中国物理学会高能物理分会第十二届全国会员代表大会暨学术年会, 广州

希格斯玻色子稀有衰变 H→Zγ 的实验研究

张铭滔

指导教师：冒亚军、周辰

北京大学

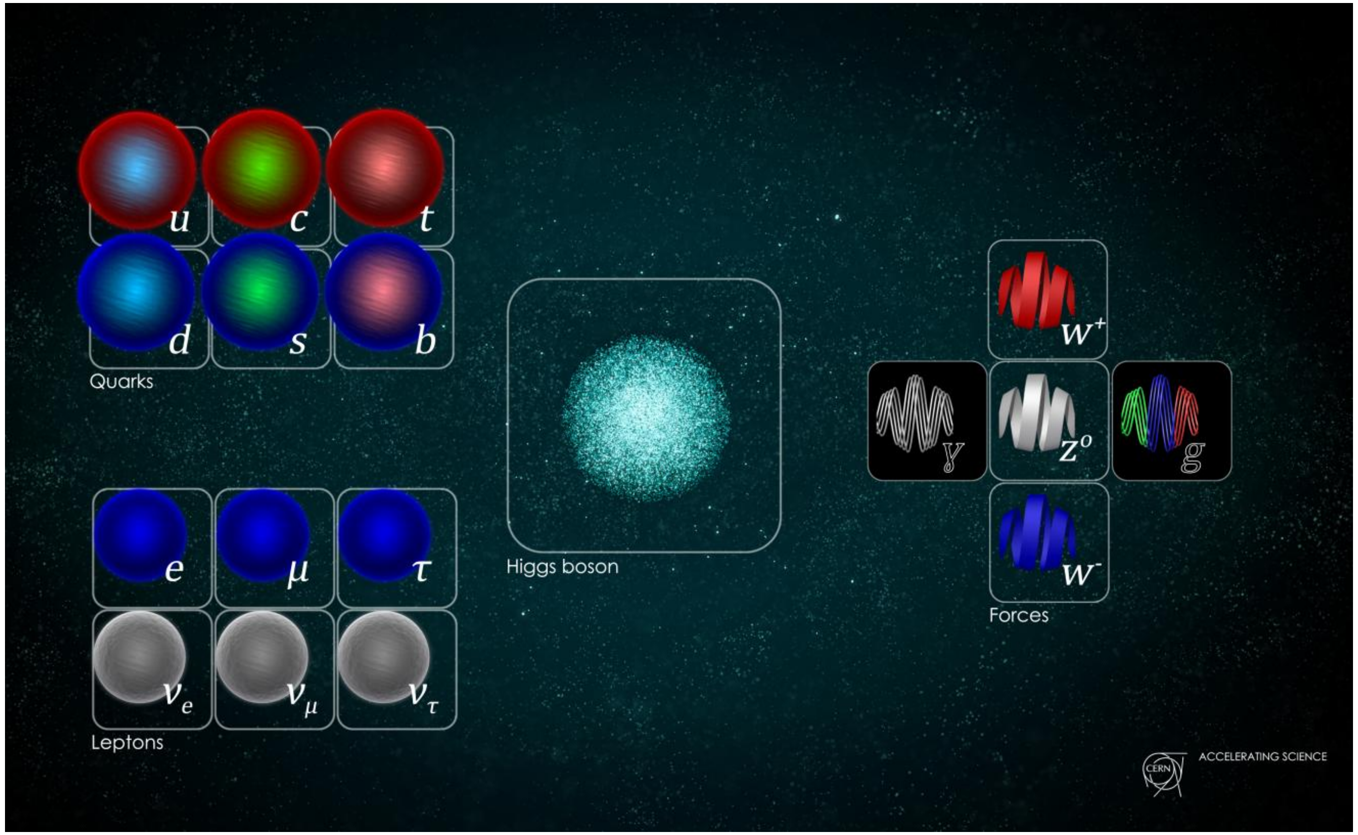
高能物理实验数据分析

CMS Collaboration

2026/07

标准模型与希格斯机制

希格斯场赋予电弱规范玻色子和费米子质量，并预言可被测量的希格斯耦合。



从质量起源到希格斯测量

电弱对称性自发破缺

希格斯场具有非零真空期望值

粒子质量与耦合相关

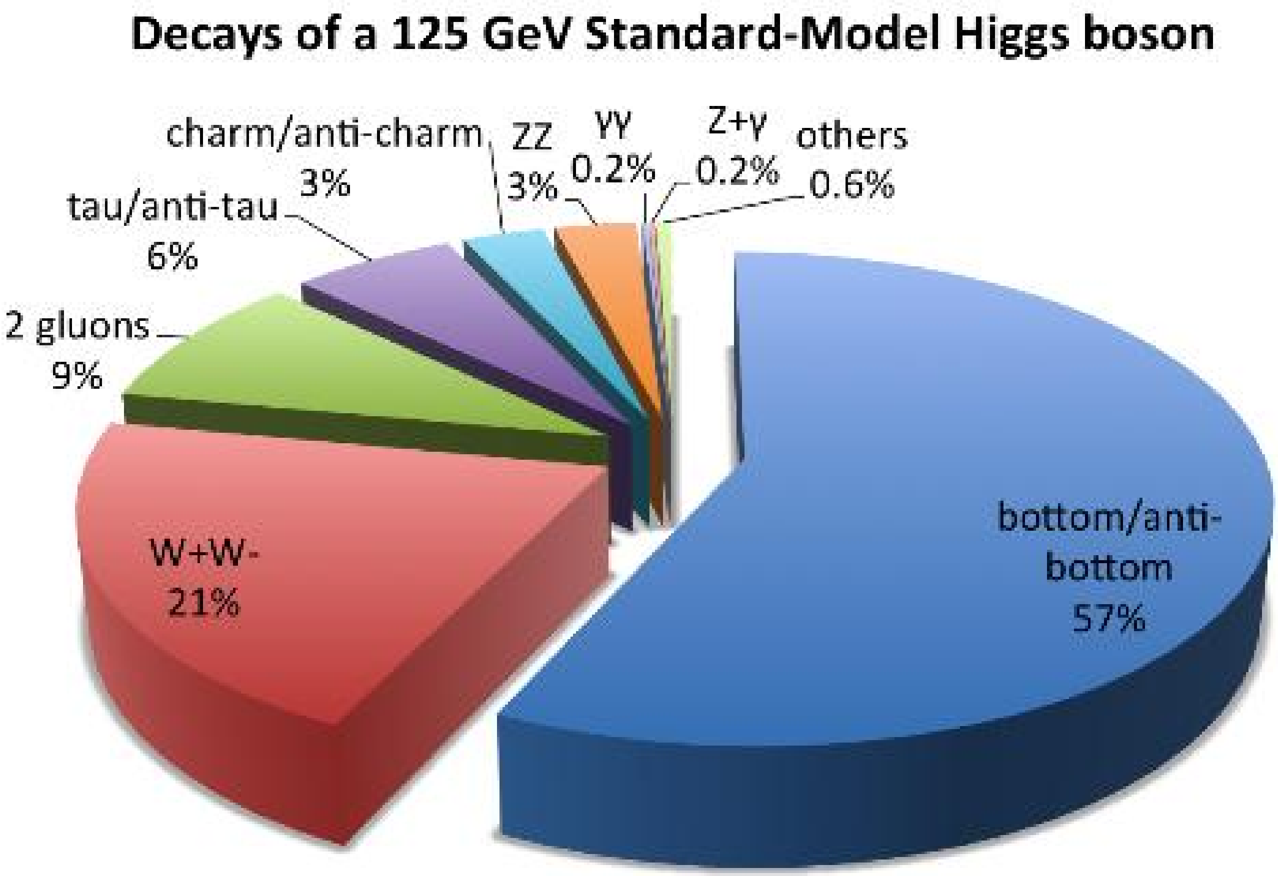
W/Z 玻色子和费米子质量由相互作用强度决定

希格斯玻色子成为可测探针

产生率和衰变率直接检验标准模型预言

标准模型中的希格斯测量

希格斯发现后，核心问题从“是否存在”转向“如何耦合”。



发现后的问题

希格斯是否完全符合标准模型预言？

测量目标

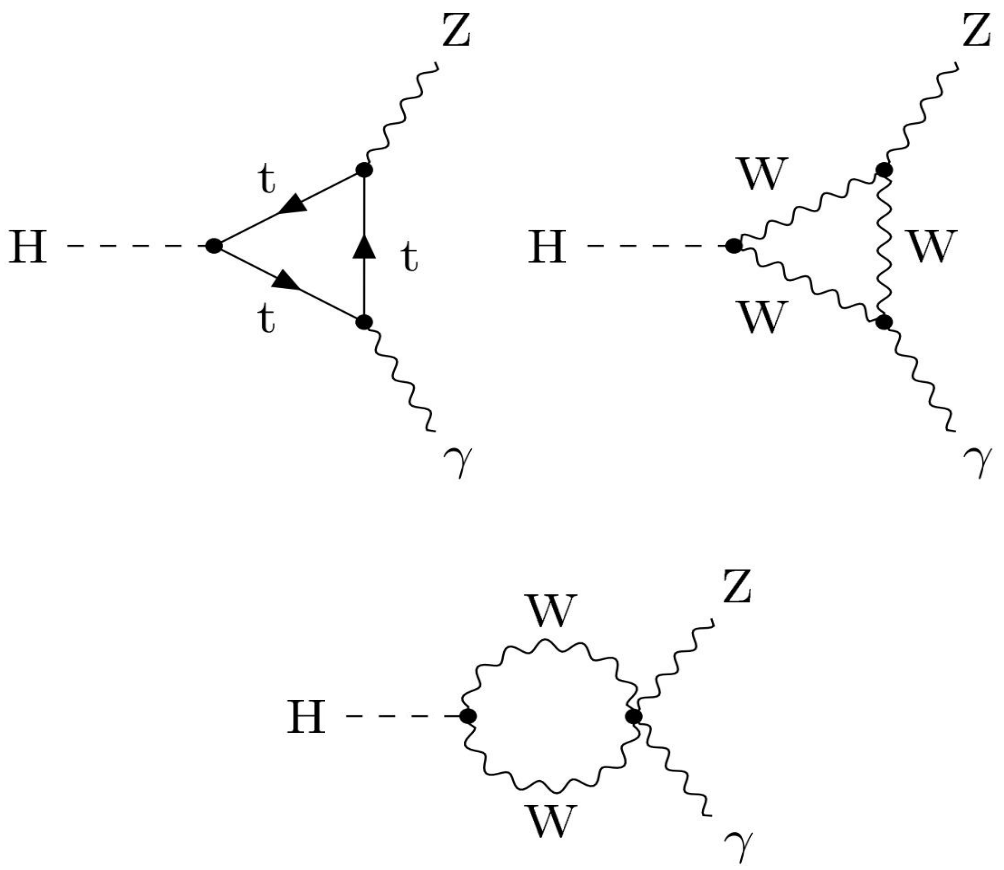
系统测量产生截面、衰变分支比
以及与规范玻色子和费米子的耦合。

稀有衰变价值

统计量小，但对超出标准模型新粒子
和异常耦合具有更高敏感性。

H→Zγ 的物理特征：电弱圈耦合探针

H→Zγ 稀有，但对新物理具有独特敏感性。



标准模型分支比

1.5×10^{-3}

实验统计量极低

为什么值得测量？

即使统计量很小，圈图过程
仍能放大虚粒子或异常耦合
对希格斯电弱相互作用的影响。

三点物理特征

1. 圈图主导

树图被禁止，量子圈图贡献可见

2. 耦合互补

与 H→γγ 与 H→ZZ 测量互补

3. 新物理窗口

约束异常多玻色子有效耦合

Run2 ATLAS + CMS 联合分析结果

Phys. Rev. Lett. 132, 021803

北京大学

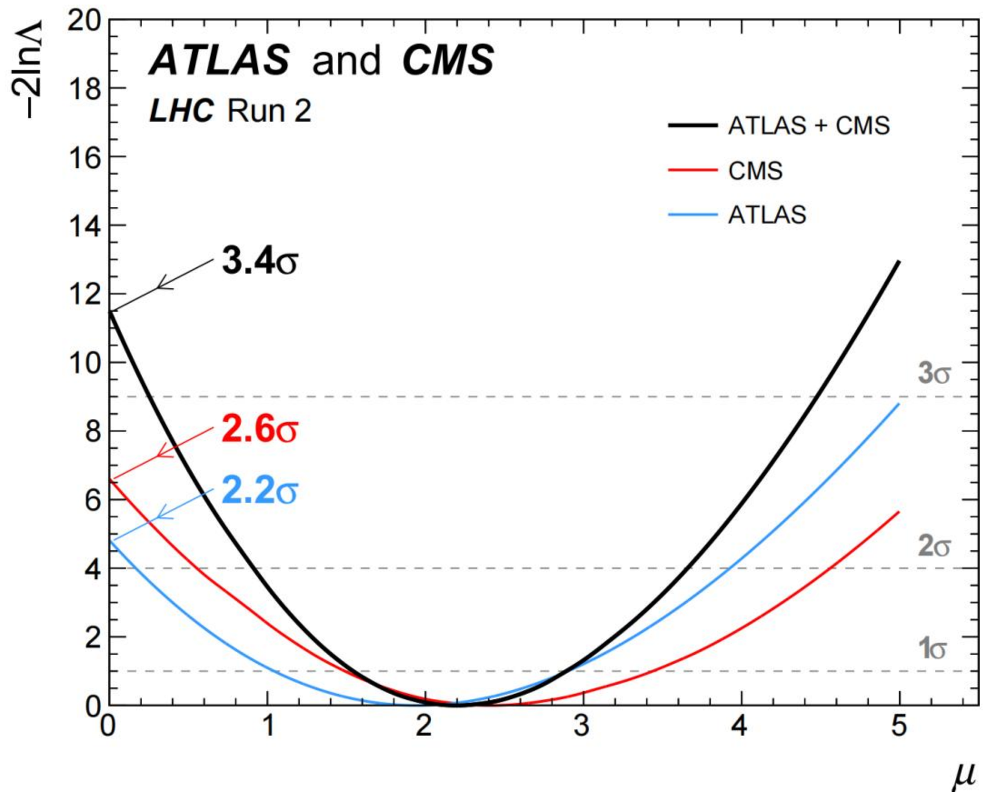
ATLAS与CMS独立地对H→Zγ衰变道进行了寻找

首个实验证据

3.4σ

信号强度

$2.2 \pm 0.6(\text{统计})^{+0.3}_{-0.2}(\text{系统})$



统计误差仍占主导

建立了 H→Zγ 衰变存在的第一手证据，但误差仍主要来自有限统计量。
使用完整 Run 3 的更高统计量继续分析，是从“证据”走向“发现”的自然路径。

ℓℓγ 末态搜索

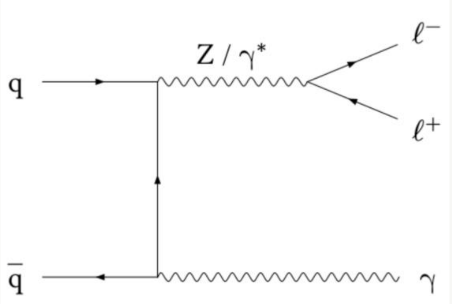
实验挑战是在巨量的连续本底上寻找希格斯质量附近的小信号窄峰。

H → Zγ → ℓℓγ

选择电子对或缪子对重建 Z，
用三体不变质量提取窄信号峰。

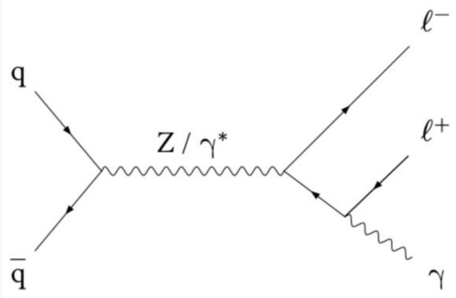
两大主要本底：

SM Zγ



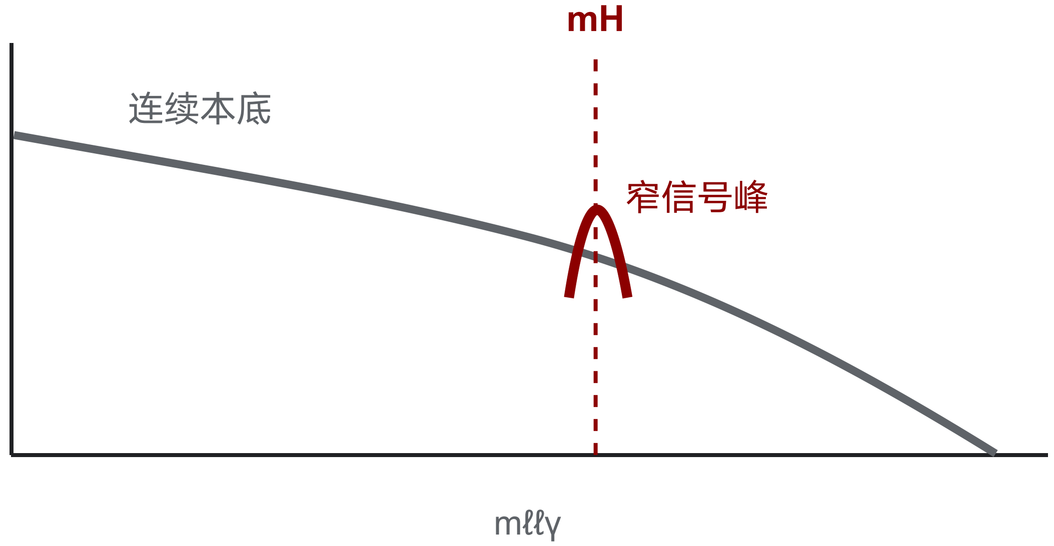
真实光子本底

Z+jets



喷注误鉴别为光子

mℓℓγ 质量谱示意



关键任务：提高信号效率、优化分类，并控制连续本底的函数偏差。

数据样本与触发

约 200 fb^{-1} 数据和新的触发策略显著扩大可用信号样本。

数据集

≈ 200 fb^{-1}

Run 2 UL 数据

2022 + 2023 早期 Run 3 数据

电子道与缪子道同时分析

触发策略

单轻子 ∪ 双轻子触发

电子道效率 **80%→93%**

缪子道效率 **87%→97%**

逻辑或组合提高接受度，同时保持本底可控

模拟样本

信号：ggF、VBF、VH、ttH

峰状本底：$H\to\mu\mu$+FSR γ

非共振本底：Zγ、Z+假光子

用途：优化、验证、拟合测试

触发效率提升直接增加可用信号样本，是后续灵敏度提升的第一步。

对象重建优化

对象选择与 Z 质量约束并列提升信号效率和质量分辨率。

电子/缪子重建

有针对性的轻子选择，
提升信号保留效率

光子选择

抑制喷注误鉴别光子
和软光子贡献

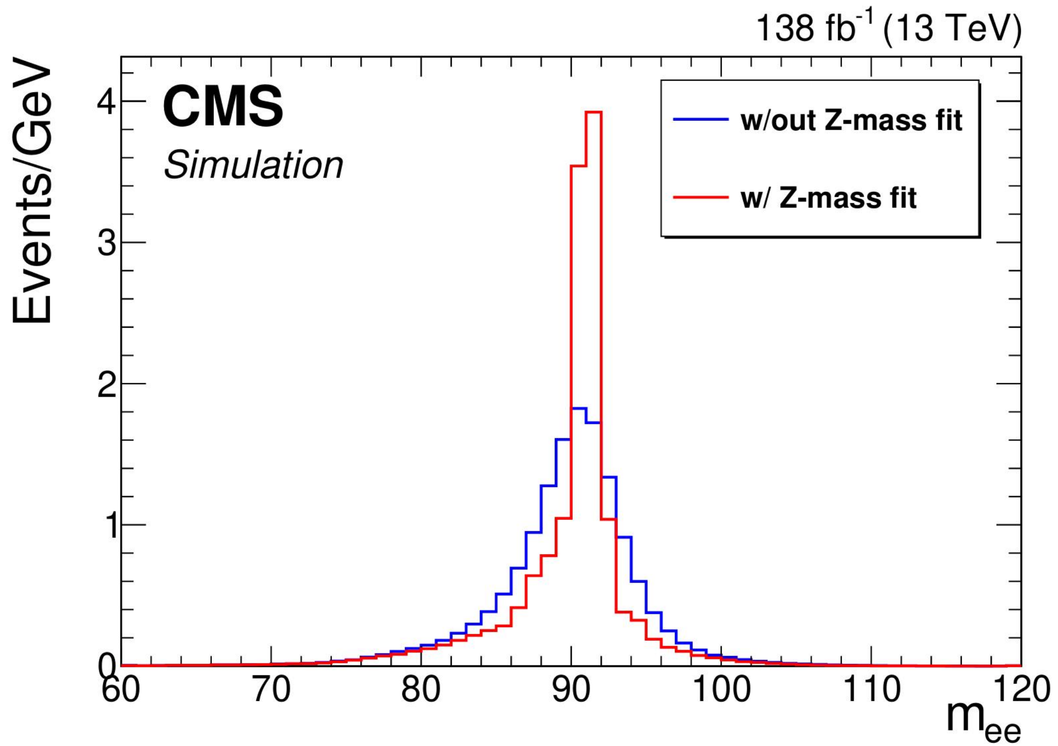
Z 线形约束拟合

利用 Z 玻色子质量线形
重新约束轻子动量
使 $m_{\ell\ell\gamma}$ 峰更窄

分辨率提升

电子道约 20%

缪子道约 10%



窄峰搜索中，分辨率改善会直接提升信号与连续本底的区分能力。

基础事例选择

基础选择压低 FSR 和软光子本底，同时保持信号效率与质量谱稳定。

选择项	具体条件	对应作用
基本末态	两个同味反号轻子 + 一个光子	保留 $H\to Z\gamma\to \ell\ell\gamma$ 信号拓扑
Z 候选	$80\ \mathrm{GeV} < m_{\ell\ell} < 100\ \mathrm{GeV}$	保留 $Z\to\ell\ell$，压低低质量轻子对本底
光子硬度	$p_T(\gamma) / m_{\ell\ell\gamma} \geq 15/110$	抑制软光子，稳定信号窗口附近形状
轻子-光子分离	$\Delta R(\ell,\gamma) > 0.4$	压低末态辐射光子贡献
FSR 抑制	$m_{\ell\ell\gamma} + m_{\ell\ell} > 185\ \mathrm{GeV}$	去除靠近 Z 峰的辐射光子事件
拟合窗口	保留 $m_{\ell\ell\gamma}$ 分析质量区间	保证连续本底可用解析函数描述

还有一些其他的筛选，例如触发器横动量要求、事件过滤器等此处并未一一列出

主要剩余本底：Zγ 初态辐射与 Z+jets 中喷注误鉴别为光子。

产生模式分类

初级分类通过锚定不同产生模式来提高整体分析灵敏度。

初级分类表：类别互斥且包含全部事例与主要产生模式

Category name	Lepton, jet, and p_T^{miss} requirements				Optimization selections		
	N_ℓ	N_{jet}	$N_{b,med.}$	p_T^{miss} [GeV]	$m_{\ell\ell}$ [GeV]	$p_T(\ell\ell\gamma)/m_{\ell\ell\gamma}$	I_{mini}
ggF	$= 2$	≤ 1		< 90	Further binned by BDT		
VBF	$= 2$	≥ 2	$= 0$		Further binned by BDT		
$t\bar{t}H$ leptonic	$= 3$	≥ 3	≥ 1				< 0.1
	≥ 4	≥ 1	≥ 1				
VH $N_\ell \geq 3$	≥ 3		$= 0$	> 30		> 0.3	< 0.15
$t\bar{t}H$ hadronic	$= 2$	≥ 5	≥ 1		[85,95]		
VH p_T^{miss}	$= 2$	≤ 1		> 90		> 0.4	
Untagged	Events passing baseline selection not in another category						

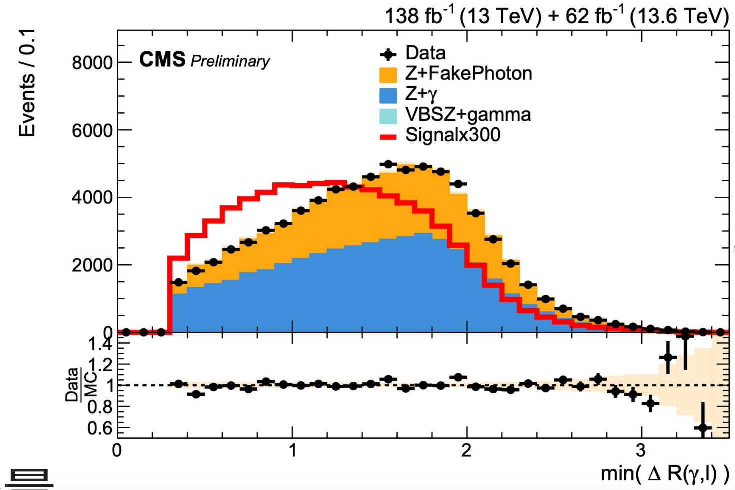
类别由喷注、轻子、缺失横能量和 b 标记等拓扑定义

剩余 ggF/VBF 主类别再进入 BDT 二次分类提升灵敏度

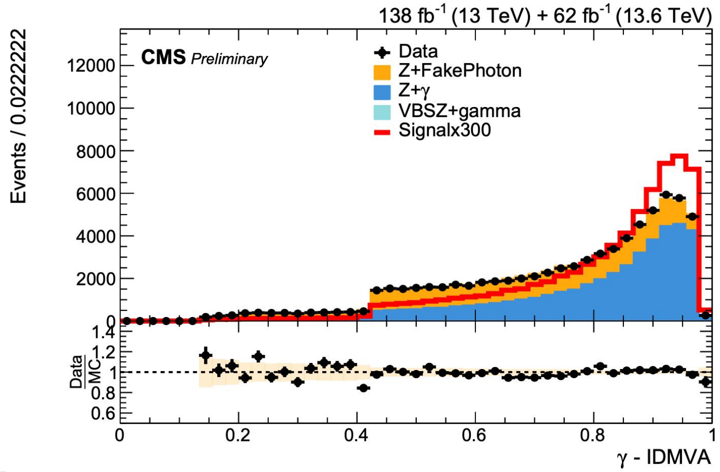
提升决策树(BDT)二次分类与四折验证

物理动机：提升最终显著度　　　　**四折交叉验证：**检查过训练与折间波动，后续偏差测试可利用全部统计量

部分输入变量

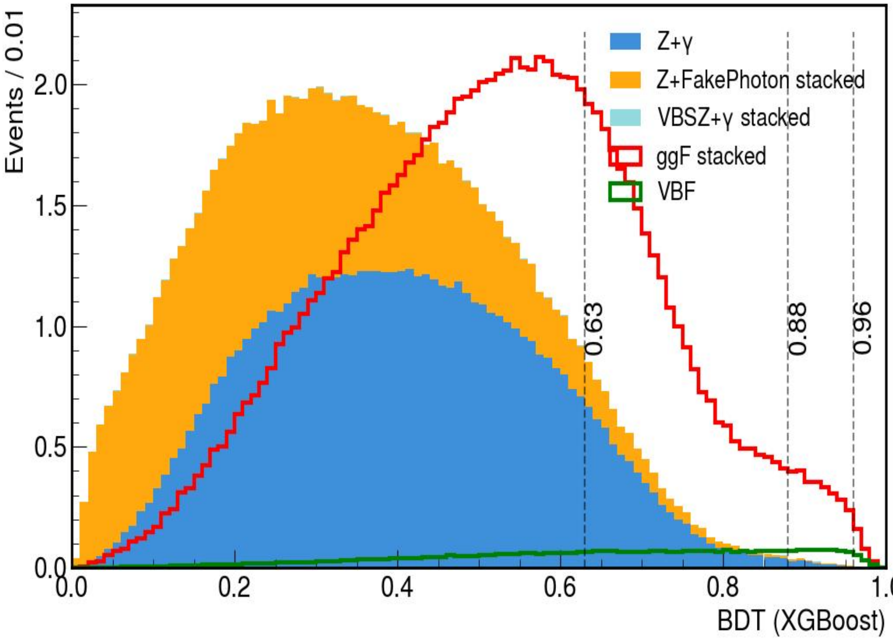


......



BDT输出

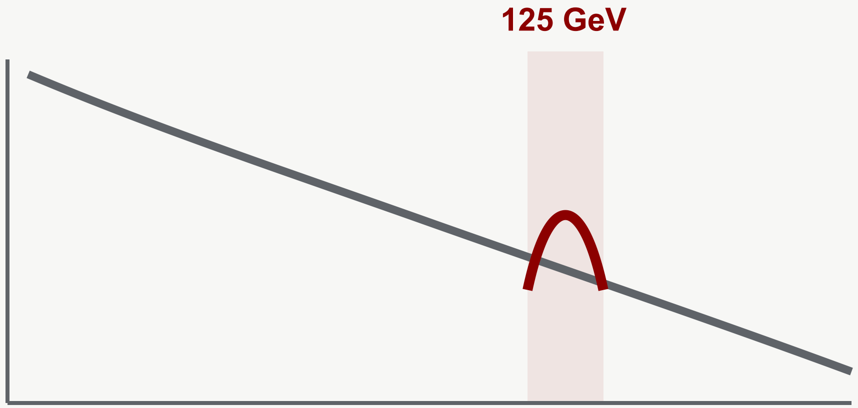
本底与信号的 BDT 分数分布



本底建模策略

本底由平滑的本底解析函数来描述，小信号提取的关键风险来自本底函数偏差。

为什么本底建模关键？



窄信号峰叠加在平滑连续谱上

建模流程

拟合优度检验
筛除不能描述
本底形状的函数

→

假信号检验
评估本底函数
制造的假信号

↓

F-检验
控制函数阶数与复杂度

←

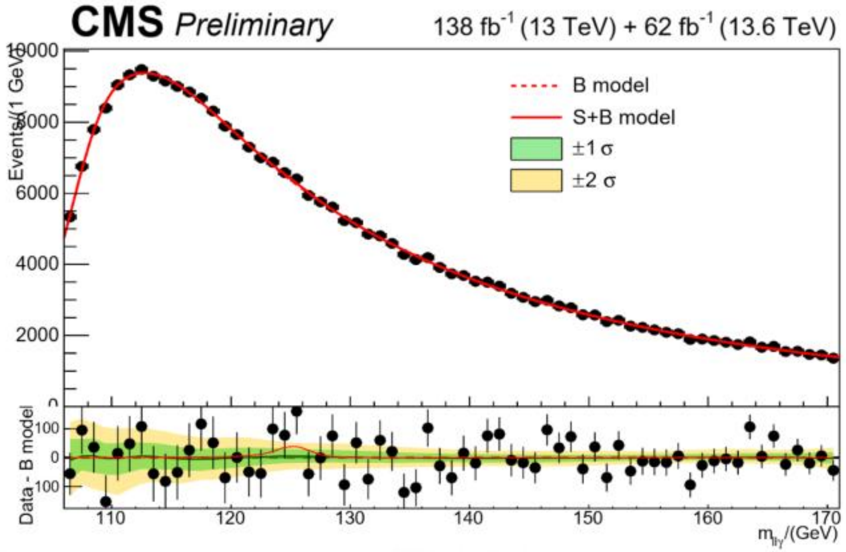
偏差测试
验证最终函数包络

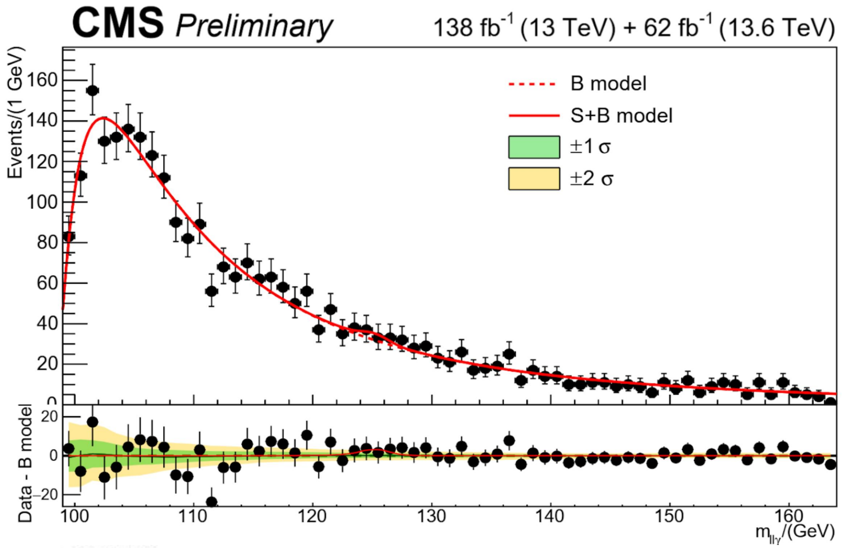
候选本底函数族： Bernstein多项式、指数函数和、幂律函数和、Laurent级数等

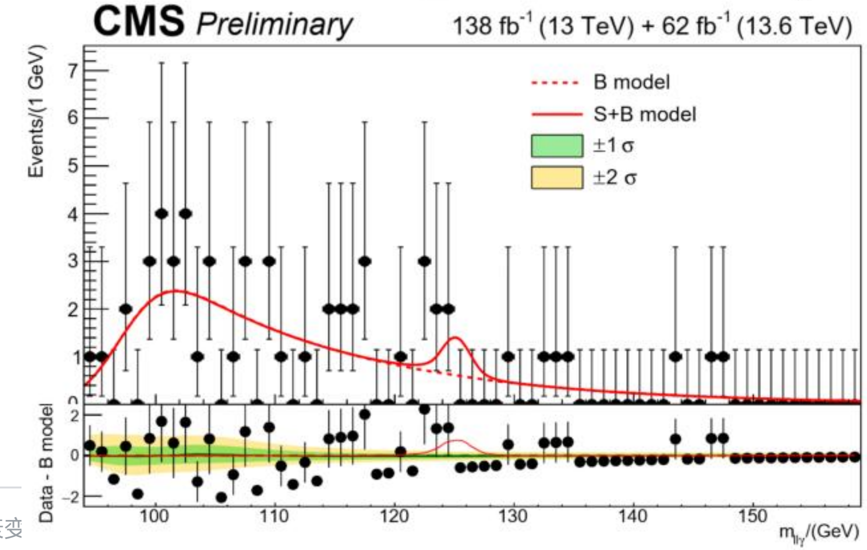
目标： 通过细致偏差估计，降低因函数选取导致的物理结果偏差。

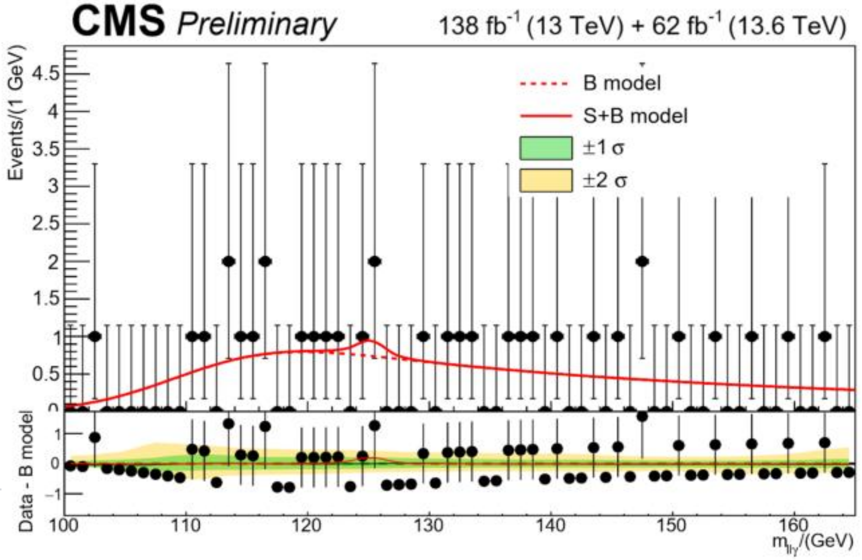
本底拟合建模

通过一系列严格的偏差测试后，最终拟合得到的结果（以其中四个子类别为例）







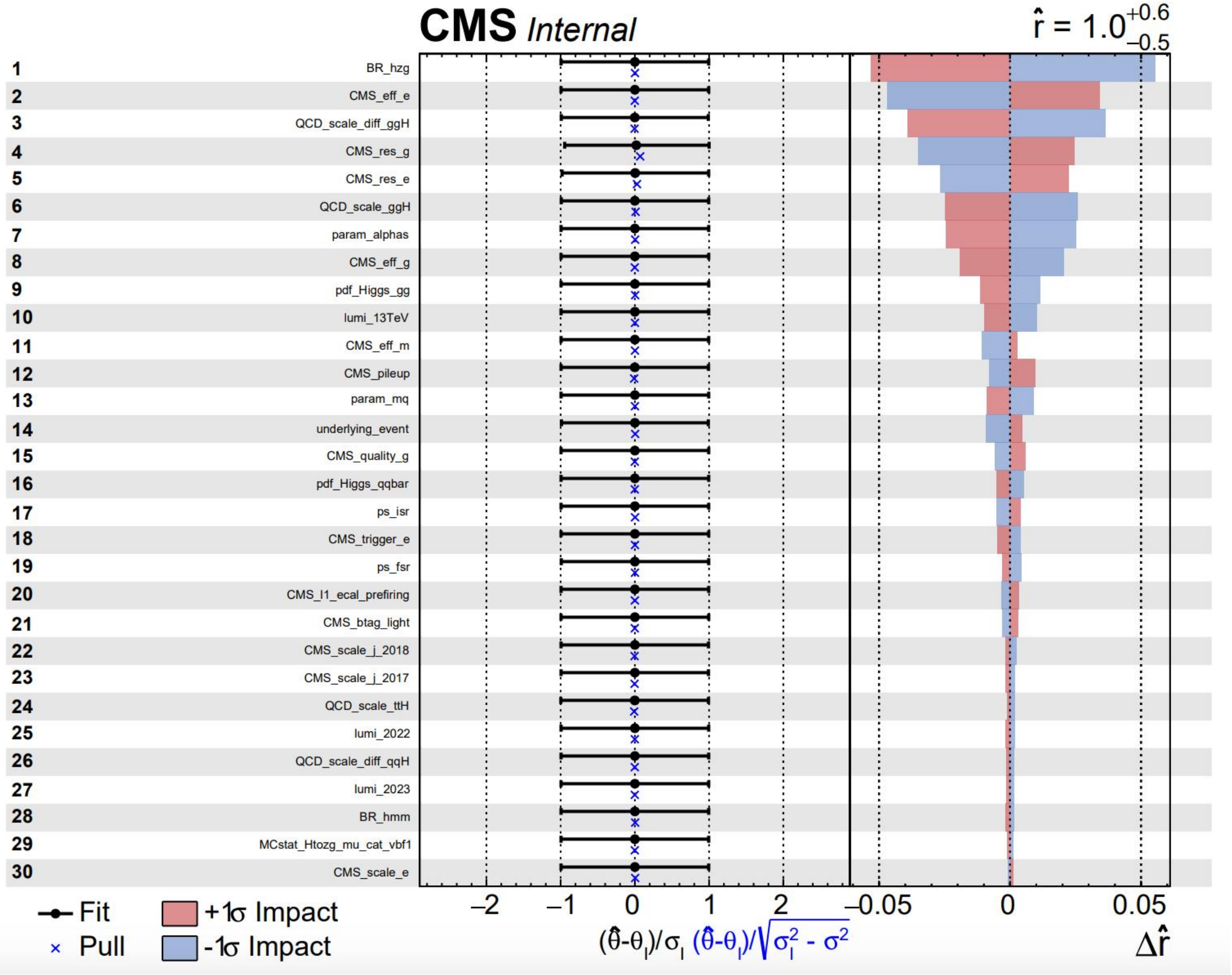


系统不确定度

系统不确定度纳入统计模型后，当前结果中统计误差仍占主导。

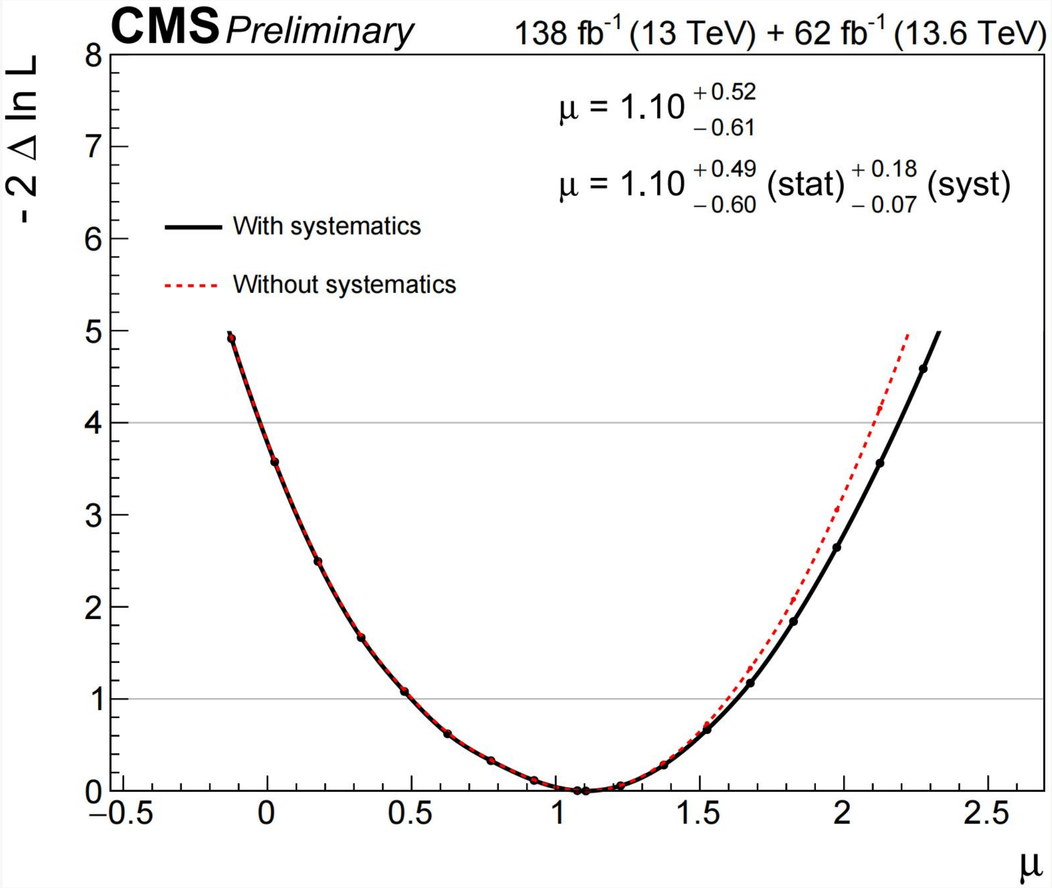
来源类别	主要内容	对信号强度的影响
QCD计算截断	触发、重建、鉴别效率	9%
尺度与分辨率	轻子/光子能量尺度、质量分辨率	4%
光子质量	光子品质变量分布	4%
模型选择	由离散轮廓方法纳入统计误差中	+8%，-36%

其他的不确定度，例如亮度、堆积顶点等效应，这里没有一一列出



信号强度结果

信号强度结果表明 CMS 结果与标准模型预言高度一致。



总信号强度

$\mu = 1.1$

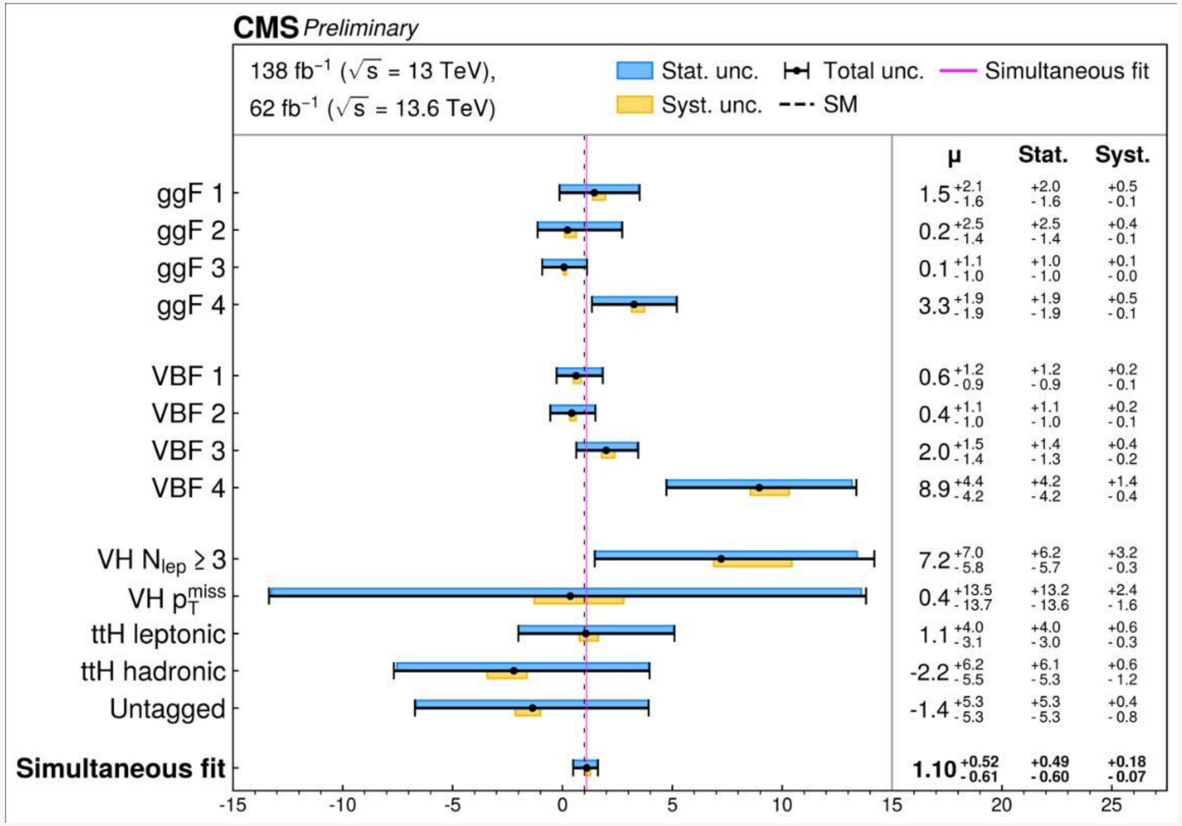
各通道兼容性

$p = 0.75$

2026 Moriond 国际会议公开

信号强度结果

信号强度结果表明 CMS 结果与标准模型预言高度一致。



总信号强度

$\mu = 1.1$

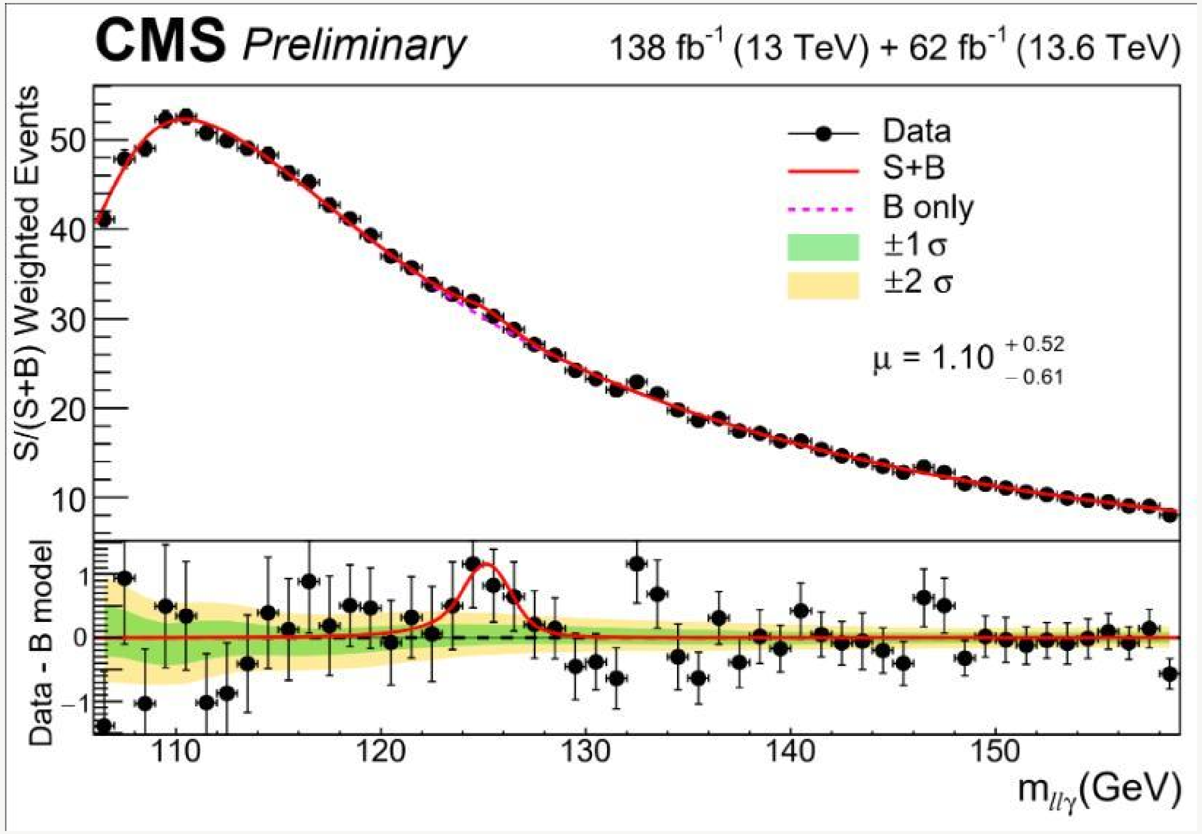
各通道兼容性

$p = 0.75$

2026 Moriond 国际会议公开

显著度结果

CMS Run 2+早期 Run 3 数据分析结果中观测到与标准模型一致的信号迹象。



观测显著度

1.9σ

预期显著度

2.3σ

结果与标准模型预言一致

总结与未来展望

完成了 CMS Run 2+早期 Run 3 数据分析（共计 ~ 200 fb^{-1}），得到与标准模型一致的结果。

Run 2联合分析	Run 2 + 部分 Run 3	Run 2 + Run 3	ATLAS + CMS
3.4σ 联合证据	**≈ 200 fb^{-1}** **观测(预期)显著度** **1.9(2.3)σ**	≈ 500 fb^{-1} 3σ“独立证据水平”	5σ “发现水平”

未来路线 用 CMS 完整 Run 3 推进实验证据，再联合 ATLAS 与 CMS 的 Run 3 独立结果。

物理目标 把 H→Zγ 从“已有证据”的稀有过程推进为可用于精密耦合检验的测量对象。

Backup

$$\mathscr{L}_{\mathrm{EW}} = \mathscr{L}_{\mathrm{gauge}} + \mathscr{L}_{\mathrm{fermion}} + \mathscr{L}_{\mathrm{Higgs}} + \mathscr{L}_{\mathrm{Yukawa}},$$

$$\mathscr{L}_{\mathrm{gauge}} = -\frac{1}{4} W^{i}_{\mu\nu} W^{\mu\nu i} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu},$$

$$\mathscr{L}_{\mathrm{fermion}} = \bar{\psi}_L i\gamma^{\mu} D_{\mu} \psi_L + \bar{\psi}_R i\gamma^{\mu} D_{\mu} \psi_R,$$

$$\mathscr{L}_{\mathrm{Higgs}} = (D^{\mu}\phi)^{\dagger} D_{\mu}\phi - V(\phi) = M_W^2 W^{\mu+} W^{-}_{\mu} (1+\frac{H}{v})^2 + \frac{1}{2} M_Z^2 Z^{\mu} Z_{\mu} (1+\frac{H}{v})^2 + \frac{1}{2} (\partial_{\mu} H)^2 - V(\phi),$$

$$\mathscr{L}_{\mathrm{Yukawa}} = -\zeta_f [\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \bar{\phi} \psi_L],$$

其中

$$W^{\pm} = \frac{1}{\sqrt{2}} (W^1 \mp i W^2),$$

$$Z = -\sin\theta_W B + \cos\theta_W W^3,$$

赝标量南部-戈德斯通玻色子模型

北京大學

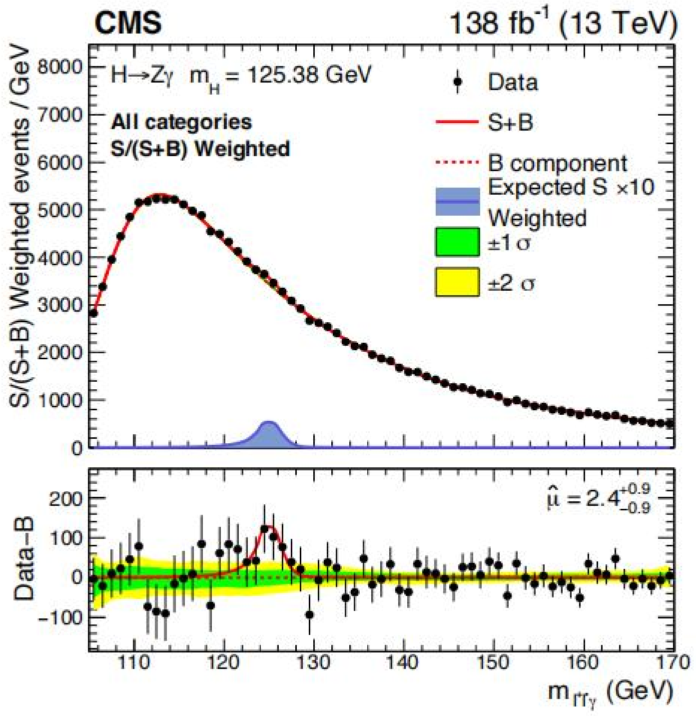
$$\mu(h\to VV^*) = \frac{\sigma_h}{\sigma_h^{SM}}\cdot\frac{\Gamma_{total}^{SM}}{\Gamma_{total}}\cdot F_{O_H}\cdot F_{PNGB},$$

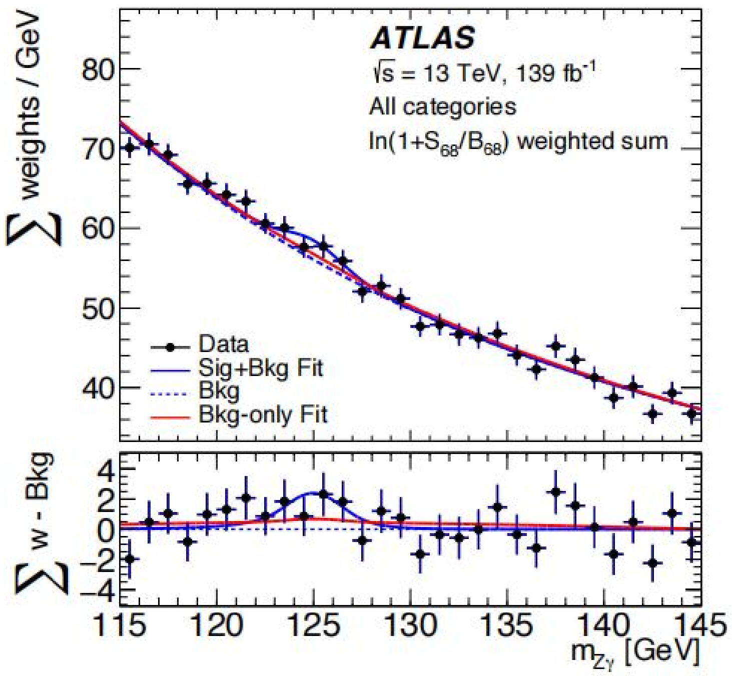
$$\mu(h\to Z\gamma) = \frac{\sigma_h}{\sigma_h^{SM}}\bullet\frac{\Gamma_{total}^{SM}}{\Gamma_{total}}\bullet F_{O_H}\bullet\frac{|F_{Z\gamma}^{t}+F_{Z\gamma}^{W}\sqrt{F_{PNGB}}+\Delta\kappa_{Z\gamma}\tan\theta_W|^2}{|F_{Z\gamma}^{t}+F_{Z\gamma}^{W}|^2},$$

$$R\equiv\frac{\mu(h\to Z\gamma)}{\mu(h\to VV^*)}=\frac{|F_{Z\gamma}^{t}+F_{Z\gamma}^{W}\sqrt{F_{PNGB}}+\Delta\kappa_{Z\gamma}\tan\theta_W|^2}{|F_{Z\gamma}^{t}+F_{Z\gamma}^{W}|^2F_{PNGB}},$$

表 2.1 m_H=125GeV 时质子-质子对撞机上标准模型预言希格斯玻色子产生截面随质心能量的变化

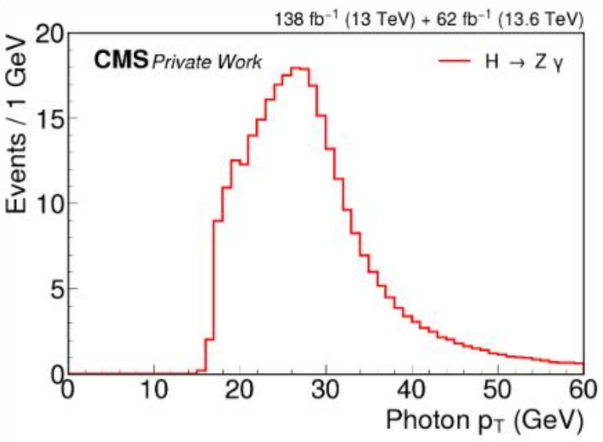
$\sqrt{s}$ (TeV)	当m_H=125GeV 时希格斯玻色子的产生截面(pb)					
	ggF	VBF	WH	ZH	ttH	共计
7	$16.9^{+4.4\%}_{-7.0\%}$	$1.24^{+2.1\%}_{-2.1\%}$	$0.58^{+2.2\%}_{-2.3\%}$	$0.34^{+3.1\%}_{-3.0\%}$	$0.09^{+5.6\%}_{-10.2\%}$	19.1
8	$21.4^{+4.4\%}_{-6.9\%}$	$1.60^{+2.3\%}_{-2.1\%}$	$0.70^{+2.1\%}_{-2.2\%}$	$0.42^{+3.4\%}_{-2.9\%}$	$0.13^{+5.9\%}_{-10.1\%}$	24.2
13	$48.6^{+4.6\%}_{-6.7\%}$	$3.78^{+2.2\%}_{-2.2\%}$	$1.37^{+2.6\%}_{-2.6\%}$	$0.88^{+4.1\%}_{-3.5\%}$	$0.50^{+6.8\%}_{-9.9\%}$	55.1
14	$54.7^{+4.6\%}_{-6.7\%}$	$4.28^{+2.2\%}_{-2.2\%}$	$1.51^{+1.9\%}_{-2.0\%}$	$0.99^{+4.1\%}_{-3.7\%}$	$0.60^{+6.9\%}_{-9.8\%}$	62.1

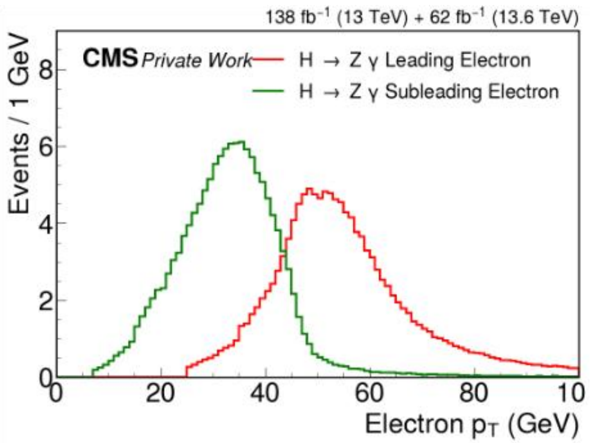


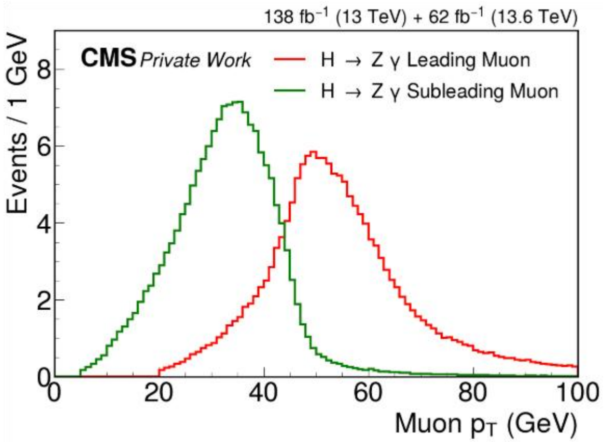


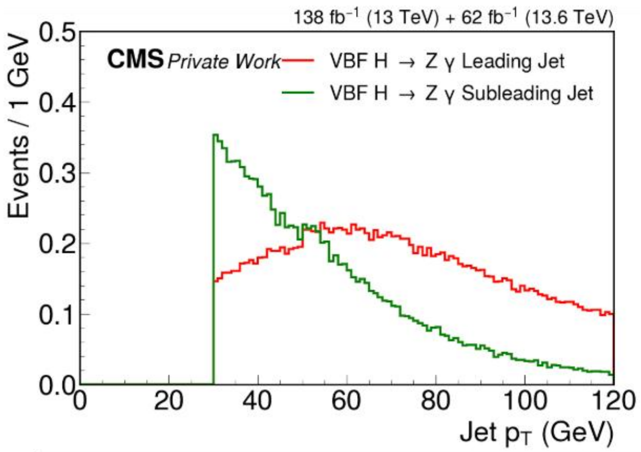
	CMS	ATLAS
No. categories	8	6
Prod modes	ggF, VBF, VH+ttH (lep)	ggF, VBF
Background uncertainty	Discrete profiling	Spurious signal
m_H	125.38 GeV	125.09 GeV
Signal strength	2.4 ± 0.9 (stat) ± 0.3 (syst)	2.0 ± 0.9 (stat) ± 0.4 (syst)
Significance Obs (Exp)	2.7 (1.2)	2.2 (1.2)

Physics Object Definitions









Photon selection
p_T >15 GeV
$\|\eta_{sc}\| < 2.5$ and not gap
$\Delta R_{min}(l,\gamma) > 0.3$ ★
Conversion-safe electron veto, γ/e cleaning requirement
Fall17v2 (run 2)/Winter22 ID (run 3) w/ ISO, WP80 (tighter & standardized) ★

Electron selection
p_T > 7 GeV (trigger pT threshold higher)
$\|\eta_{sc}\| < 2.5$
$\|d_z\| < 1.0$ cm, $\|d_{xy}\| < 0.5$ cm
Fall17v2 ID WPL (run 2) HToZZ(2018) ID with 2022 WP (run 3)

Muon selection
p_T > 5 GeV (trigger pT threshold higher)
$\|\eta\| < 2.4$
$\|d_z\| < 1.0$ cm, $\|d_{xy}\| < 0.5$ cm, SIP3D < 4
Loose ID
I_{rel}<0.35

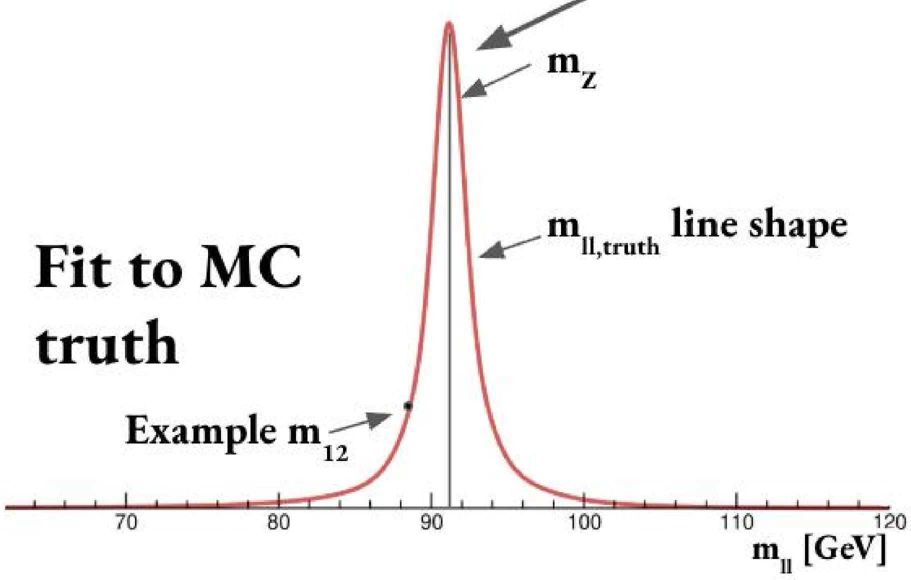
Jet selection
p_T > 30 GeV
$\|\eta\| < 4.7$
Tight jetID, loose PUID
Jet veto (run 3), HEM veto, horn veto

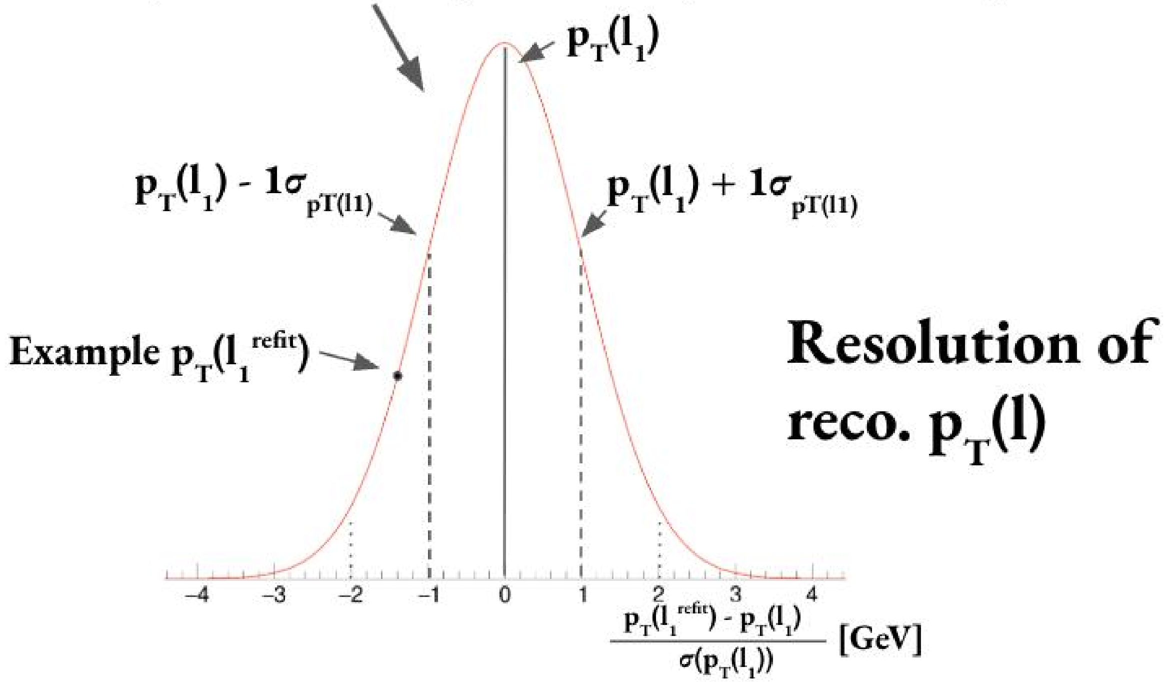
b-tagging with deepJET

Mass resolution improvement using Z-constrained fit

- Allow $p_T(l)$ to vary within resolution to improve m_{ll} using a parametric fit to $m_{ll,truth}$
  1. Fit $m_{ll,truth}$ distribution in signal MC
  2. Unbinned maximum likelihood fit of data constraining m_{ll} to pdf from fit to signal MC

$$L\left(p_{\mathrm{T}}^{\mathrm{refit},1}, p_{\mathrm{T}}^{\mathrm{refit},2}, m_{12} \mid p_{\mathrm{T}}^{1}, p_{\mathrm{T}}^{2}, \sigma_{p_{\mathrm{T}}^{1}}, \sigma_{p_{\mathrm{T}}^{2}}, \vec{\beta}\right) = L\left(m_{12} \mid \vec{\beta}\right) \cdot \mathrm{Gauss}\left(p_{\mathrm{T}}^{\mathrm{refit},1} \mid p_{\mathrm{T}}^{1}, \sigma_{p_{\mathrm{T}}^{1}}\right) \cdot \mathrm{Gauss}\left(p_{\mathrm{T}}^{\mathrm{refit},2} \mid p_{\mathrm{T}}^{2}, \sigma_{p_{\mathrm{T}}^{2}}\right)$$



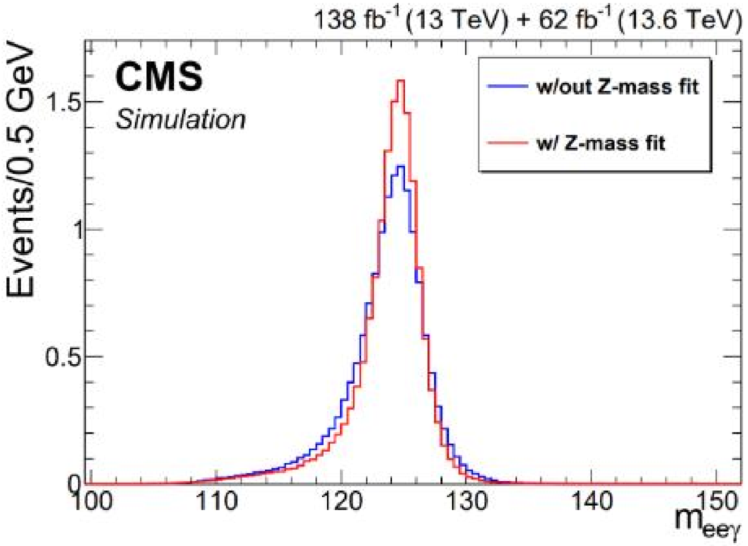


- FSR recovery for $Z \rightarrow \mu\mu$ included in kinematic refit procedure

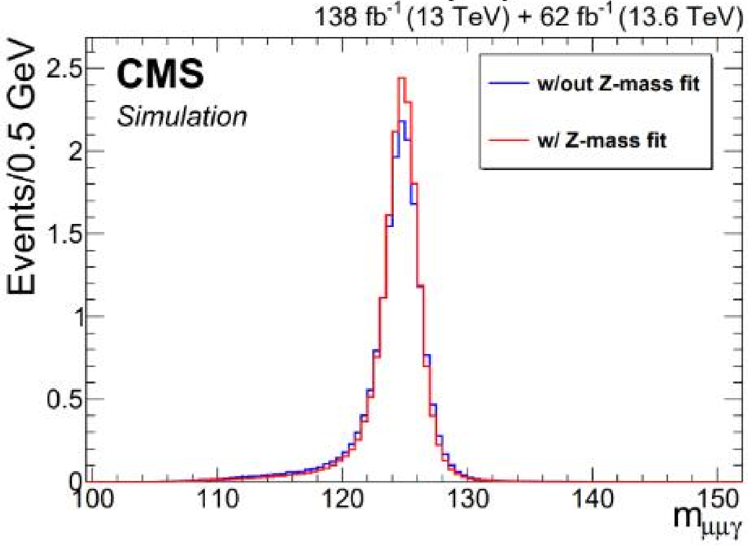
Signal resolution and constrained fit to Z line shape

- Good $m_{ll\gamma}$ signal resolution is crucial for signal extraction
- Constrained fit to Z line shape to improve signal mass resolution
- Signal resolution metric is half of minimum interval containing 68.3% of signal MC ($\sigma_{68.3\%}$)

$Z \rightarrow e^{+}e^{-}$



$Z \rightarrow \mu^{+}\mu^{-}$



Signal mass resolution by category

	$Z \rightarrow e^{+}e^{-}$		$Z \rightarrow \mu^{+}\mu^{-}$	
Category	$\sigma_{68.3\%}$ [GeV]	$\sigma_{68.3\%,\text{refit}}$ [GeV]	$\sigma_{68.3\%}$ [GeV]	$\sigma_{68.3\%,\text{refit}}$ [GeV]
ggF	2.6	2.1	1.7	1.5
VBF	2.5	2.0	1.8	1.5
$t\bar{t}$H leptonic	2.4	1.9	1.7	1.5
VH $N_{\ell} \geq 3$	2.6	2.1	1.8	1.6
$t\bar{t}$H hadronic	2.3	1.8	1.7	1.5
VH p_{T}^{miss}	2.7	2.0	1.9	1.6
Untagged	2.8	2.2	2.0	1.7

- Improves $m_{ll\gamma}$ resolution by 20% (10%) for $Z \rightarrow ee$ ($Z \rightarrow \mu\mu$) events
- Observe minimal to no change in $m_{ll\gamma}$ background shape

基准筛选
初始
$N_e \geq 2 \,\|\|\, N_\mu \geq 2$
触发筛选
触发器横动量筛选
$N_\gamma \geq 1$
$80 \leq m_{ll} \leq 100$ GeV
$p_T(\gamma)/m_{ll\gamma} > 15/110$
$m_{ll} + m_{ll\gamma} > 185$ GeV
$95 \leq m_{ll\gamma} \leq 180$ GeV
事件过滤器

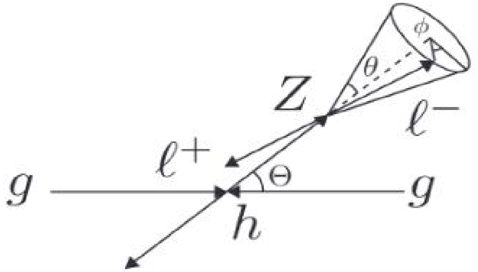
Baseline Selection (w/ N-1 eff.)

Selection	ε_S	ε_B
$N_\gamma \geq 1$, $N_e \geq 2$ (OS) OR $N_\mu \geq 2$ (OS)	–	–
Triggers efficiency ★ plateau $p_T(l)$ selections	0.99	0.99
80 GeV < m_{ll} <100 GeV ★	0.95	0.86
$p_T(\gamma)/m_{ll\gamma}$ > 15/110	0.95	0.83
m_{ll}+ $m_{ll\gamma}$ >185 GeV	0.99	0.99
95 GeV < $m_{ll\gamma}$ < 180 GeV	0.99	0.85
Pass quality (p_T^{miss}) Filters	0.99	0.99

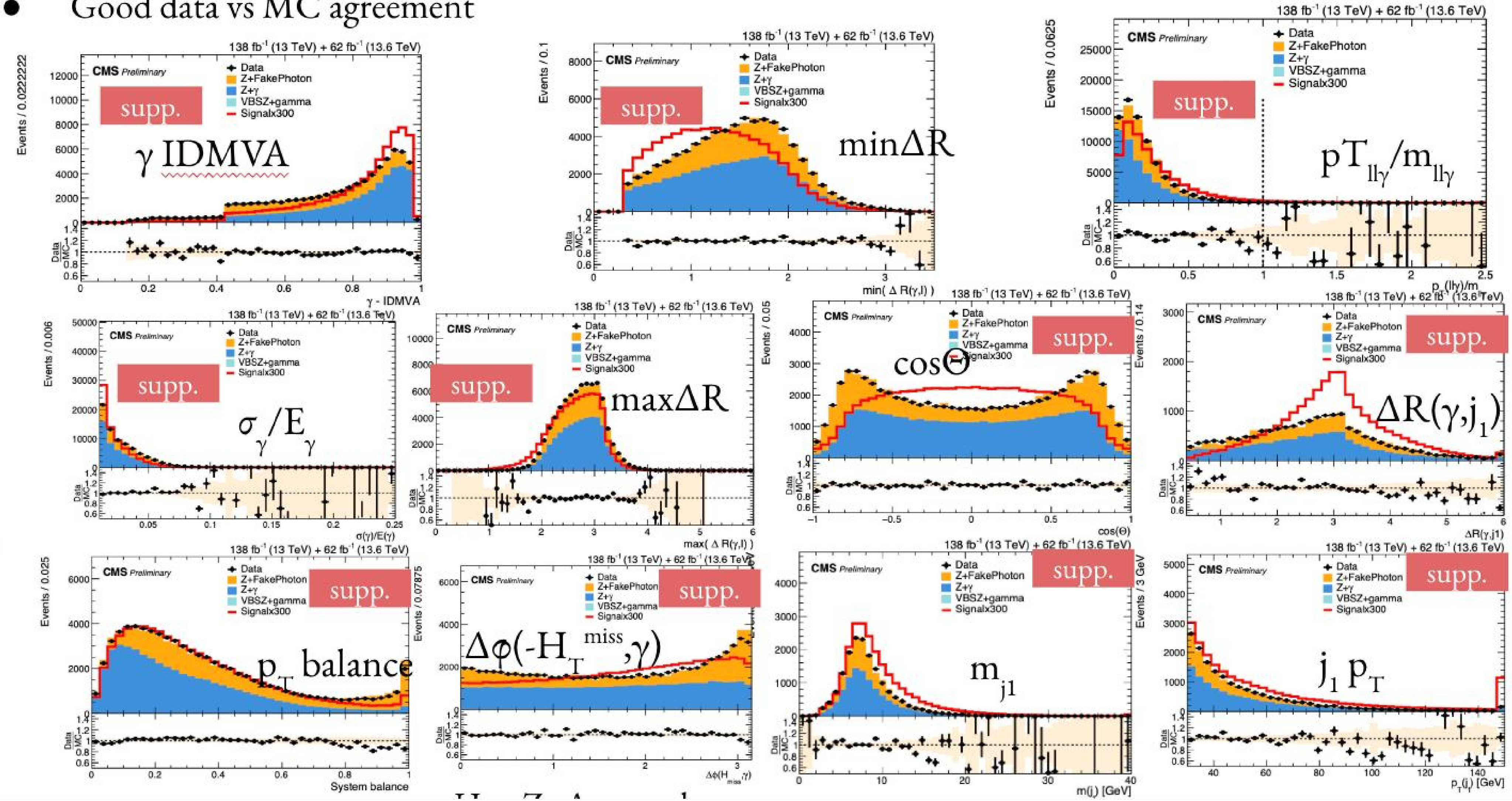
ggF BDT Input Variables (Data v. MC: 120-130 GeV)

Type	Variables
Boosted Higgs	$pT_{ll\gamma}/m_{ll\gamma}$
Photon Quality	photon ID MVA
	σ_γ/E_γ
Angles Between Final State Particles	min $\Delta R(l,\gamma)$
	max $\Delta R(l,\gamma)$
	$\cos\Theta$
	$\cos\theta$
	ϕ
	$\lvert\eta_\gamma-\eta_{j1}\rvert$
	$\Delta R(j_1,\gamma)$
Angles Of Final State Particles	$\eta(l_1)$
	$\eta(l_2)$
	$\eta(\gamma)$
	$\eta(j_1)$
Jet Kinematics	j_1 m
	j_1 p_T
	N_{jets}
System Balance	system p_T balance
	$\Delta\phi(H^{miss}{}_T,\gamma)$
	$\Delta\phi(ll\gamma,j_1)$

Higgs Variables | Jet Variables



- Variables selection optimized by signal and background separation and no correlation with $m_{ll\gamma}$
- Plots of variables below ranked by feature importance
- Good data vs MC agreement

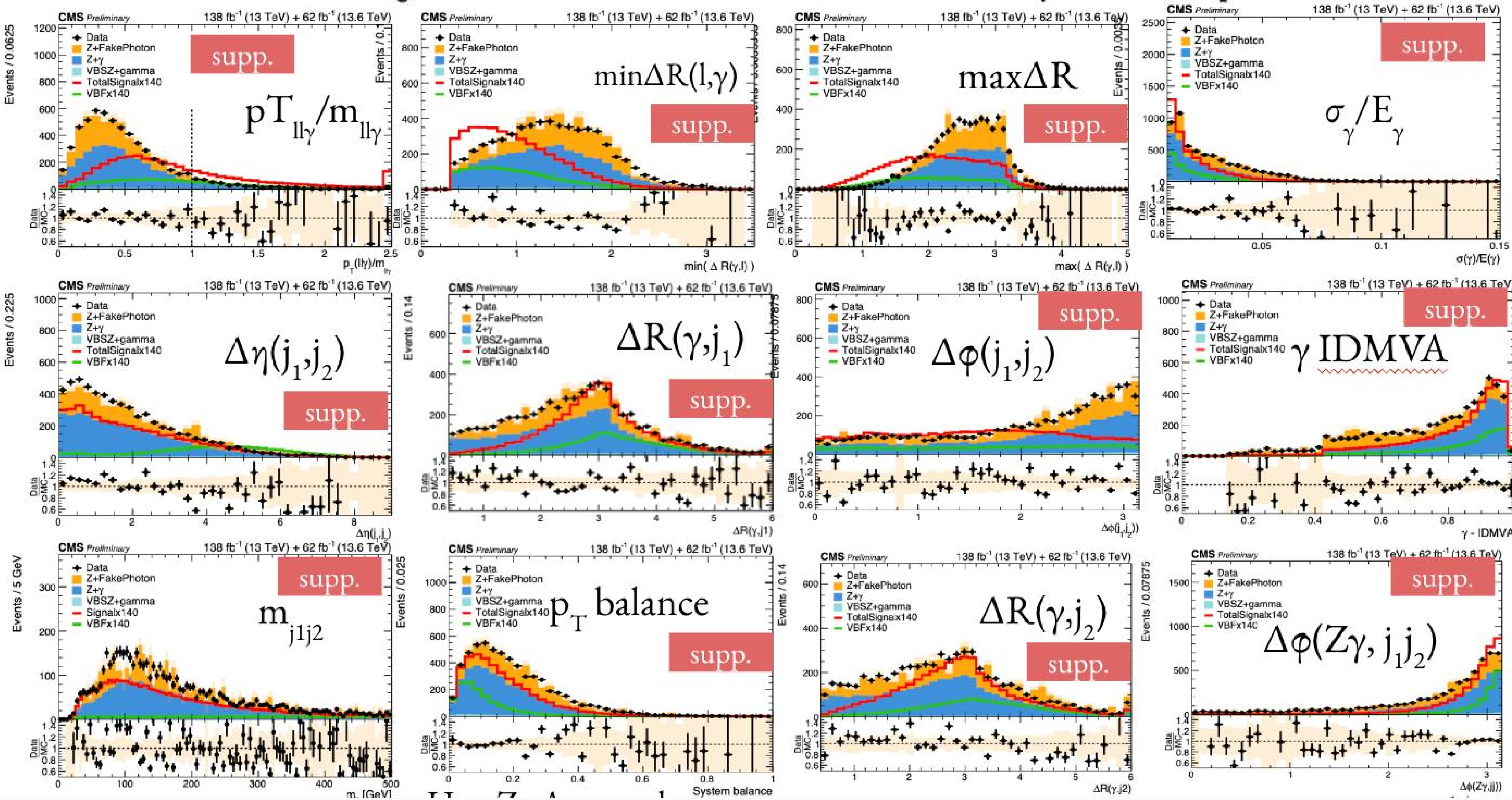


VBF BDT Input Variable (Data v. MC: 120-130 GeV)

Type	Variables
Boosted Higgs	$pT_{ll\gamma}/m_{ll\gamma}$
Photon Quality	photon ID MVA
	σ_γ/E_γ
Angles Between Final State Particles	min $\Delta R(l,\gamma)$
	max $\Delta R(l,\gamma)$
	$\cos\Theta$
	$\cos\theta$
	ϕ
	photon zeppenfeld
	$\Delta R(j_1,\gamma)$
	$\Delta R(j_2,\gamma)$
Angles Of Final State Particles	$\eta(l_1)$
	$\eta(l_2)$
	$\eta(\gamma)$
Jet Kinematics	$m(j_1,j_2)$
	$\Delta\eta(j_1,j_2)$
	$\Delta\phi(j_1,j_2)$
	j_1 p_T
	j_2 p_T
	N_{jets}
System Balance	system p_T balance
	$\Delta\phi(H^{miss}_T,\gamma)$
	$\Delta\phi(ll\gamma,jj)$

Higgs Variables | Jet Variables

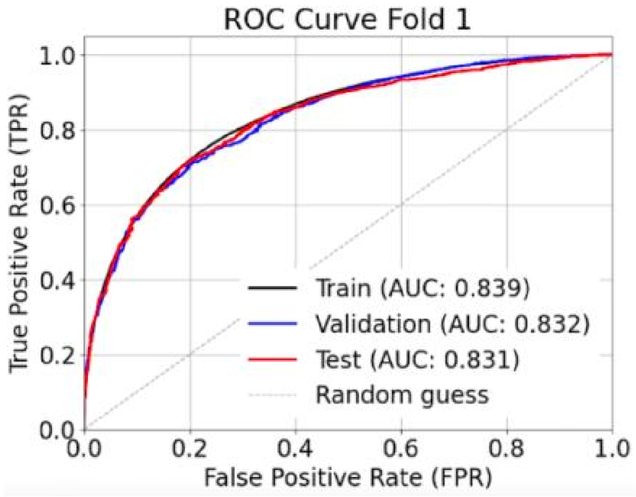
- Background: **Z+γ**, **Z+fake γ**, **VBS Z+γ.** VBF signal marked in **green**, total signal marked in **red**
- Good data v. MC agreement. Plots of variables below ranked by feature importance



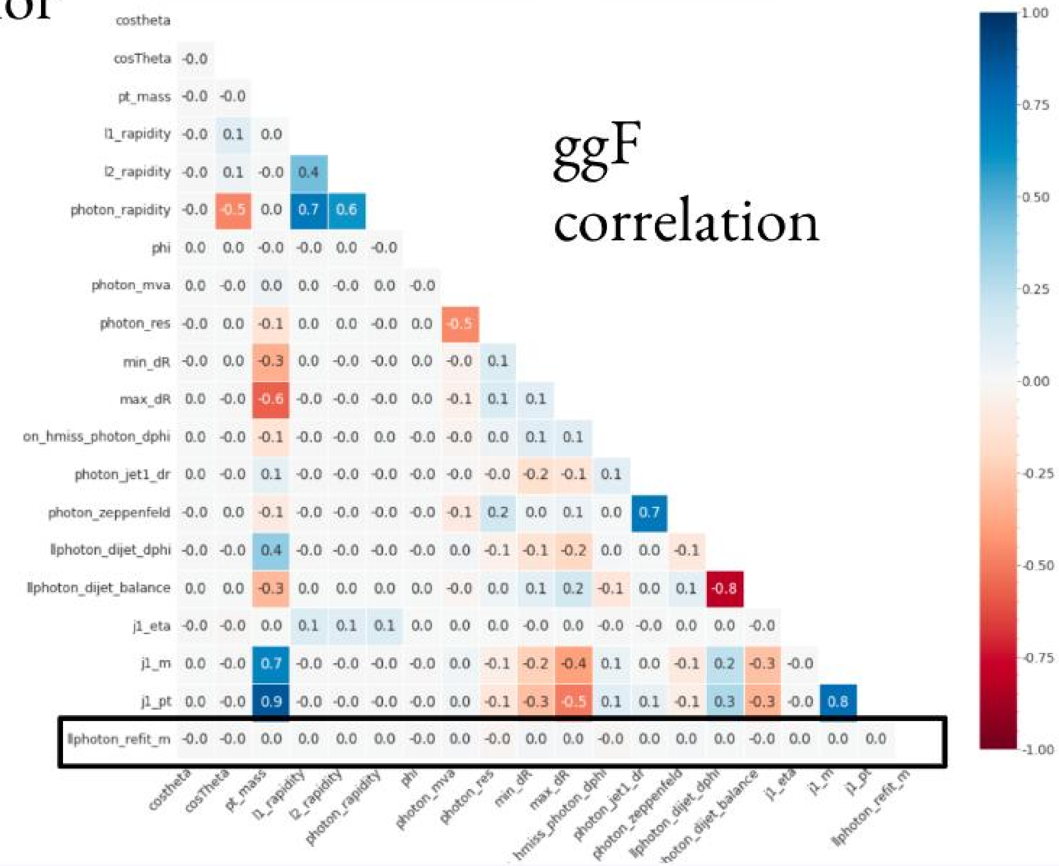
ggF & VBF BDT Optimization

- Input variables selection optimized by S vs B separation power & correlation with $m_{ll\gamma}$
- Hyperparameters optimized to minimize overtraining (validation vs training) and to ensure consistent behavior among folds
- Number of bins optimized by

$$Z_{cc} = \sqrt{2[(N_S + N_B) \cdot \ln(1 + N_S/N_B) - N_S]}$$

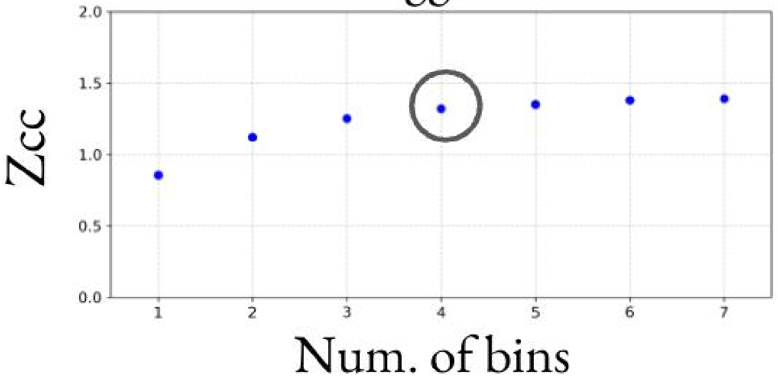


VBF ROC

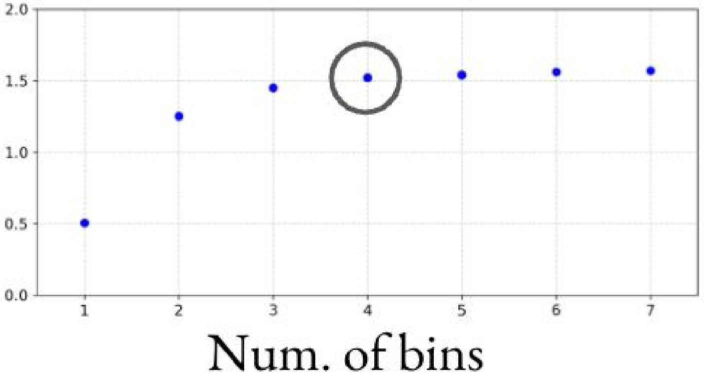


ggF correlation

ggF



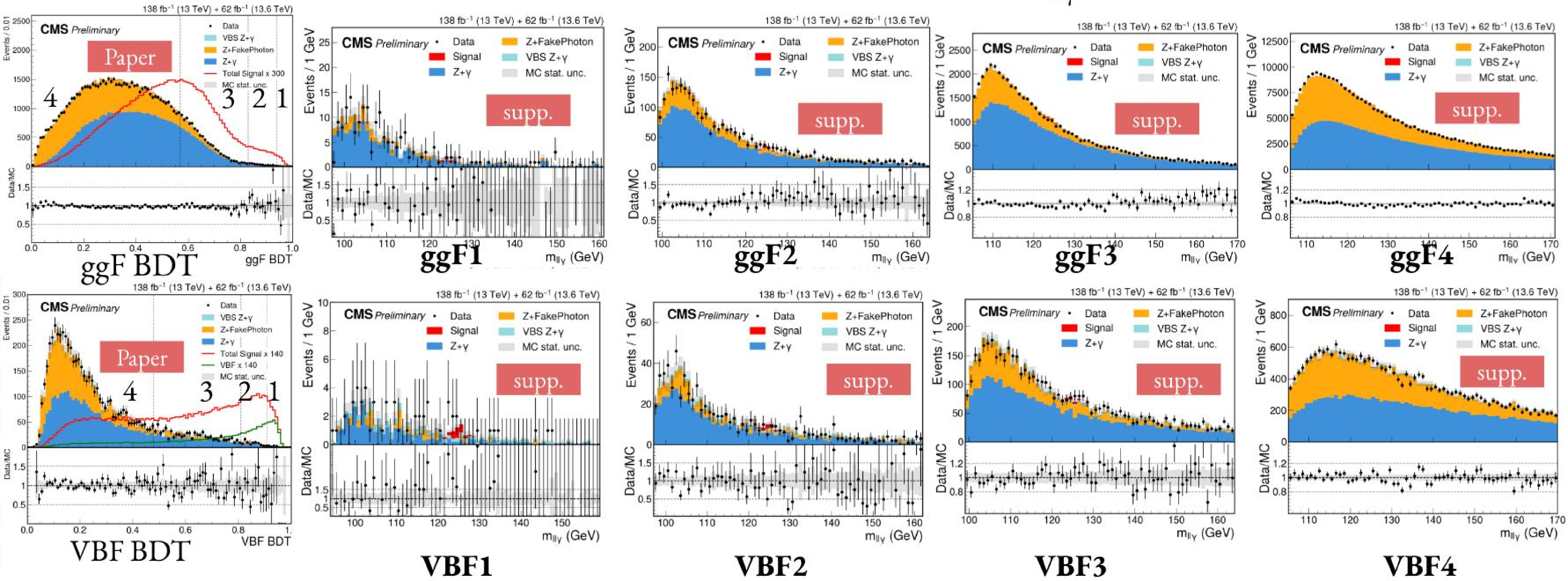
VBF



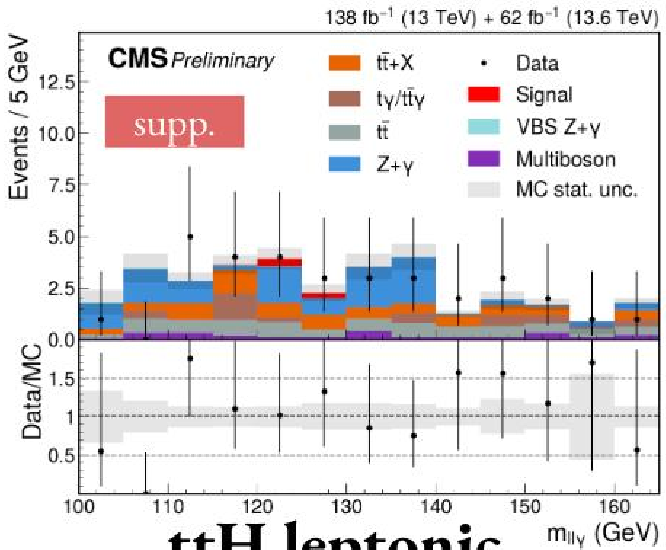
ggF & VBF Categories BDT & $m_{ll\gamma}$ Data v. MC

- Simultaneously move all BDT boundaries and optimize Z_{cc} (120-130)
- Background MC normalized to data sideband
- Good data v. MC observed in BDT score distribution and $m_{ll\gamma}$

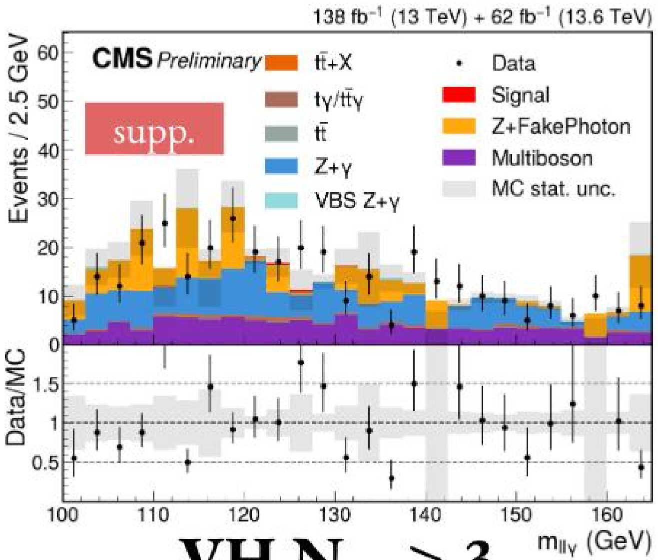
$$Z_{cc} = \sqrt{2[(N_S + N_B) \cdot \ln(1 + N_S/N_B) - N_S]}$$



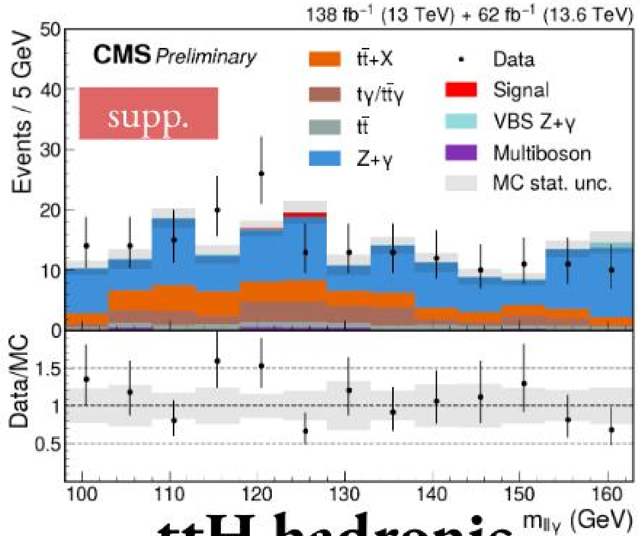
VH & ttH Categories $m_{ll\gamma}$ Data v. MC



ttH leptonic

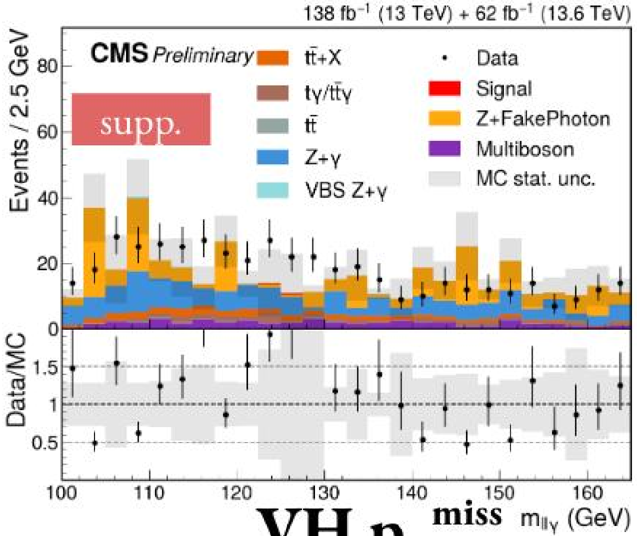


VH $N_{lep} \geq 3$

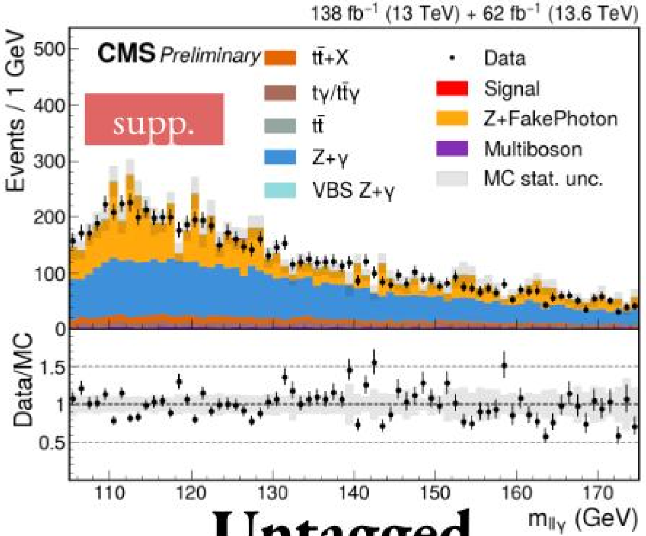


ttH hadronic

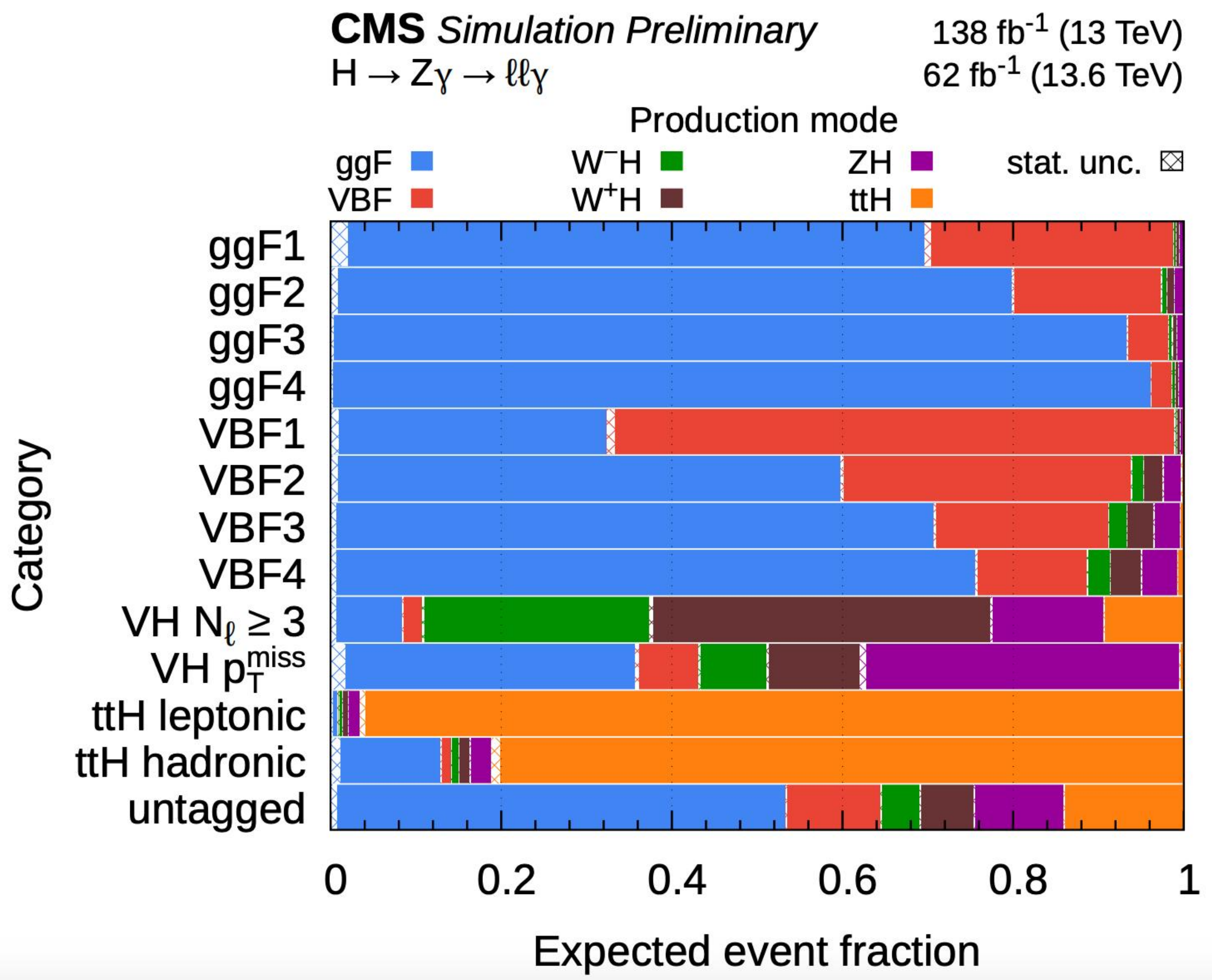
- VH & ttH categories optimized for specific production mode(s)
  - See [AN] for details
- Good data vs. MC agreement

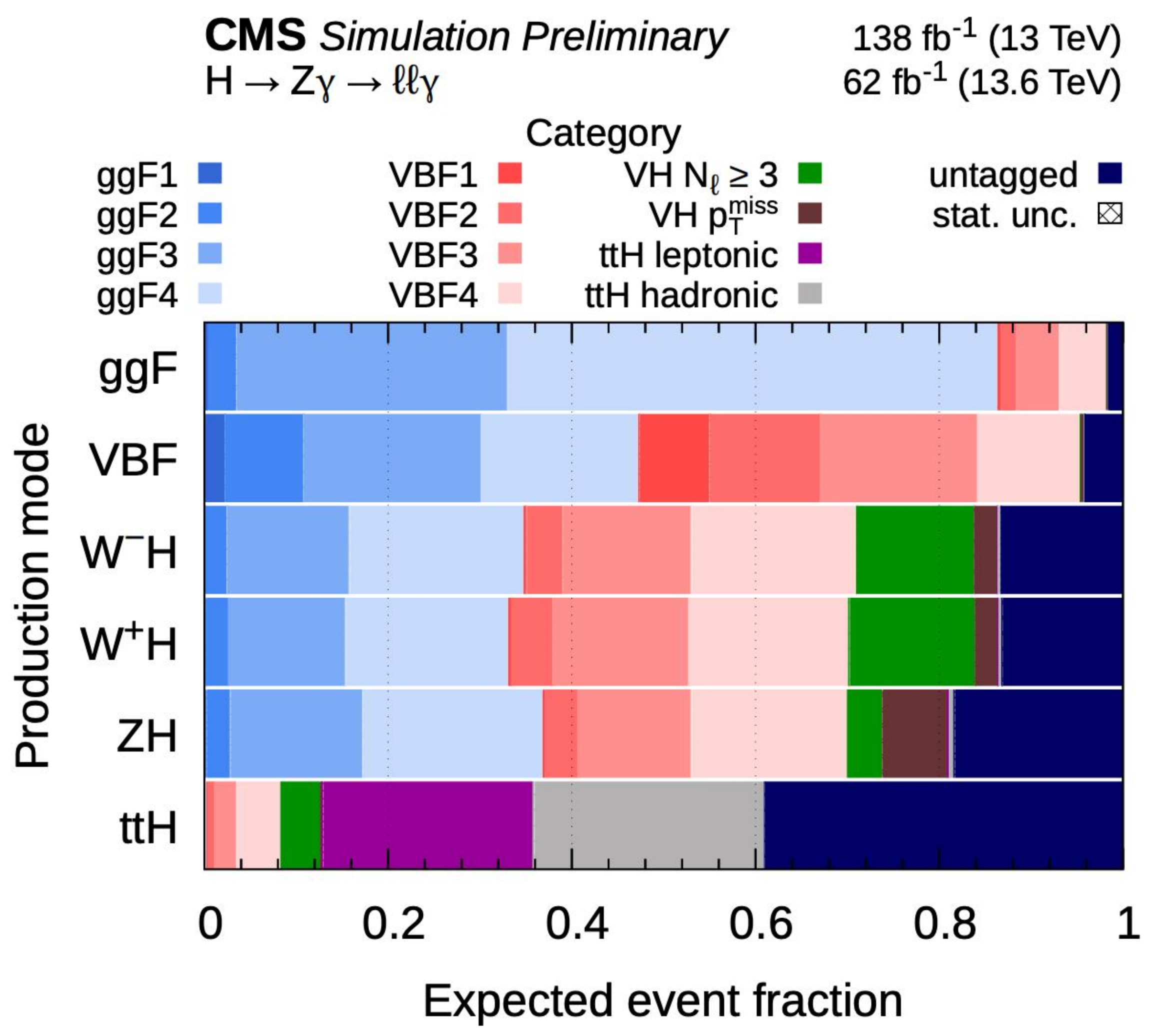


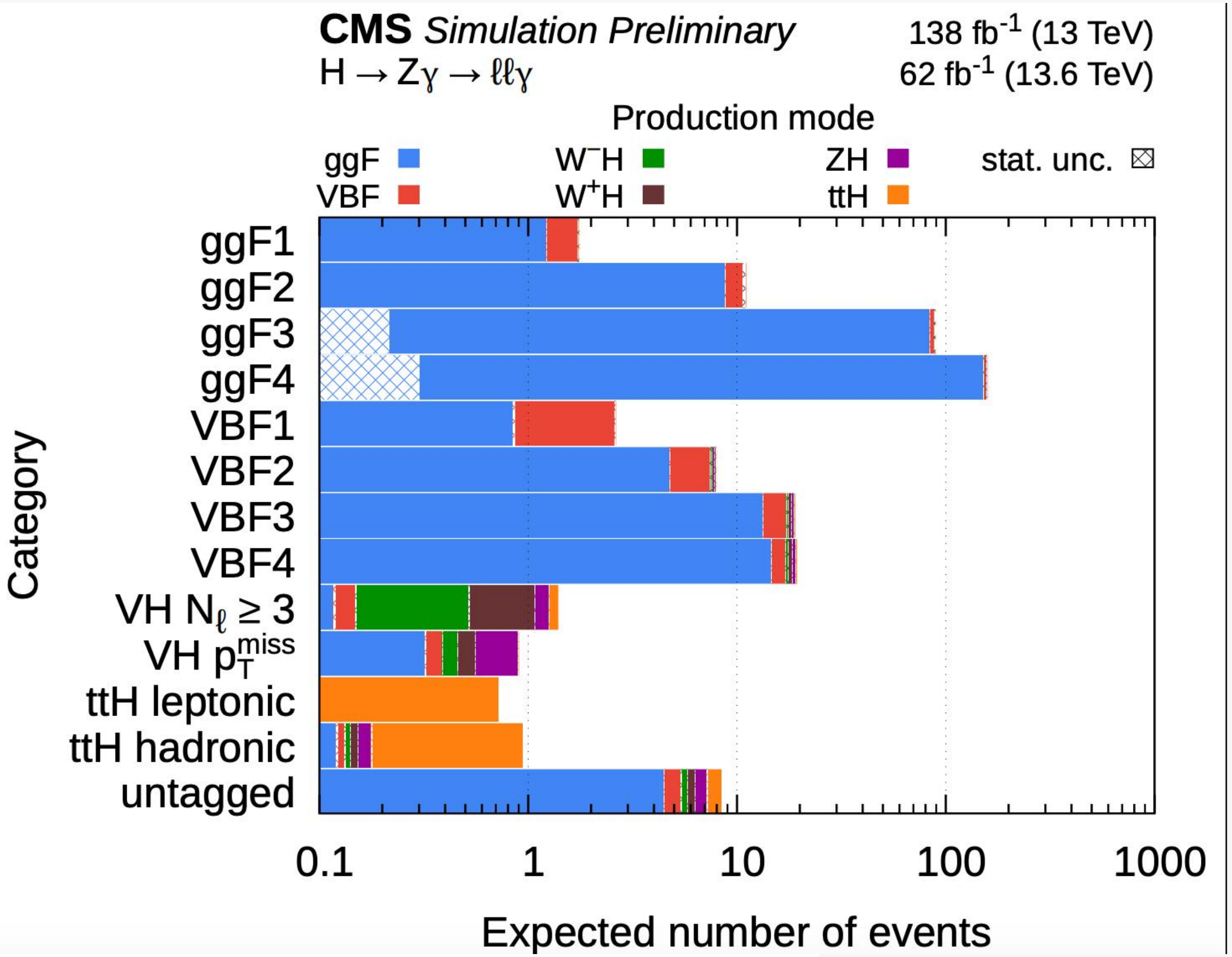
VH p_T^{miss}

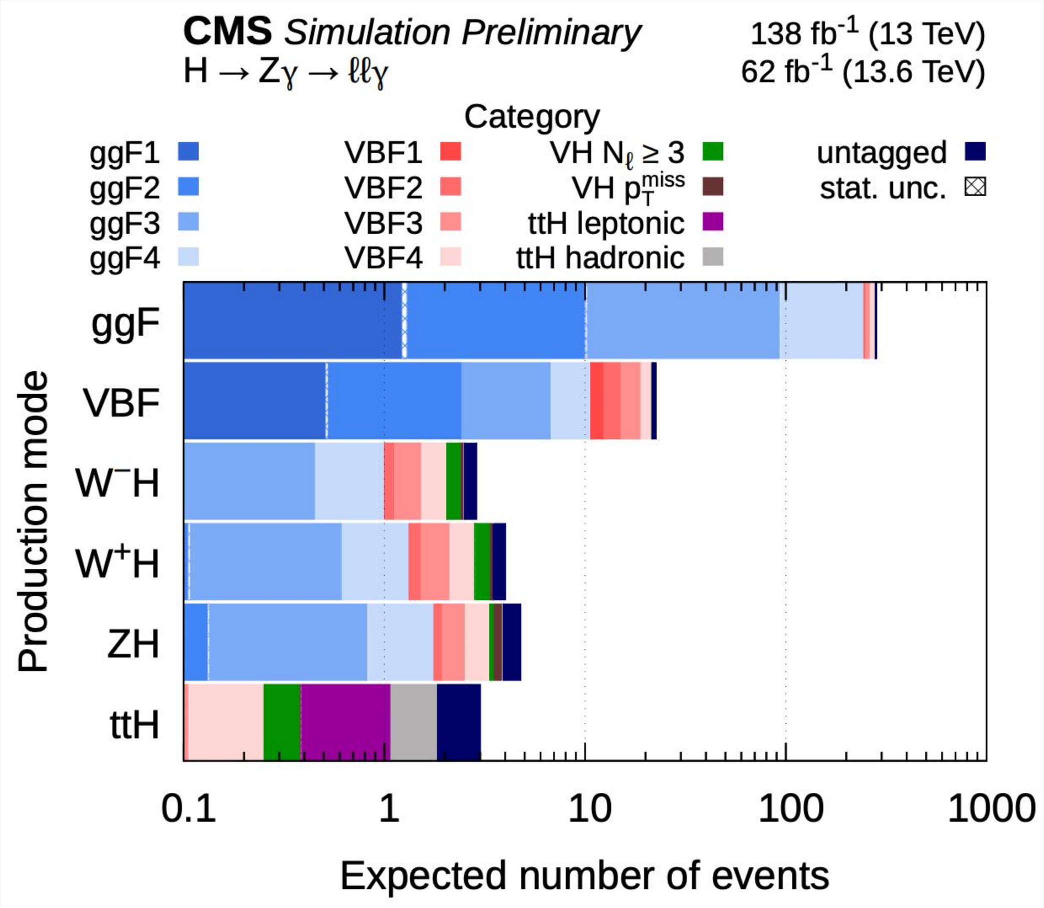


Untagged


CMS Simulation Preliminary
H → Zγ → ℓℓγ
138 fb-1 (13 TeV)
62 fb-1 (13.6 TeV)
Production mode
ggF
VBF
W−H
W+H
ZH
ttH
stat. unc.
Category
ggF1
ggF2
ggF3
ggF4
VBF1
VBF2
VBF3
VBF4
VH Nℓ ≥ 3
VH pTmiss
ttH leptonic
ttH hadronic
untagged
0
0.2
0.4
0.6
0.8
1
Expected event fraction


CMS Simulation Preliminary
H → Zγ → ℓℓγ
138 fb-1 (13 TeV)
62 fb-1 (13.6 TeV)
Category
ggF1
ggF2
ggF3
ggF4
VBF1
VBF2
VBF3
VBF4
VH Nℓ ≥ 3
VH pTmiss
ttH leptonic
ttH hadronic
untagged
stat. unc.
Production mode
ggF
VBF
W−H
W+H
ZH
ttH
0
0.2
0.4
0.6
0.8
1
Expected event fraction


CMS Simulation Preliminary
138 fb-1 (13 TeV)
62 fb-1 (13.6 TeV)
H → Zγ → ℓℓγ
Production mode
ggF
VBF
W⁻H
W⁺H
ZH
ttH
stat. unc.
ggF1
ggF2
ggF3
ggF4
VBF1
VBF2
VBF3
VBF4
VH Nℓ ≥ 3
VH pTmiss
ttH leptonic
ttH hadronic
untagged
Category
0.1
1
10
100
1000
Expected number of events

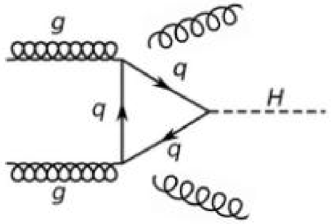

CMS Simulation Preliminary
138 fb-1 (13 TeV)
62 fb-1 (13.6 TeV)
H → Zγ → ℓℓγ
Category
ggF1
ggF2
ggF3
ggF4
VBF1
VBF2
VBF3
VBF4
VH Nℓ ≥ 3
VH pTmiss
ttH leptonic
ttH hadronic
untagged
stat. unc.
ggF
VBF
W⁻H
W⁺H
ZH
ttH
Production mode
0.1
1
10
100
1000
Expected number of events

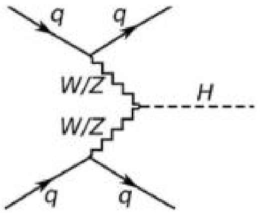
Paper

Category	ggF1	ggF2	ggF3	ggF4	VBF1	VBF2	VBF3	VBF4	VH $N_\ell \geq 3$	VH p_T^{miss}	ttH had.	ttH lep.	untag.
SM Signal Yield													
-ggF	1.18	8.30	78.1	139	0.78	4.43	12.1	12.1	0.10	0.24	0.09	0.00	3.56
-VBF	0.49	1.83	4.07	3.52	1.68	2.58	3.63	2.25	0.03	0.06	0.01	0.00	0.84
-VH	0.02	0.28	1.46	1.93	0.03	0.44	1.48	1.83	0.97	0.47	0.04	0.01	1.49
-ttH	0.00	0.00	0.00	0.00	0.00	0.02	0.07	0.13	0.13	0.00	0.68	0.62	1.02
-Total	1.69	10.4	83.6	144	2.49	7.48	17.3	16.3	1.23	0.77	0.82	0.63	6.91
SM Bkg Yield	12.5	313	10367	68306	7.77	94.5	756	5237	48.9	44.1	19.7	3.02	1541
Data Yield	12	350	9787	64432	9	77	739	5058	77	92	33	7	1654
+ $e^+e^-\gamma$ σ_{eff} (GeV)	1.47	1.59	1.65	1.83	1.45	1.53	1.63	1.80	1.70	1.70	1.56	1.62	1.69
+ $\mu^+\mu^-\gamma$ σ_{eff} (GeV)	1.23	1.26	1.29	1.50	1.29	1.32	1.36	1.50	1.47	1.34	1.34	1.38	1.42
Z_{cc}	0.5	0.6	0.8	0.6	0.9	0.8	0.6	0.2	0.2	0.1	0.2	0.4	0.2
Combined Z_{cc}	1.2				1.3				0.5				
Total Z_{cc}	1.9												

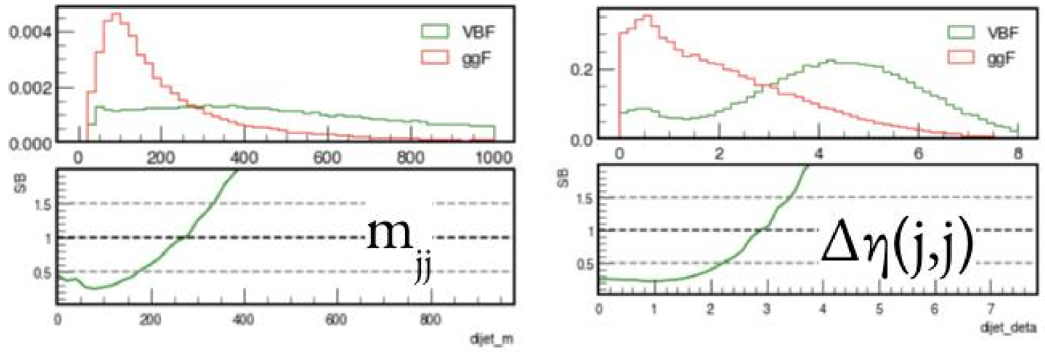
ggF/VBF separation

- ggF and VBF signal have different topologies and 3x difference in number of events with at least two jets. Difficult to find a simple cut to reduce enough ggF signal for a truly VBF-like category without discarding significant VBF signal





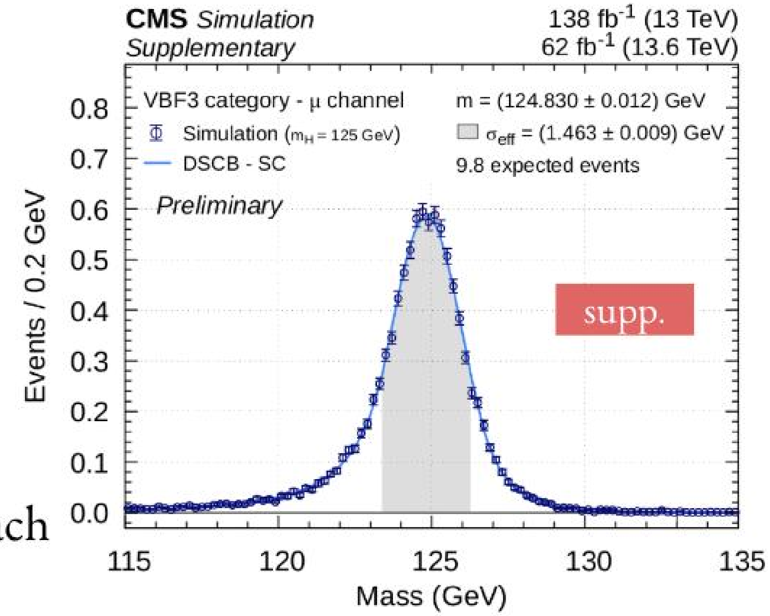
ggF+2j (~39 events w. N_{jet}>=2) VBF (~13 events w. N_{jet}>= 2)

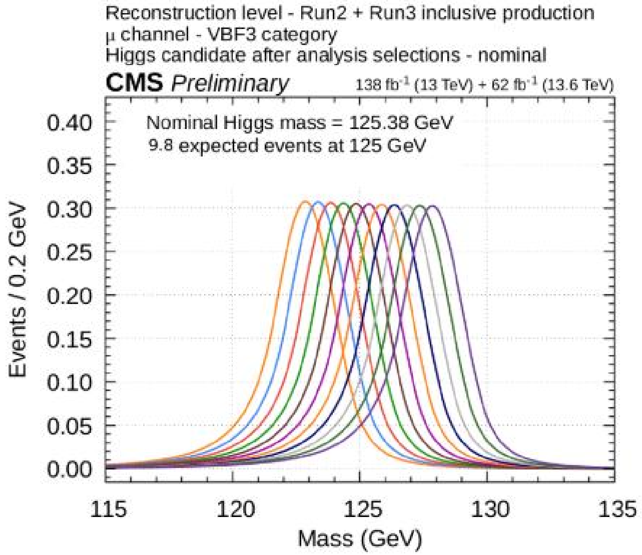


- Explored several strategies for splitting ggF and VBF production modes to balance purity, sensitivity, and simplicity
  - Sort jets differently by dijet kinematics — Worse performance, shapes background to be more VBF-like (backup)
  - Split based on N_j≥2 only — Chosen strategy, good performance, but lower VBF purity
  - Split based on N_j≥2 and m_{jj} — Optimal m_{jj} is 0 (~10% worse performance at m_{jj}>350 GeV)
  - Split using dedicated BDT — Similar performance, higher VBF purity, but more complicated (backup)
  - Do not split — Similar performance, but no dedicated "VBF" category
- Conclusion: stick to a simple cut at Njet >=2. Treat both ggF and VBF as signal in the training
- GGF-targeted category: N_{jet} <= 1; VBF-targeted category: N_{jet} >= 2.

Signal model

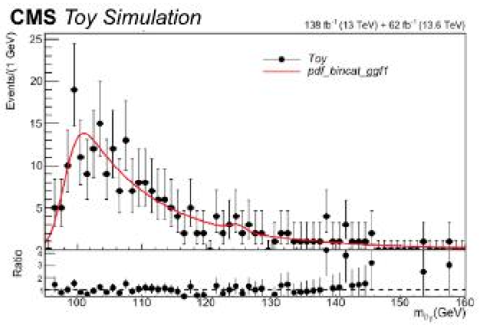
- Model: double-sided Crystal Ball with symmetric core
- Available simulation at 120 GeV, 125 GeV and 130 GeV
- Due to low amount of simulated events:
  - Sum all production modes and campaigns, independent model for each combination of category and lepton flavour
  - Fix tails parameters of 120 GeV and 130 GeV fits to values of 125 GeV
- Parametric model as a function of the mass
  - Linear fit between of 120, 125 and 130 GeV to extrapolate model at 125.38 GeV (backup)
  - In the final fit, all the parameters are fixed or function of mH (fixed)
  - Nuisances on mean and sigma to account for MC scale/resolution uncertainties
- Same model from the 125 GeV simulation for the small resonant background from $H \rightarrow \mu\mu + \gamma_{FSR}$ (backup)



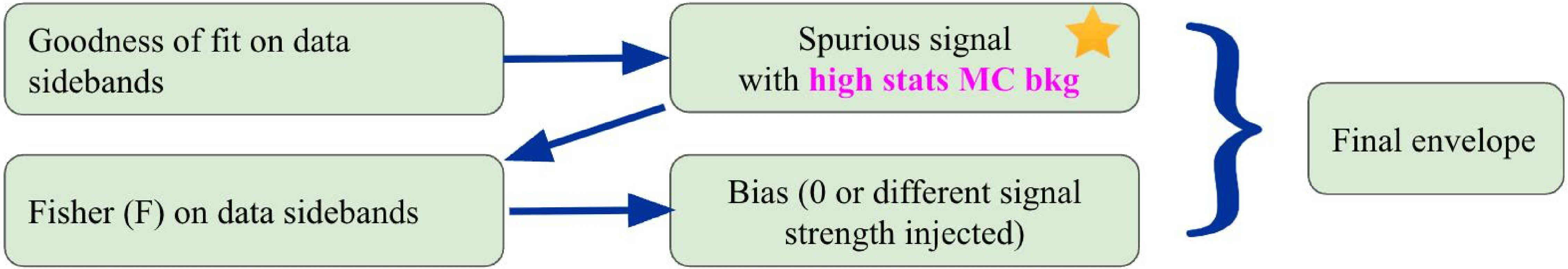


Background model

- Model: "envelope" of functions, discrete profiling method (details)
  - One envelope per category since background turn-on at different locations
- Two kind of function families
  - Convolution models $\mathcal{F}(m_{\ell\ell\gamma};\mu_G,\sigma_G,s,\vec{\alpha}) \equiv \int_{-\infty}^{\infty} \mathcal{N}(m_{\ell\ell\gamma}-t;\mu_G,\sigma_G)\Theta(t;s)f(t;\vec{\alpha})dt,$
    where N is the Gaussian function and f is a function from selected families (Bernstein polynomials, power laws, exponentials and Laurent series)
  - Asymmetric gaussian variants
- Emphasis on the robustness and reliability of the fit, multiple tests to select functions:



Toy data from one category (ggF1) showing what S+B fit might look like



Spurious Signal Test ⭐

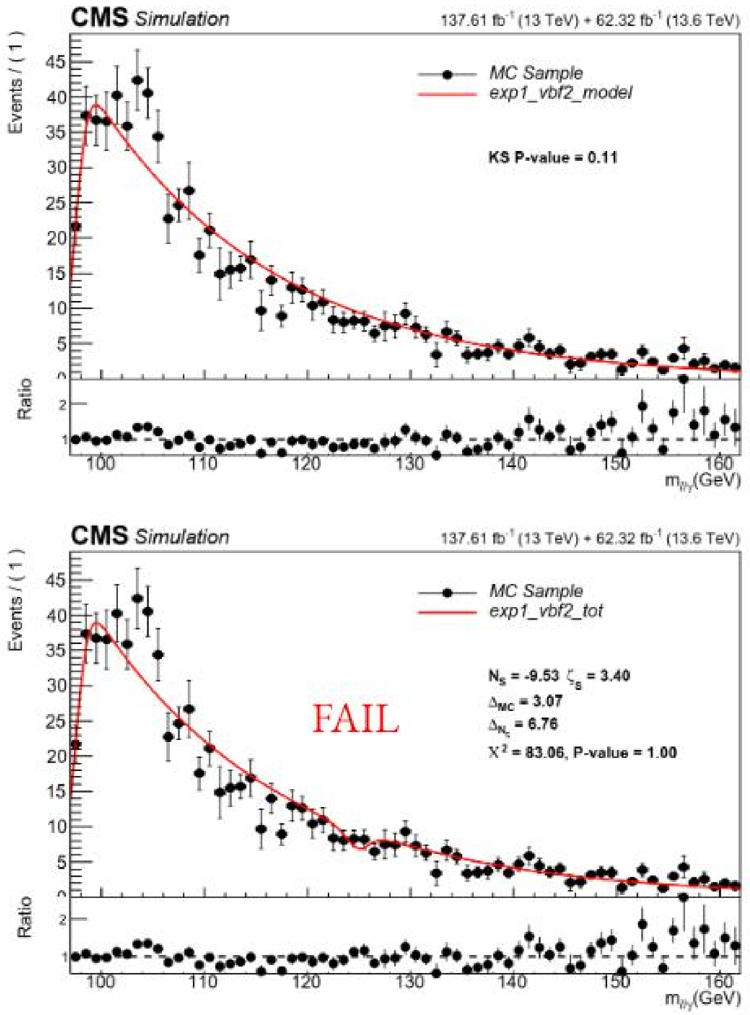
- Use the **background-only high statistics simulation** to test whether the S+B fit with a given background model would bias the signal measurement, giving “spurious signal”

- The spurious signal is defined as $|\zeta_s| = \max(0, |N_s| - 2\Delta_{MC})$
  - N_s is the signal yield from the fit
  - Δ^s_{MC} is the error of the signal yield representing the statistical uncertainty of the MC sample
  - ΔN_s is also the error of the signal yield but representing the expected uncertainty of the data

- To pass the test, on top of a good fit (χ^2 p-value > 0.05),

$$|\zeta_s| < 20\% \times \Delta N_s \quad \text{or} \quad |\zeta_s| < 10\% \times N_{\text{exp signal}}$$

(example plots in the backup)



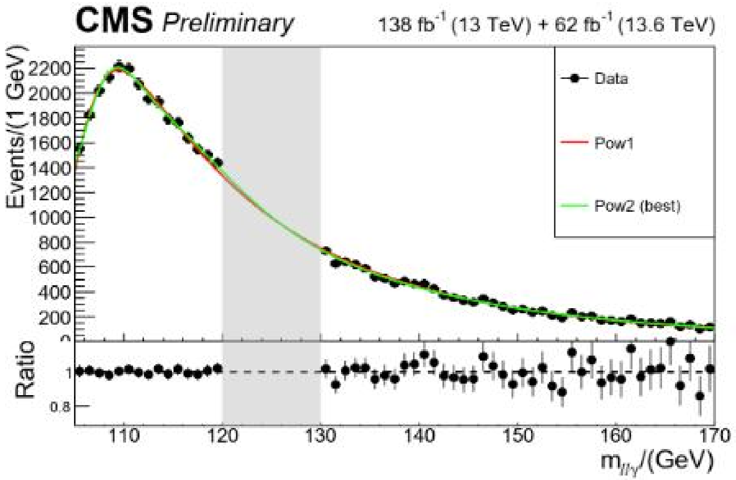
B-only fit and S+B fit of Exp1 bkg model over the high-stat MC of VBF2. Exp1 fails the test with $\zeta = 3.40$

Fisher (F) Test - Degrees of Freedom Optimization

- F test is used to determine how many degrees of freedom to use to minimize overfitting
- Data sideband NLL fit of the background functions from the same family, use the difference between NLL to select functions
- Quantity F follows the χ^2(k = Δn) distribution where Δn is the difference of DOF of two models in the same family. P-value of the test is:

$$p = P(F > 2(NLL(n) - NLL(n + \Delta n))|\chi^2(\Delta n))$$

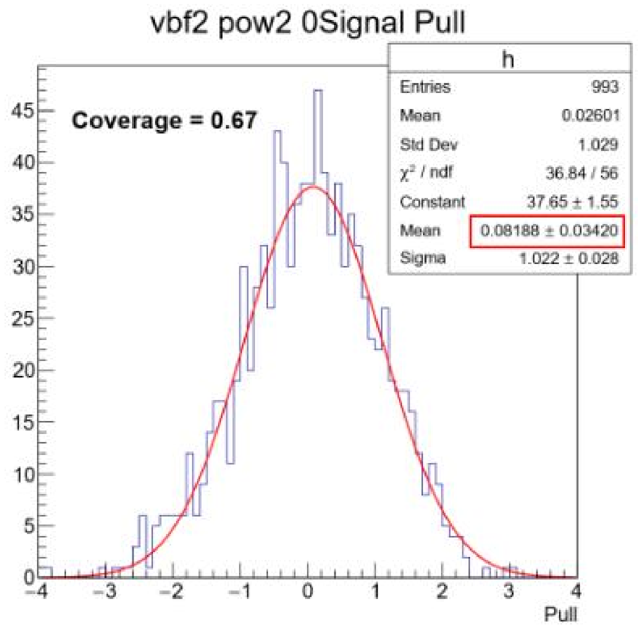
- If p-value < 0.05, the model with the higher DOF is preferred, otherwise only the lower DOF model is kept in the envelope.



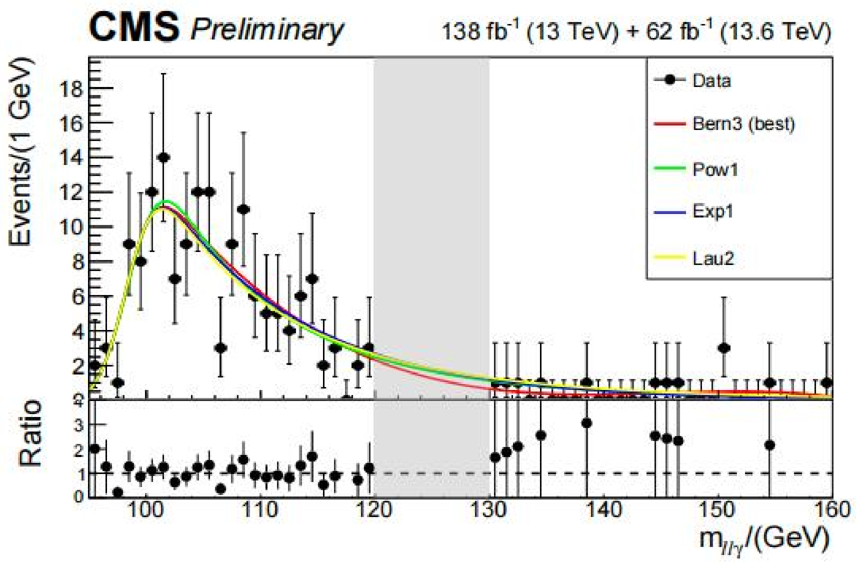
Fisher test in the Power series family in ggF3 category. Pow2 has noticeable improvement over Pow1, thus both Pow1 and Pow2 are kept in the envelope.

Profile Bias and Signal Injection Tests

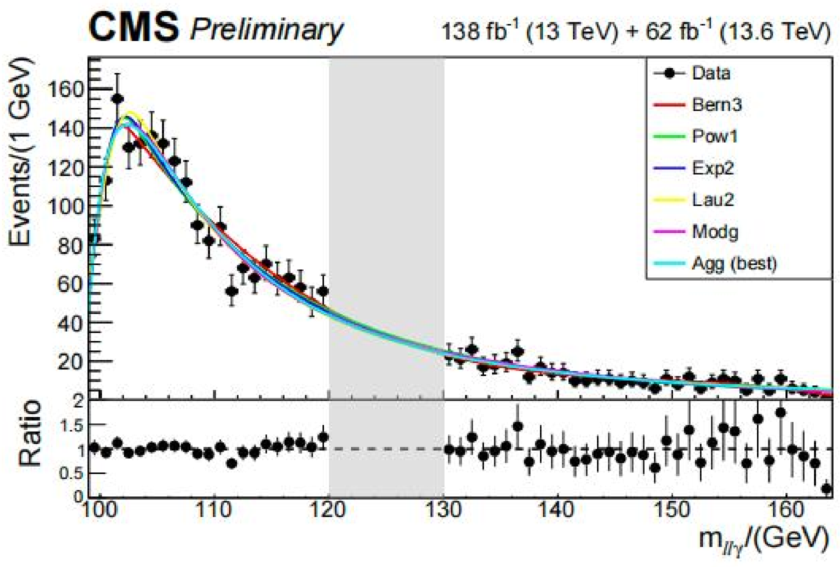
- In the **profile bias test**, we generate 1000 background-only toys for each model in the envelope and estimate the envelope bias by performing a S+B fits on the toys
  - Toys fit with **the envelope**
  - The pull = $(N_{\text{sig fit}} - N_{\text{sig exp}})/\sigma_{\text{sig fit}}$ should form a Gaussian whose mean is an estimate of the bias for the envelope
- To pass the test, all the members of the envelope should have bias smaller than 0.2
  - In cases with high bias, a heat map is used to determine the functions responsible for the biased fits
- **Signal injection test**, in principle, is the same test as the profile bias test, but where the toys are generating assuming 1x or 5x the nominal signal



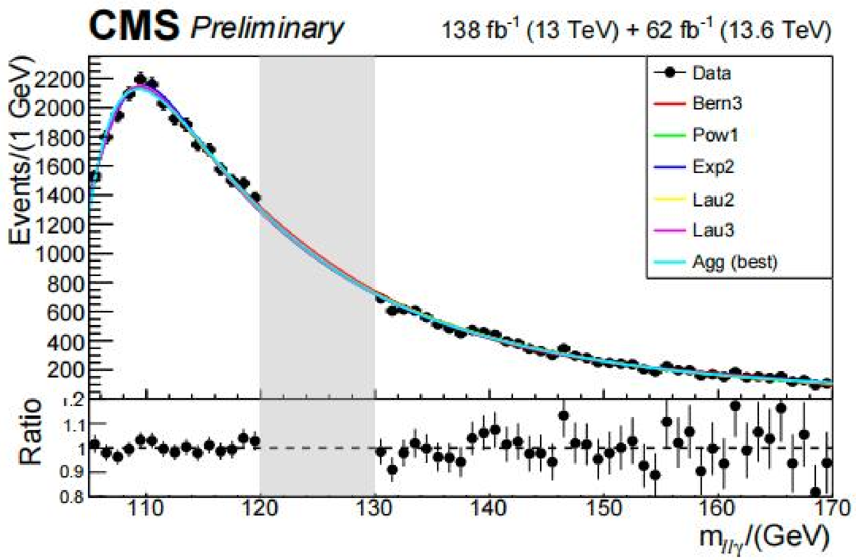
Pull distribution of Pow2 toys in VBF2.
Bias is 0.08



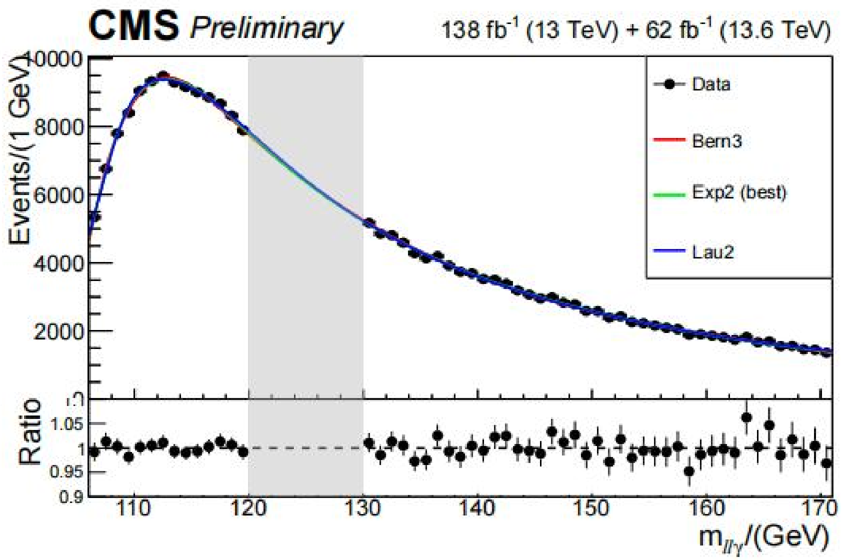
(a) ggF1 data sideband and final envelope



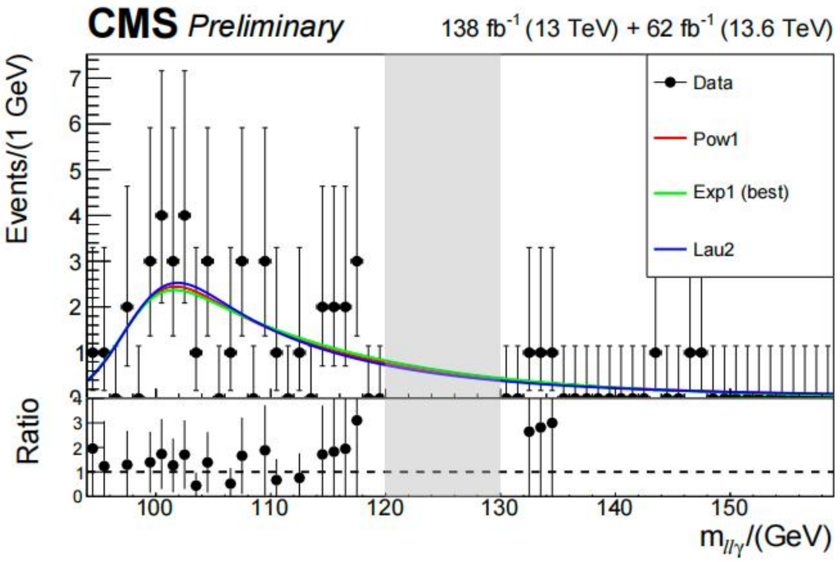
(b) ggF2 data sideband and final envelope



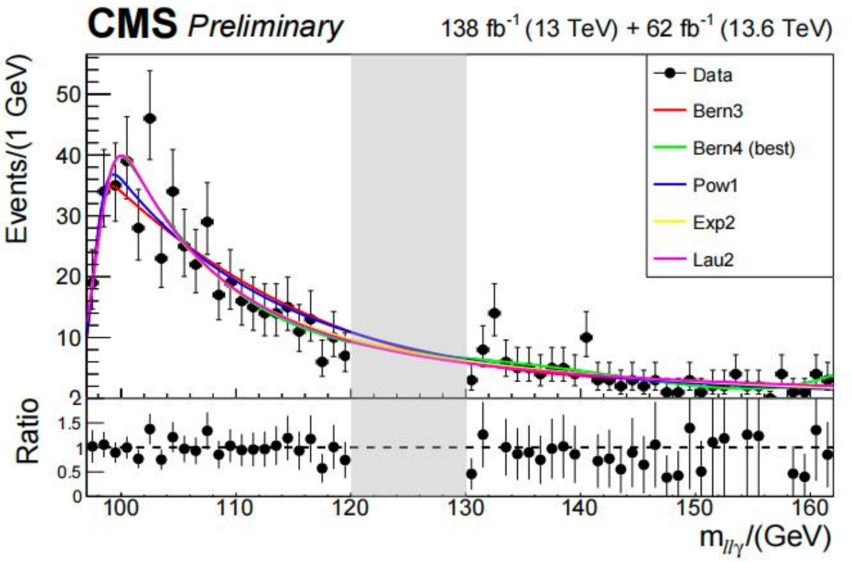
(c) ggF3 data sideband and final envelope



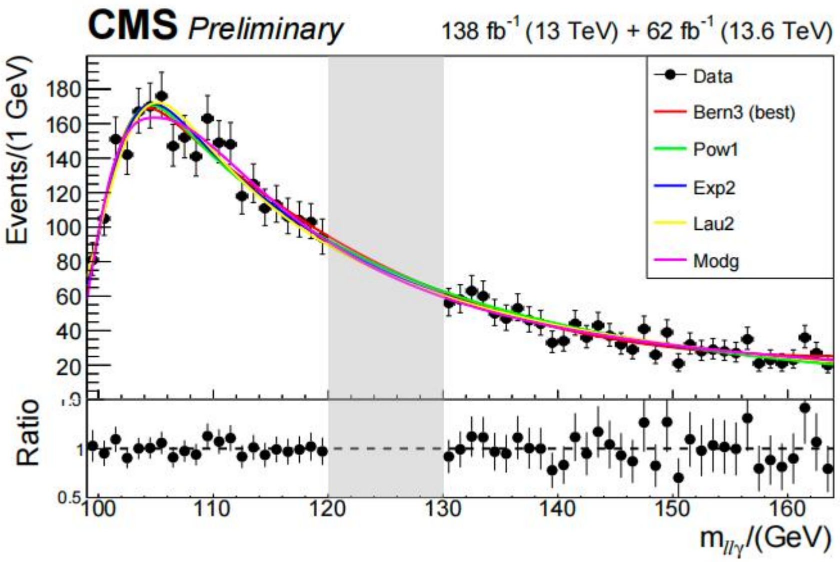
(d) ggF4 data sideband and final envelope



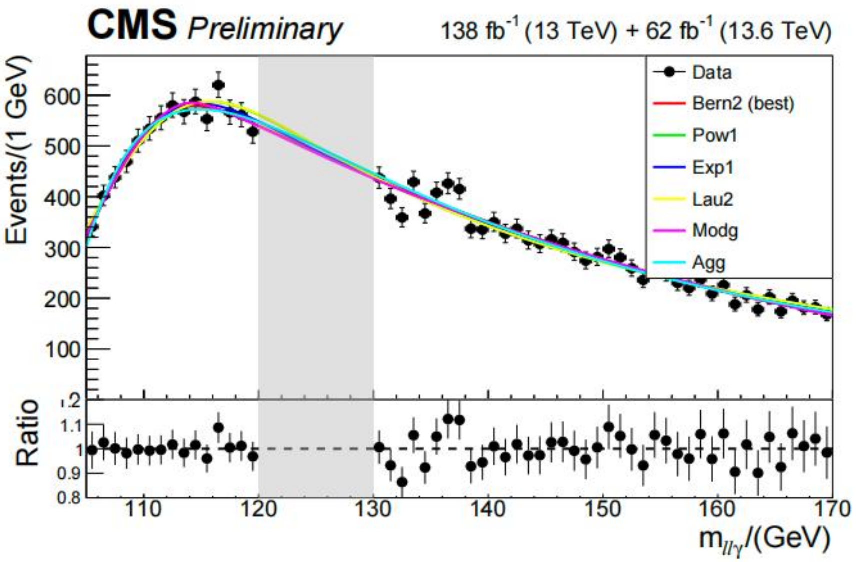
(a) VBF1 data sideband and final envelope



(b) VBF2 data sideband and final envelope

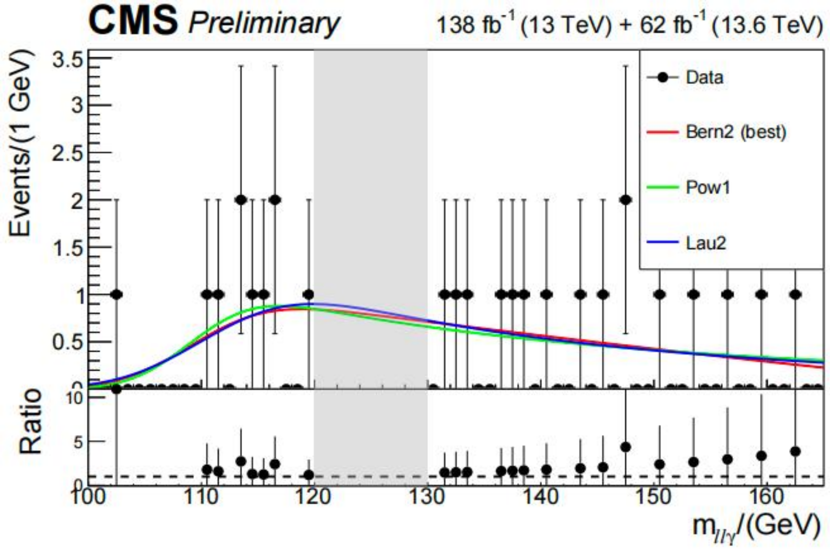


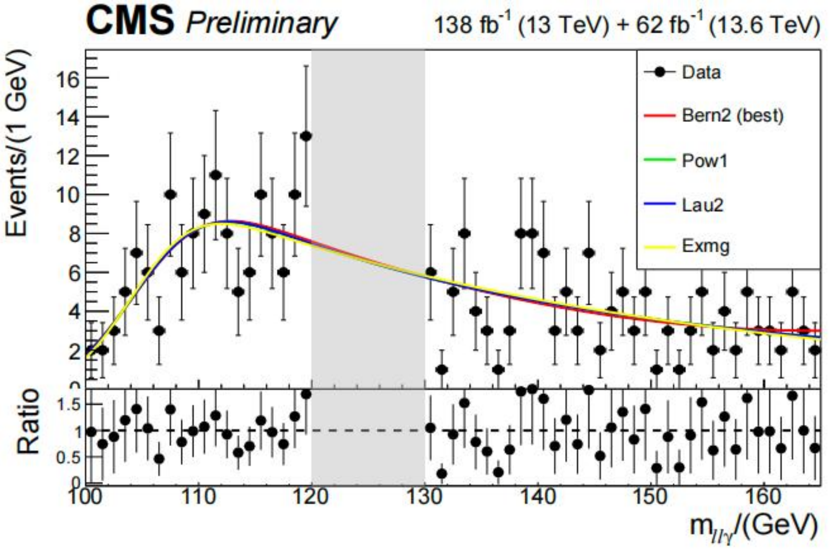
(c) VBF3 data sideband and final envelope

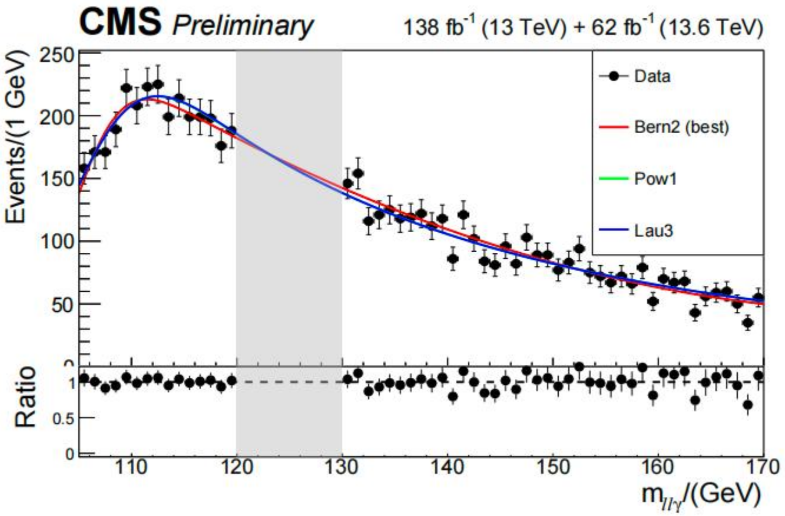


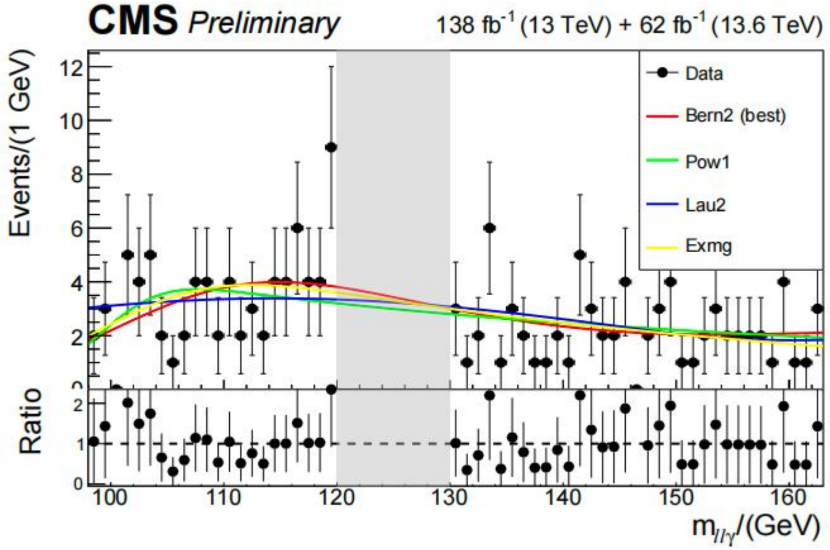
(d) VBF4 data sideband and final envelope

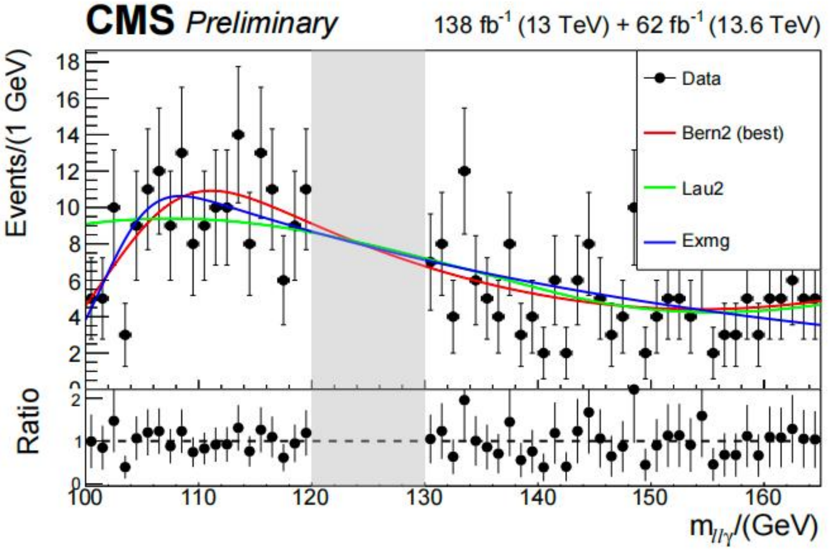
北京大學

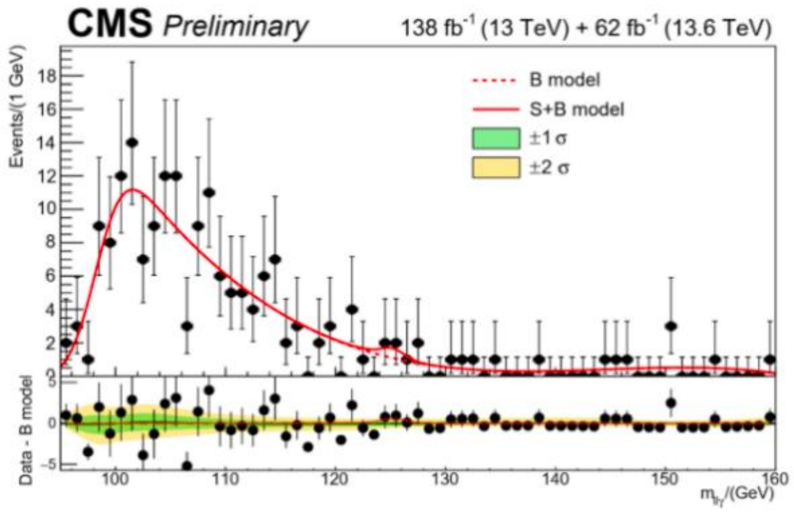

CMS Preliminary
138 fb^{-1} (13 TeV) + 62 fb^{-1} (13.6 TeV)
Data
Bern2 (best)
Pow1
Lau2
Events/(1 GeV)
Ratio
$m_{ll\gamma}$/(GeV)


CMS Preliminary
138 fb^{-1} (13 TeV) + 62 fb^{-1} (13.6 TeV)
Data
Bern2 (best)
Pow1
Lau2
Exmg
Events/(1 GeV)
Ratio
$m_{ll\gamma}$/(GeV)

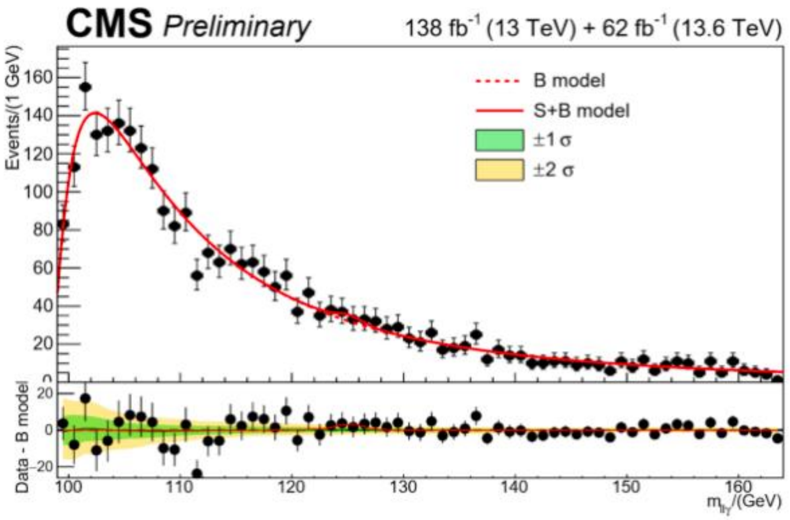

CMS Preliminary
138 fb^{-1} (13 TeV) + 62 fb^{-1} (13.6 TeV)
Data
Bern2 (best)
Pow1
Lau3
Events/(1 GeV)
Ratio
$m_{ll\gamma}$/(GeV)


CMS Preliminary
138 fb^{-1} (13 TeV) + 62 fb^{-1} (13.6 TeV)
Data
Bern2 (best)
Pow1
Lau2
Exmg
Events/(1 GeV)
Ratio
$m_{ll\gamma}$/(GeV)

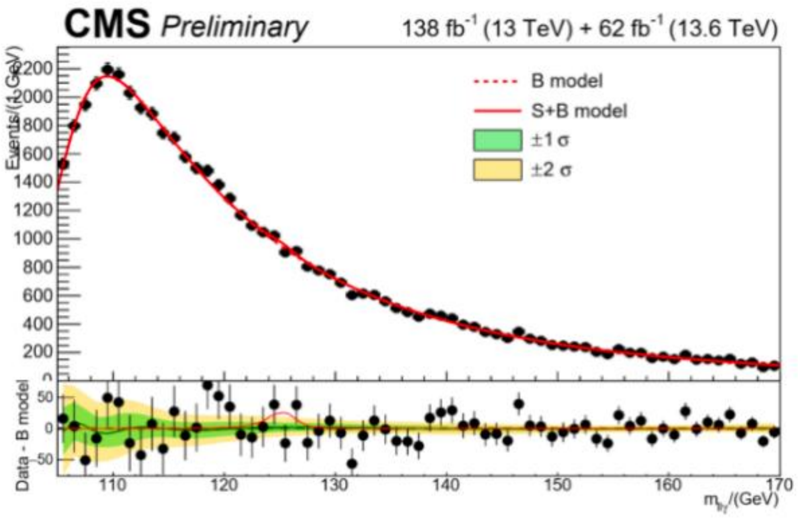

CMS Preliminary
138 fb^{-1} (13 TeV) + 62 fb^{-1} (13.6 TeV)
Data
Bern2 (best)
Lau2
Exmg
Events/(1 GeV)
Ratio
$m_{ll\gamma}$/(GeV)



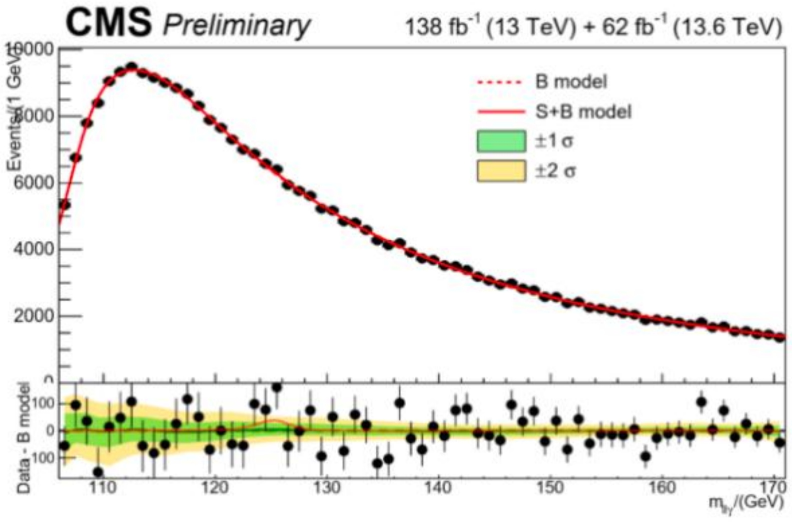
(a) ggF1



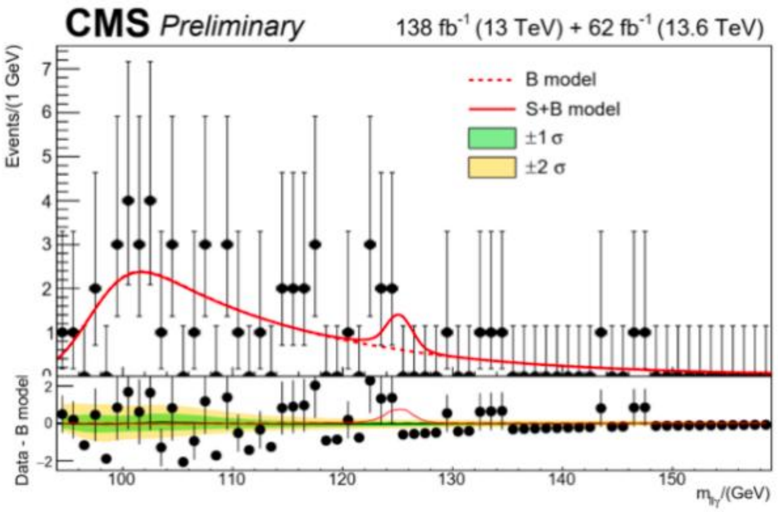
(b) ggF2



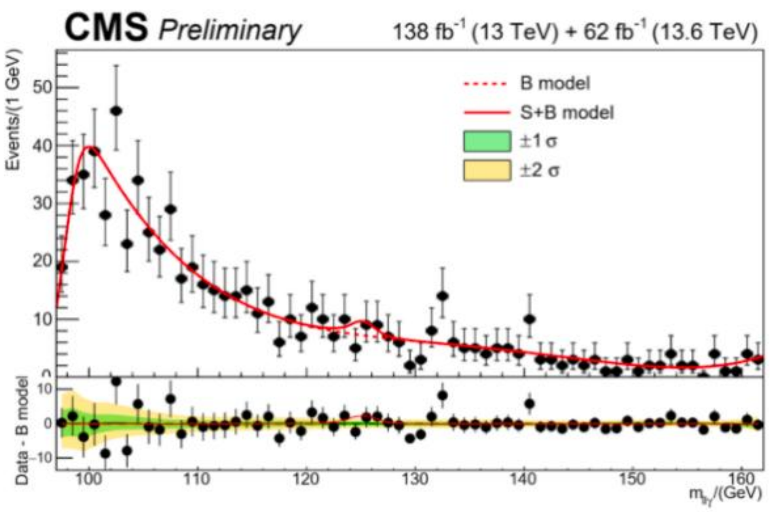
(c) ggF3



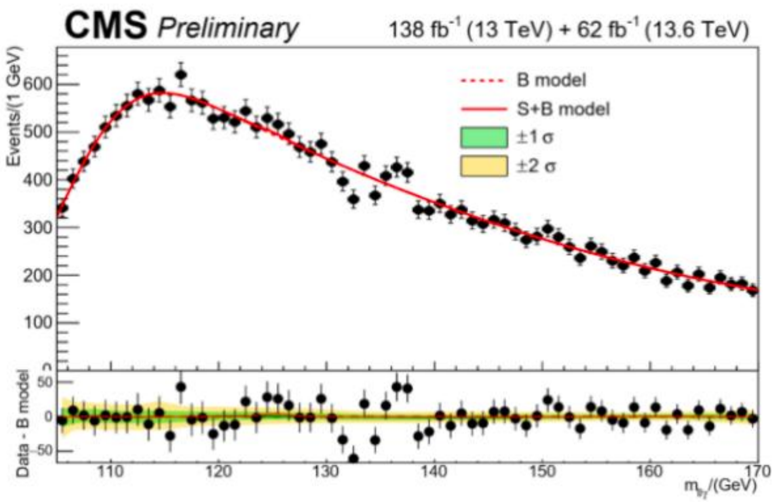
(d) ggF4



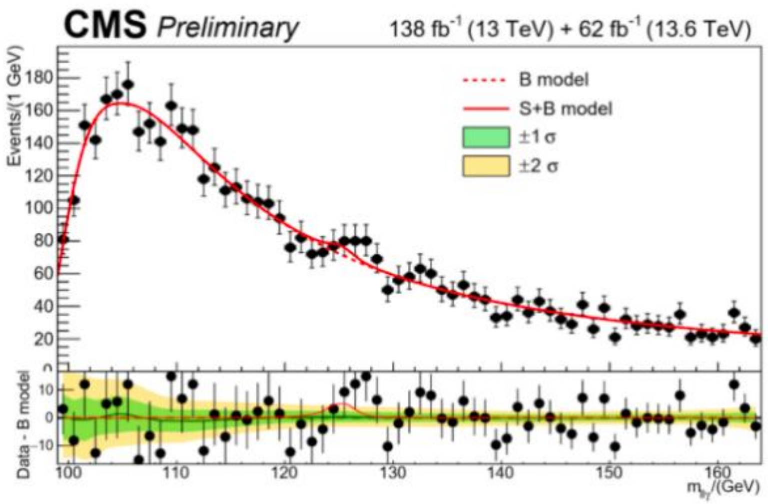
(e) VBF1



(f) VBF2

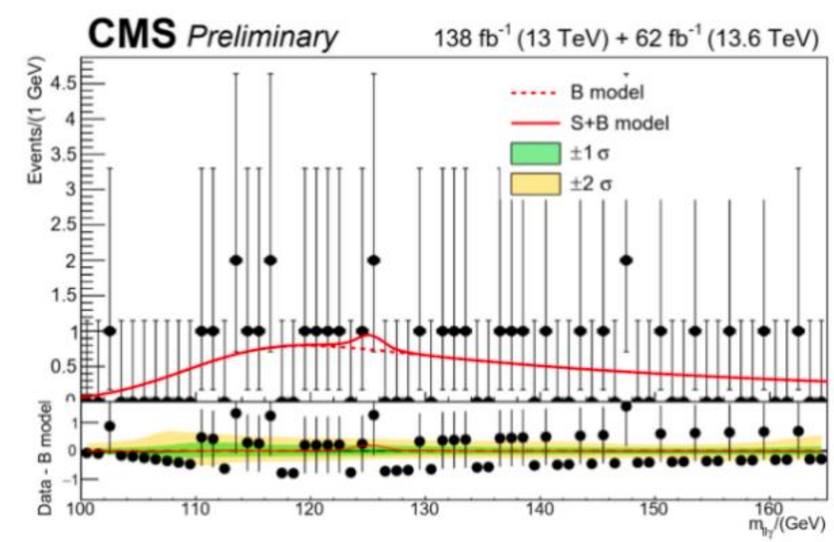


(g) VBF3

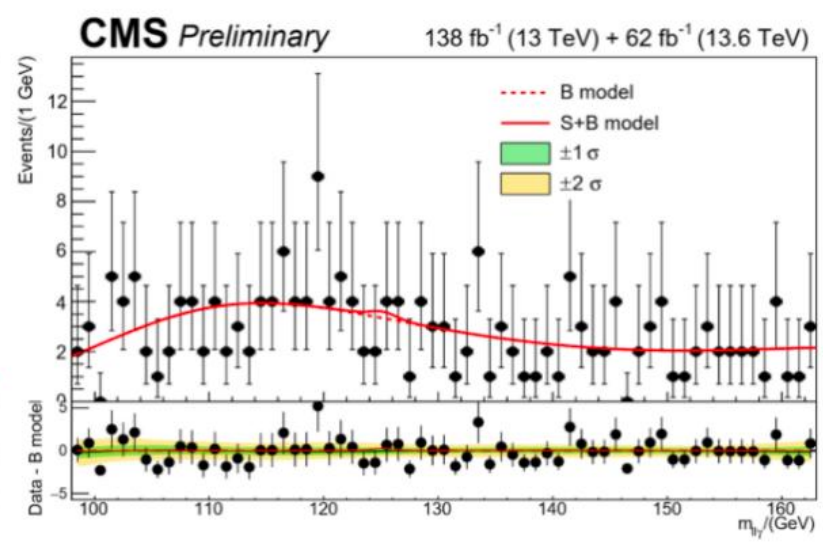


(h) VBF4

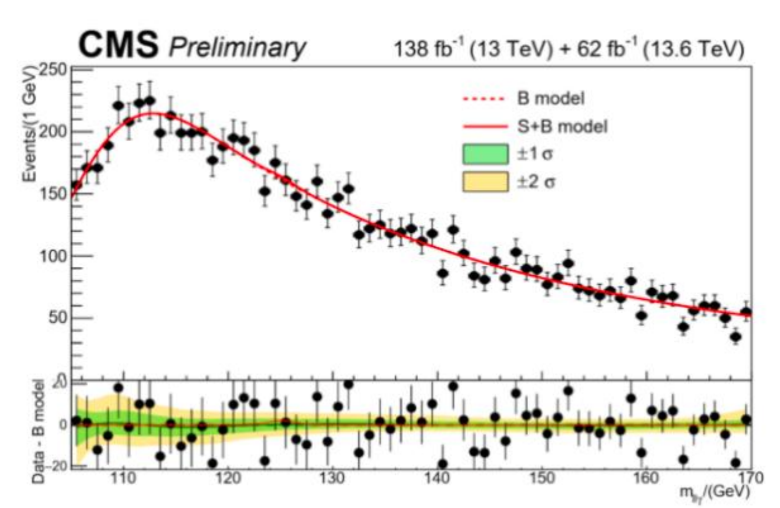
北京大學



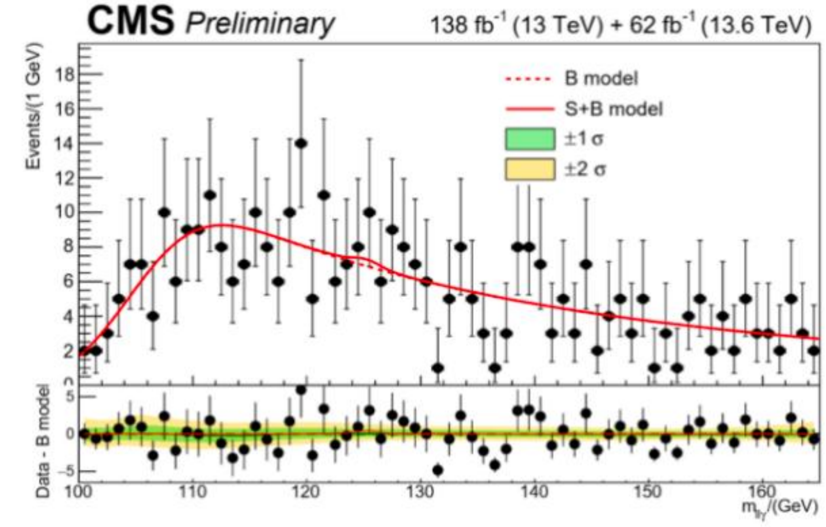
(j) tthlep



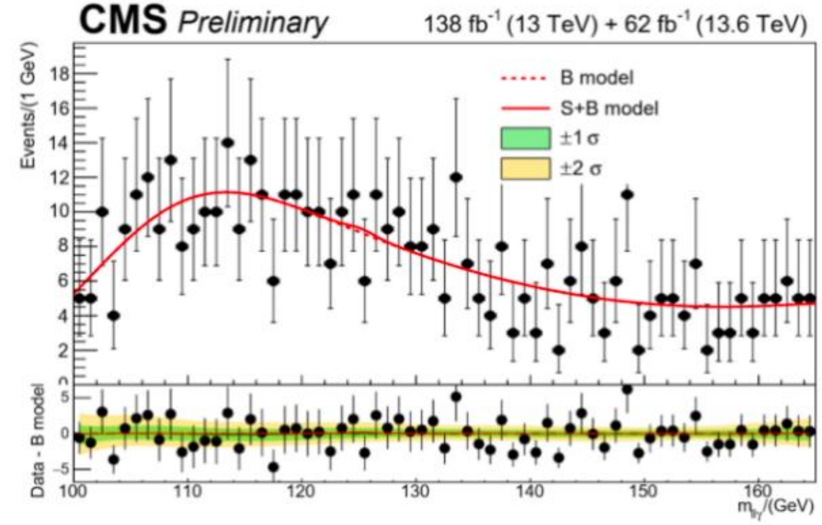
(i) tthhad



(m) Untagged

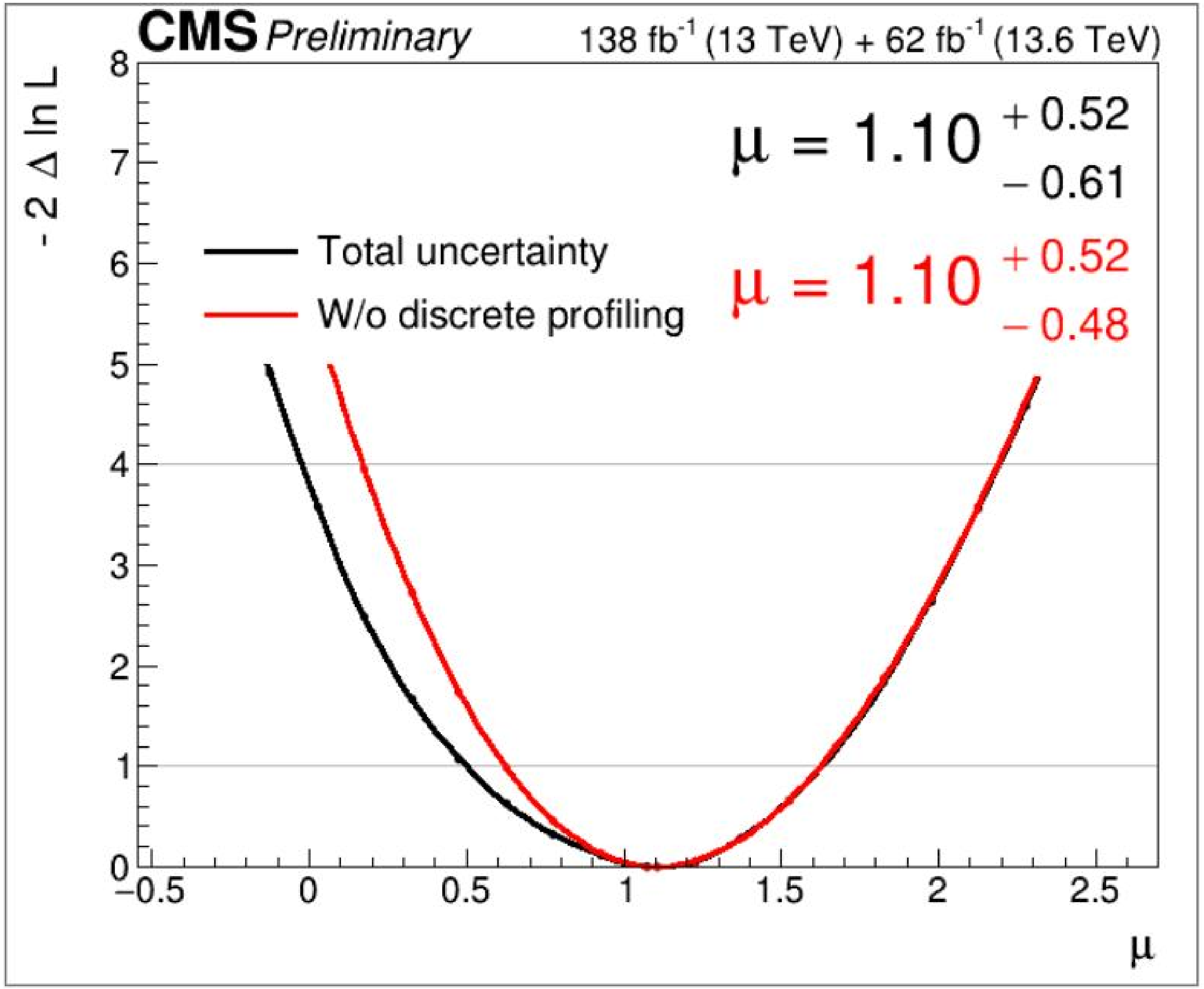


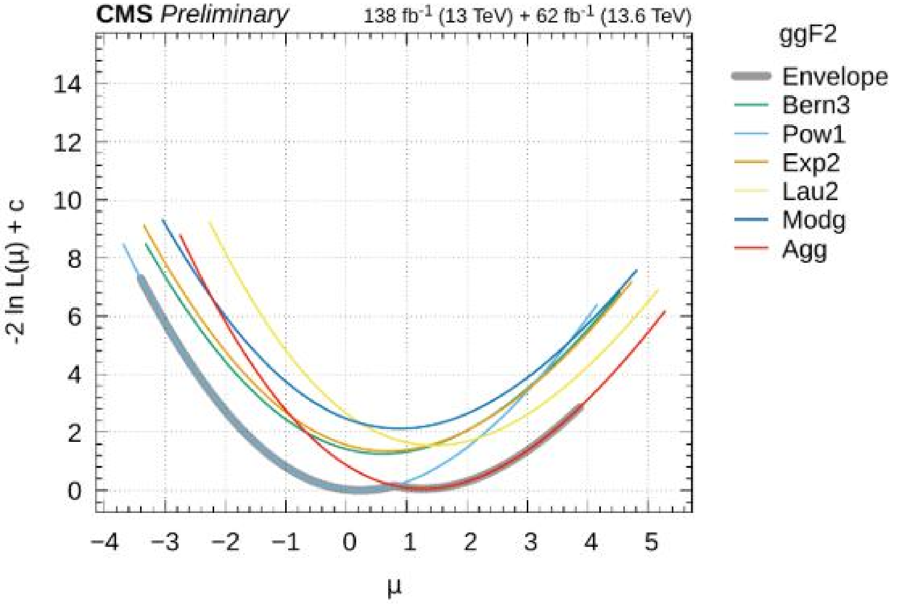
(l) VH3l

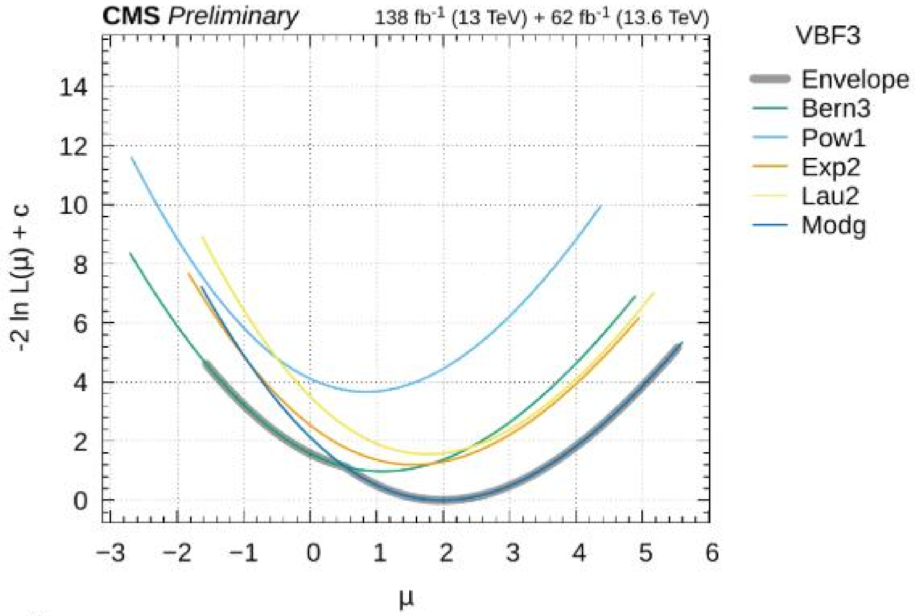


(k) VHpTmiss

Impact of discrete profiling







不确定度来源	对信号强度的影响
统计误差	+0.50, -0.60
本底形状	+0.08, -0.36
信号系统误差	0.11
固定阶数计算截断	0.09
轻子和光子能量刻度和分辨率	0.04
光子品质	0.04
部分子分布函数、强耦合系数和夸克质量	0.03
轻子和光子重建、鉴别和触发效率	0.02
亮度和堆积效应	0.01
Underlying event 和部分子 shower	0.01
喷注和丢失横动量的能量刻度、b 喷注标记、模拟样本统计量	< 0.01

北京大學

PEKING UNIVERSITY 1898