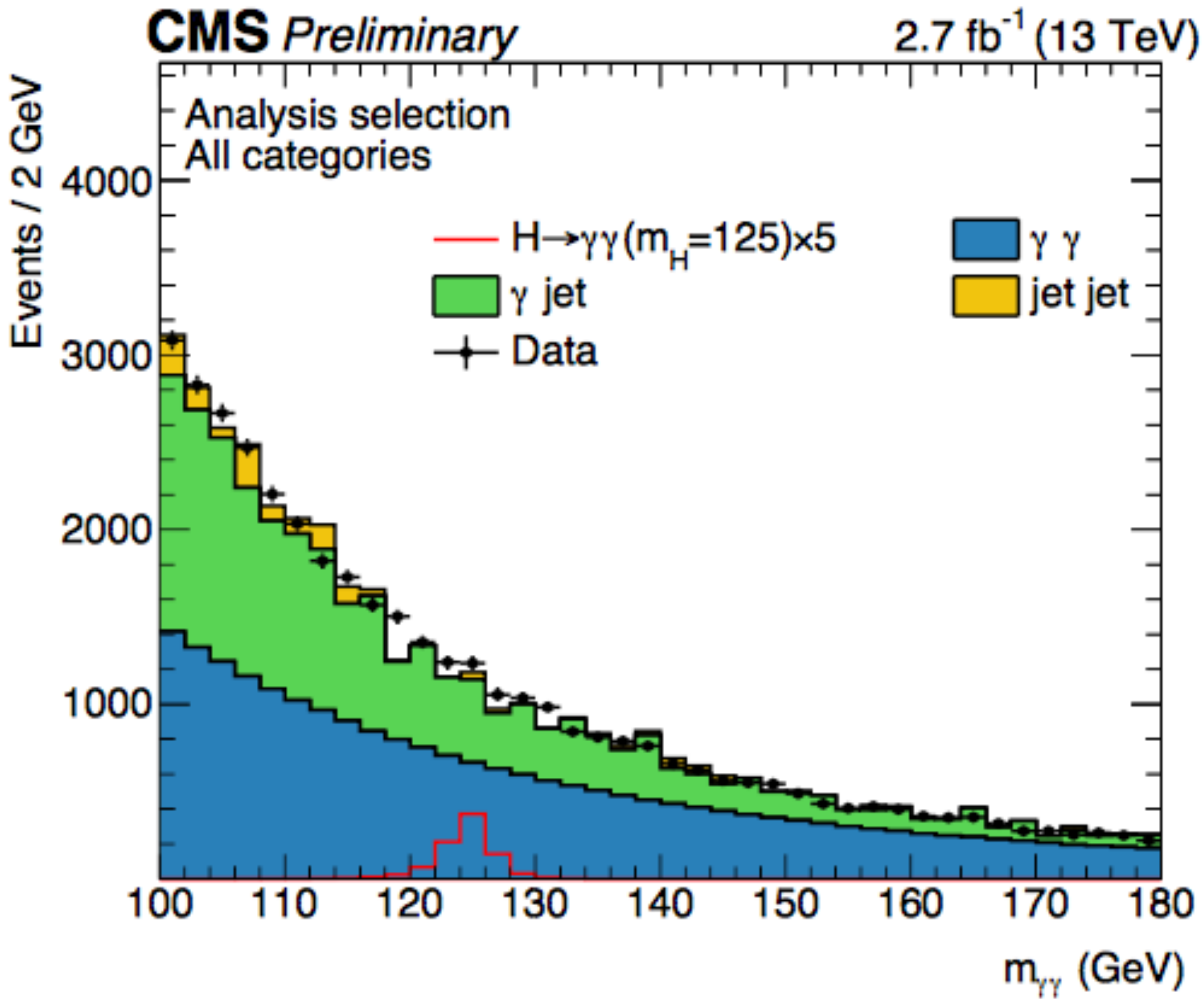
- Photon identification aims at selecting **prompt photons against $\pi^0/\eta \rightarrow \gamma\gamma$ and electrons**
- **multivariate approach** with BDT, trained on γ+jet samples, combining shower shape and isolation variables

Diphoton MVA check



- Check of diphoton BDT on Z→ee events where the electrons are reconstructed as photons.

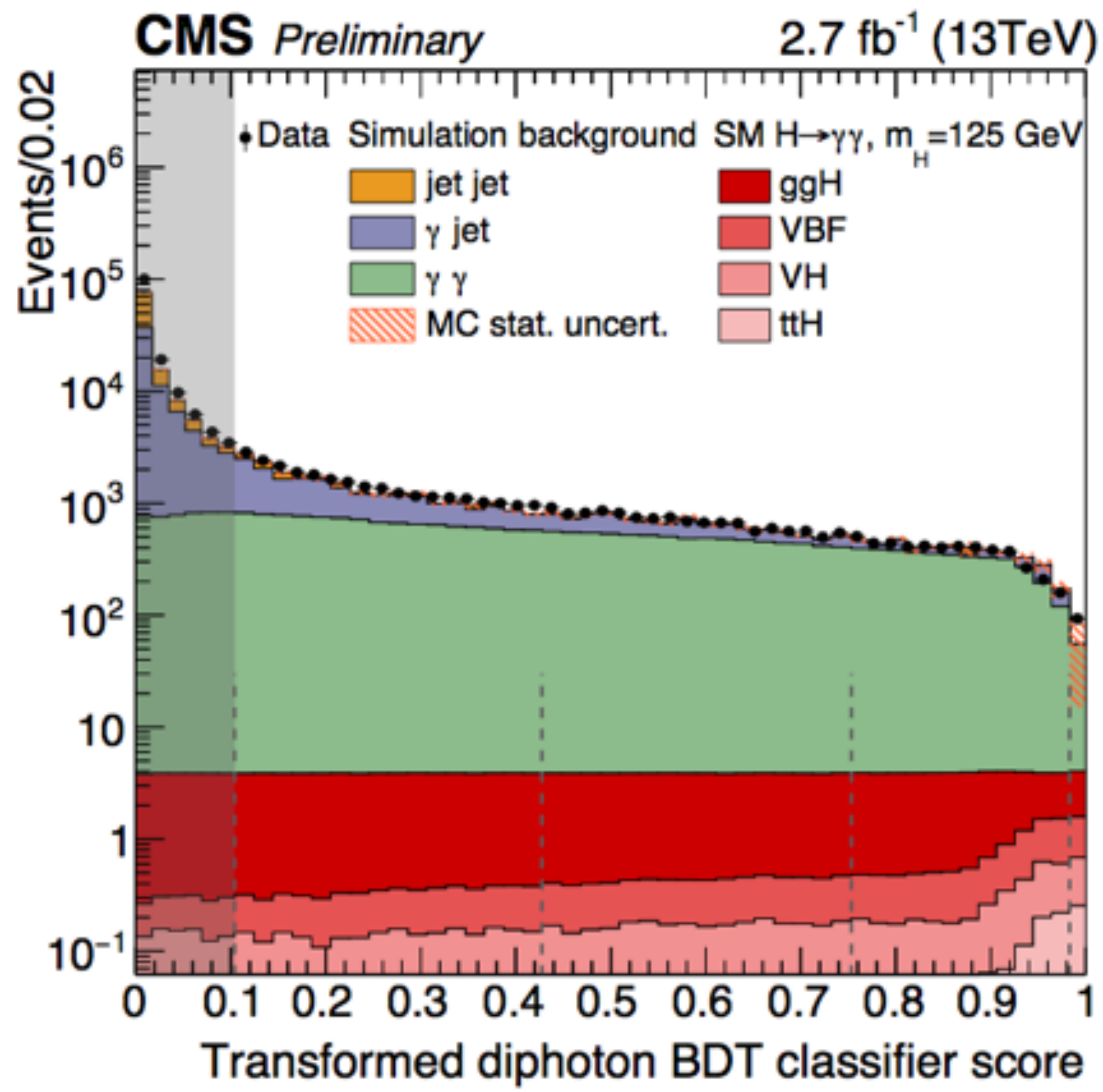
Diphoton invariant mass distribution



- Illustrative diphoton invariant mass distribution for data, and simulated signal and background.

Diphoton Selection and Event Classification

- **MVA discriminator is used to identify signal-like diphoton pairs:** kinematics, high photon ID scores and good mass resolution from background
- Diphotons split into categories, exploiting different S/B ratios and mass resolution => maximum sensitivity:
  - TTH categories: make cuts on photon quality and requirements on bTags, Jets and absence/presence of leptons .
  - VBF categories: use an MVA to identify VBF-like events with dijets. A further MVA using the diphoton and dijet MVAs as inputs is used to classify VBF events by sensitivity into VBFTags 0 and 1.
  - Untagged categories: mostly populated by ggH, bring the largest contribution to analysis' sensitivity. Category boundaries defined by Diphoton MVA.
- Event tagging sequence is defined as follows:

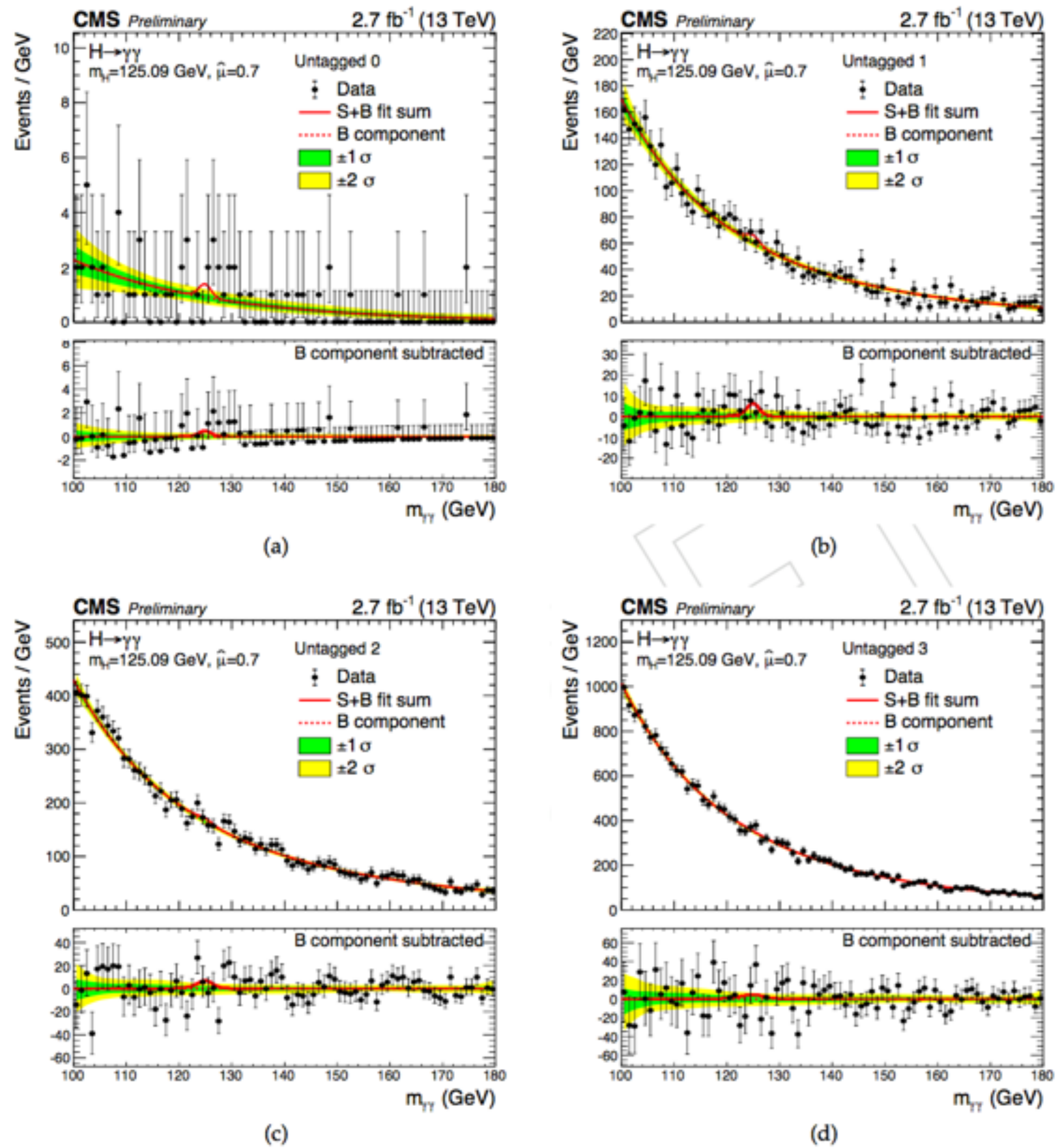


ttHLeptonic → **ttHHadronic** → **VBF0-1** → **Untagged0-3**

Background Model Plots Unblinded

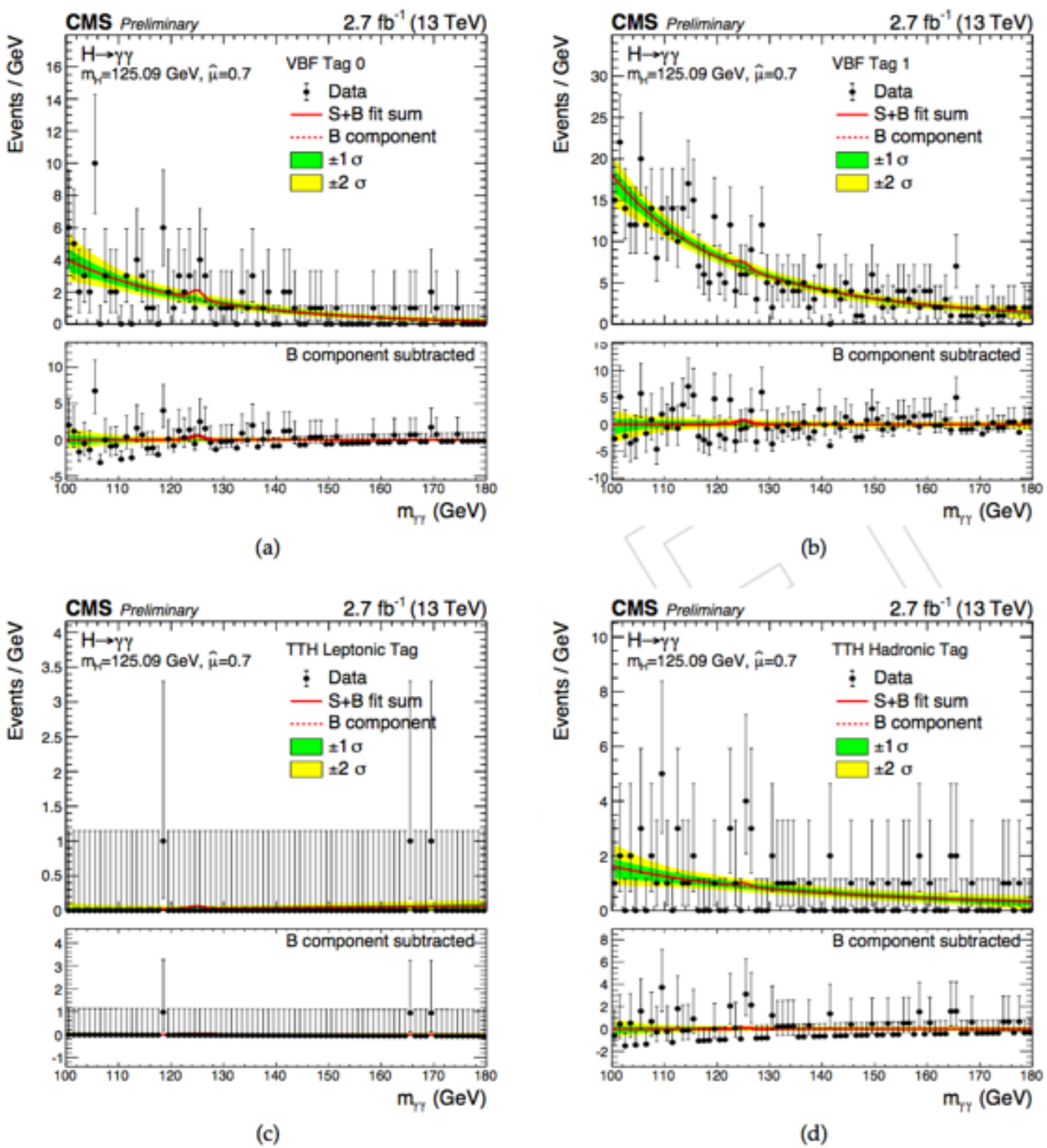
Imperial College London

- Data and Signal + Background Fits shown, in the scenario where **m_H is fixed to the Run 1 best fit value of 125.09 GeV**
- Figures also show the Background-subtracted distributions.
- Uncertainty bands achieved by throwing toys from the post-Fit distributions and finding locations of relevant quantiles.



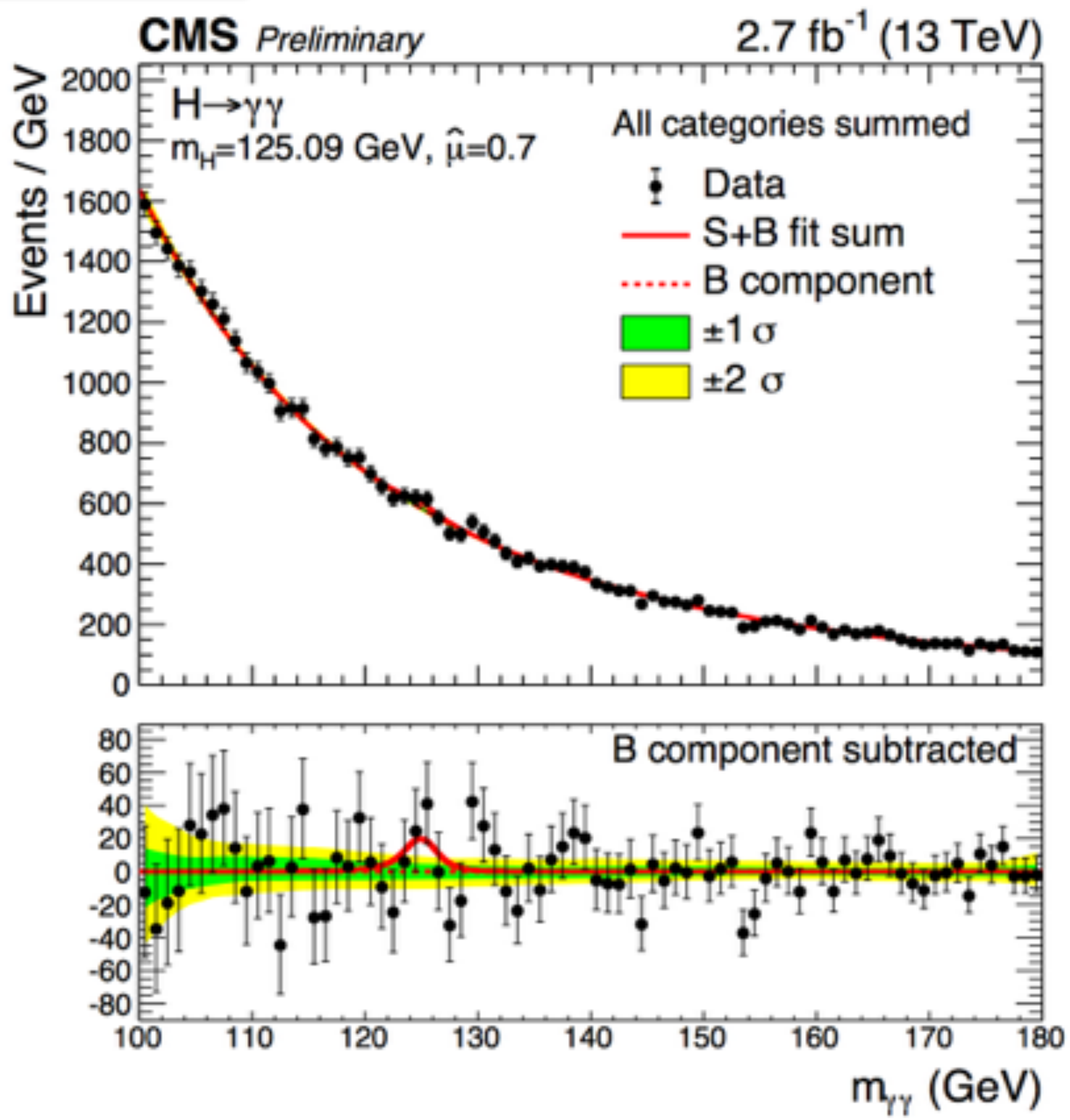
Background Model Plots Unblinded

- Data and Signal + Background Fits shown, in the scenario where **m_H is fixed to the Run 1 best fit value of 125.09 GeV**
- Figures also show the Background-subtracted distributions.
- Uncertainty bands achieved by throwing toys from the post-Fit distributions and finding locations of relevant quantiles.

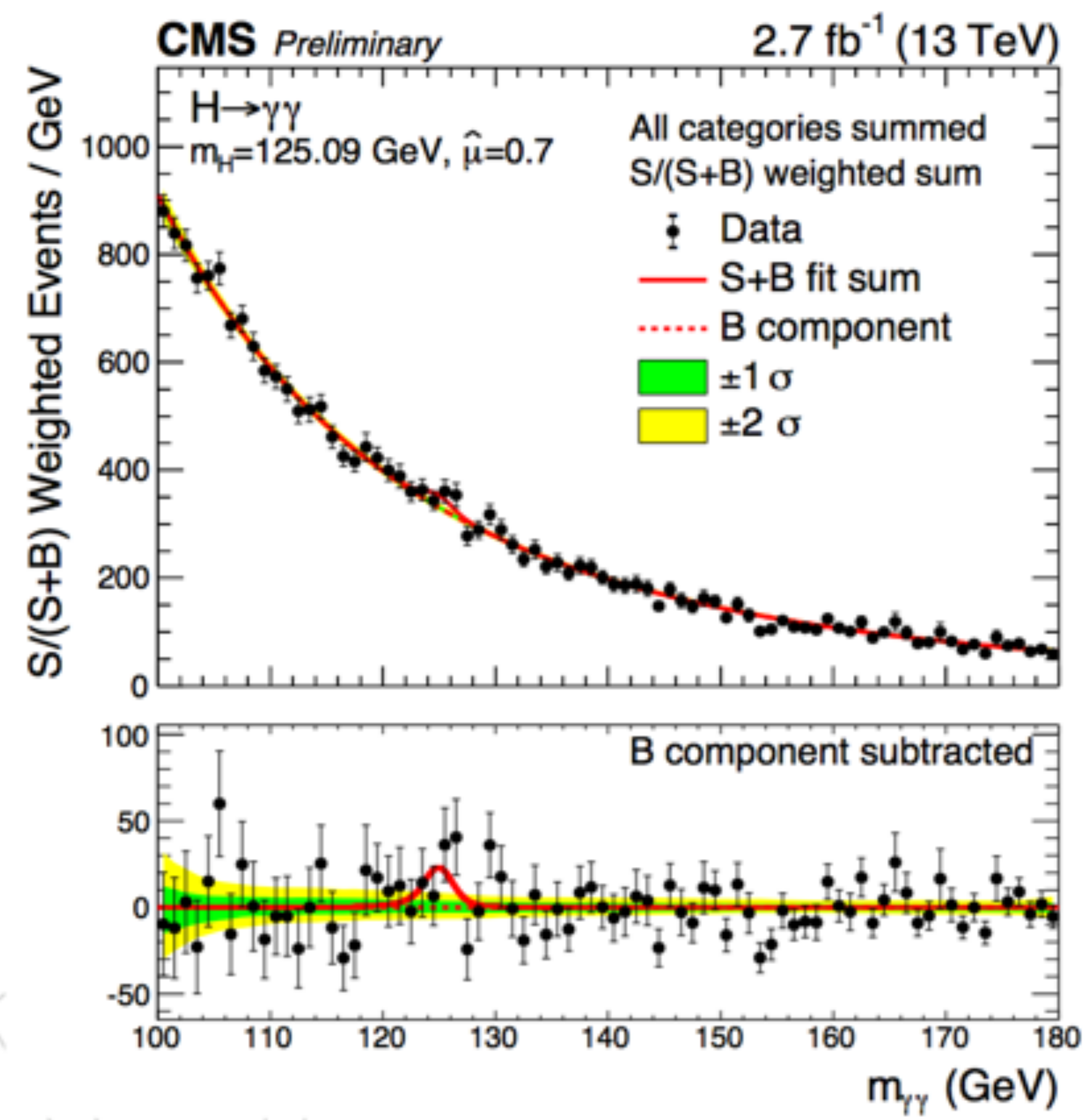


Weighted S/S+B and P-Value

Unweighted

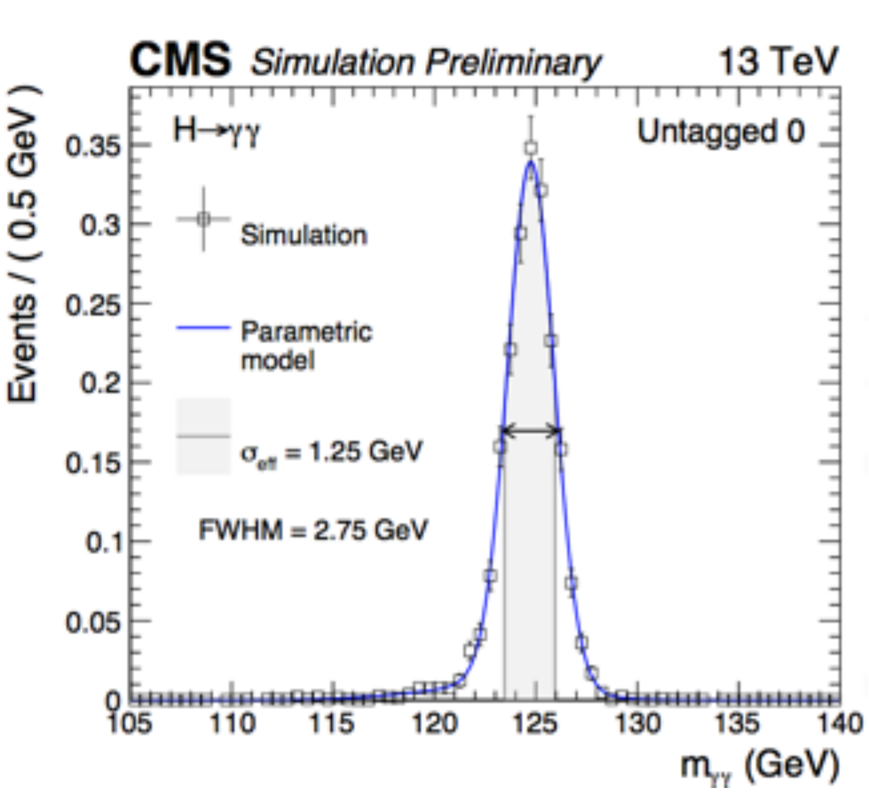


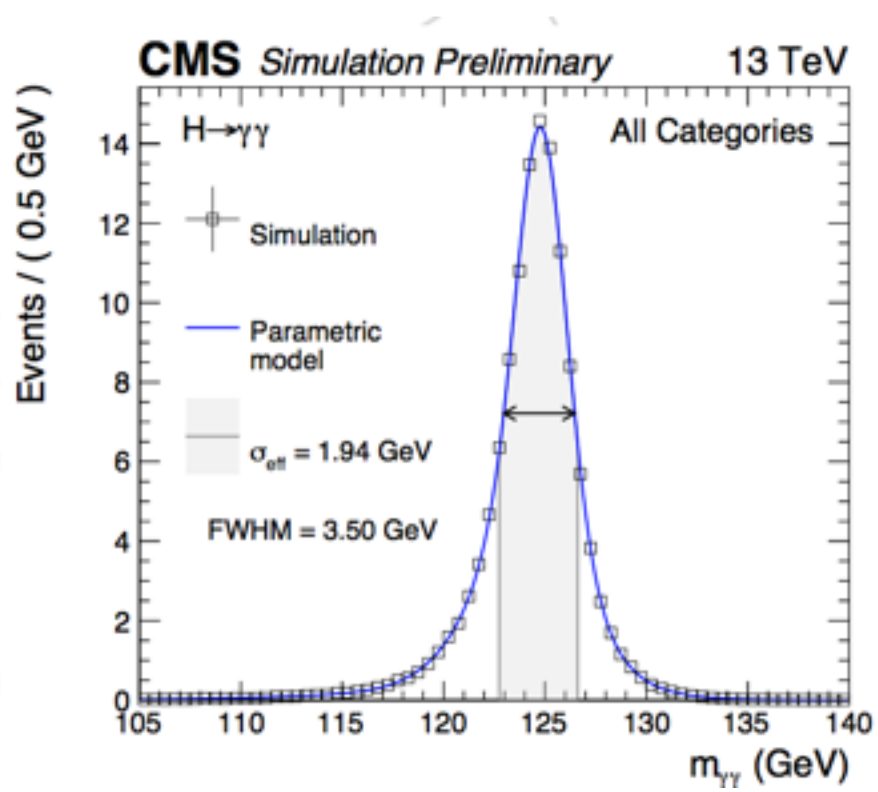
Weighted

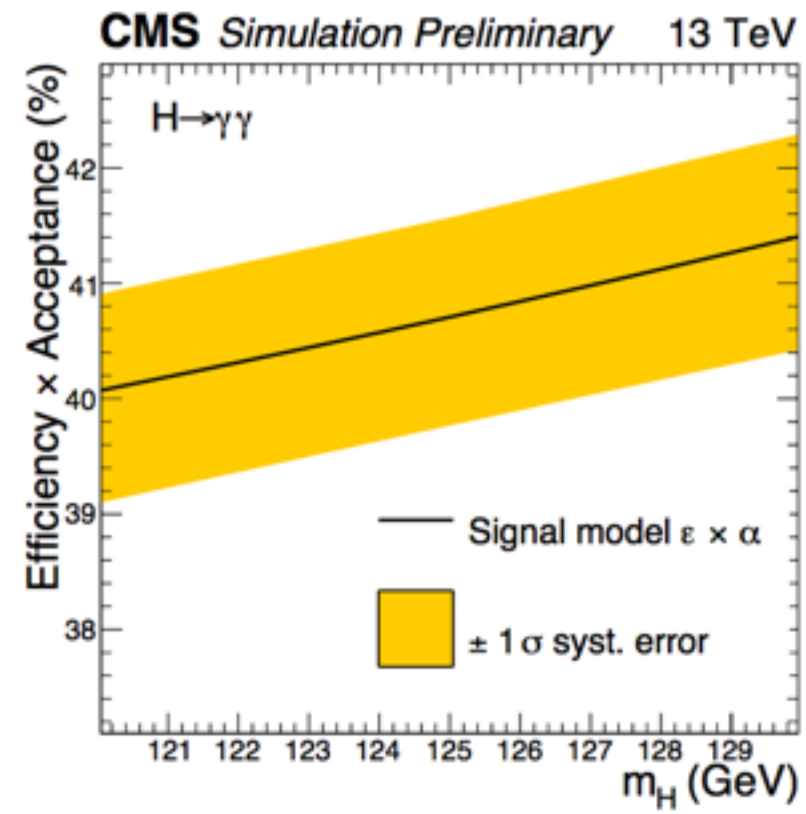


- Shown here is the total invariant mass distribution where we reweight using the factor S/(S+B) in a 1GeV window around the point.

Signal Model

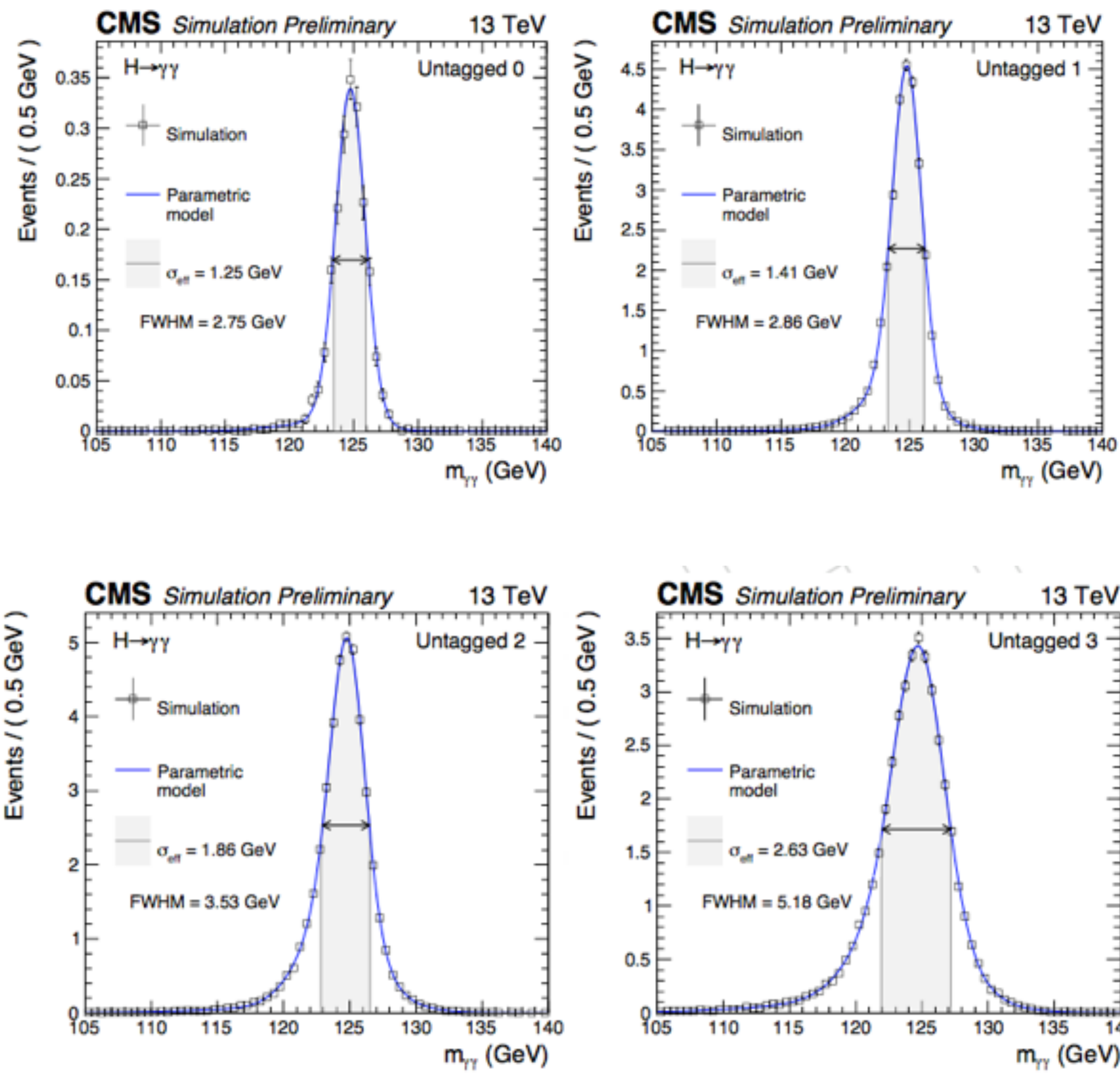




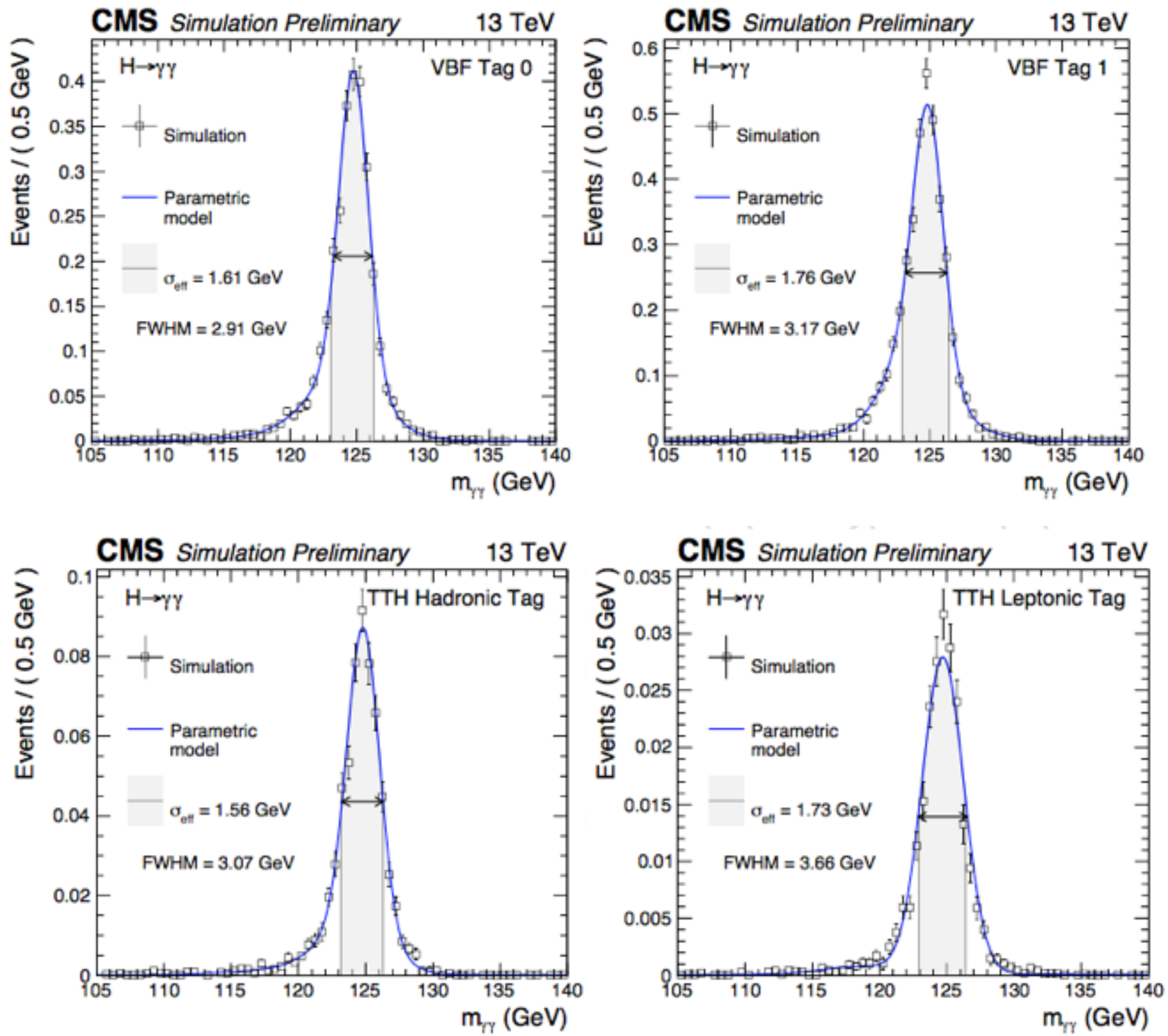


Event Categories	SM 125 GeV Higgs boson expected signal yield								Bkg (GeV^{-1})
	Total	ggH	VBF	WH	ZH	$t\bar{t}H$	σ_{eff} (GeV)	σ_{HM} (GeV)	
Untagged 0	2.08	**76.19** %	10.06 %	7.45 %	3.98 %	2.32 %	1.25	1.17	0.93
Untagged 1	30.44	**86.24** %	7.13 %	3.73 %	2.12 %	0.79 %	1.41	1.22	61.19
Untagged 2	43.36	**91.16** %	4.80 %	2.39 %	1.29 %	0.36 %	1.86	1.50	165.52
Untagged 3	42.18	**92.18** %	4.21 %	2.05 %	1.16 %	0.40 %	2.63	2.20	350.94
VBF Tag 0	3.00	35.28 %	**63.48** %	0.68 %	0.19 %	0.36 %	1.61	1.24	1.57
VBF Tag 1	4.08	53.14 %	**43.62** %	1.69 %	0.85 %	0.69 %	1.77	1.35	6.85
TTH Hadronic Tag	0.64	8.76 %	0.41 %	1.66 %	2.10 %	**87.06** %	1.56	1.31	0.90
TTH Leptonic Tag	0.23	0.14 %	0.09 %	2.91 %	1.31 %	**95.55** %	1.73	1.56	0.03
Total	126.00	86.92 %	7.87 %	2.62 %	1.45 %	1.14 %	1.94	1.49	587.92

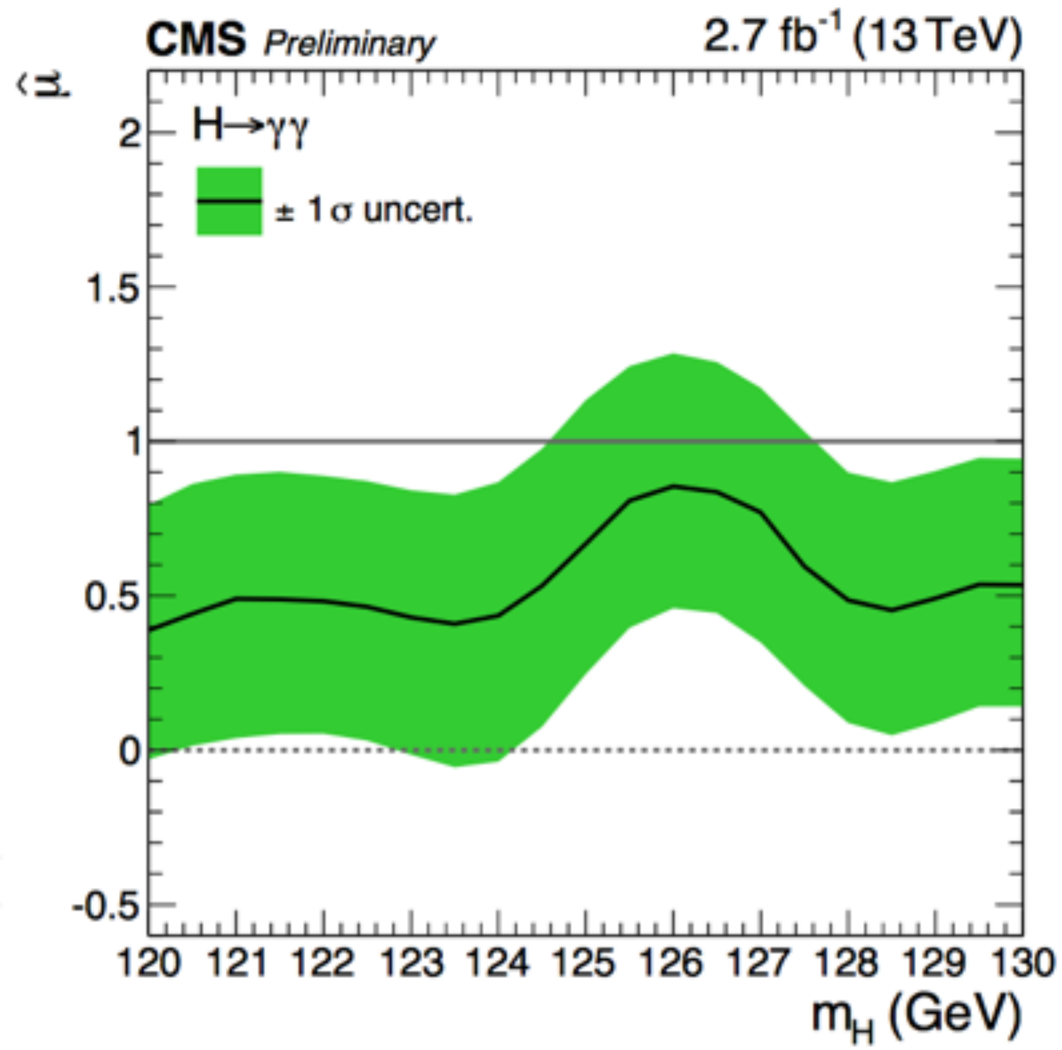
Signal Models (Untagged Categories)



Signal Models (VBF/TTH Categories)



production rates as a function of imposed m_H



- Shows the best fit μ as a function of the imposed m_H condition .