Search for h→aa→2μ2b/2τ2b with the CMS experiment

Pallabi Das
Princeton University, USA

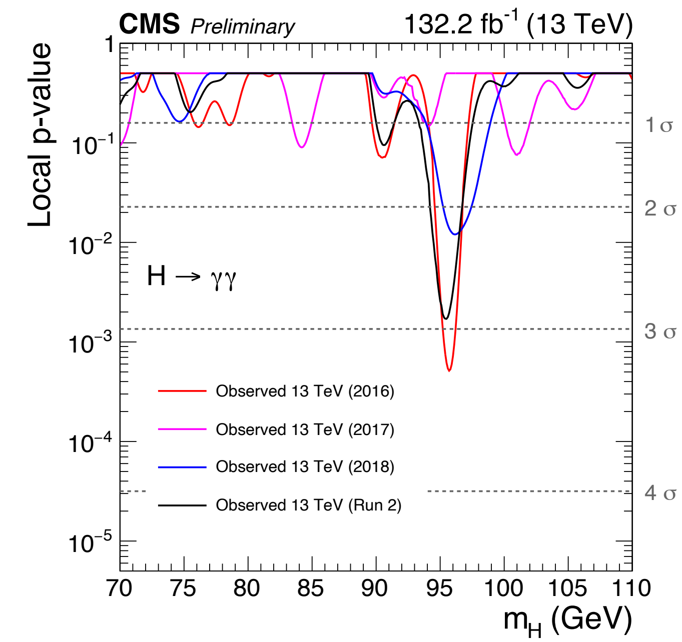
The 20th Workshop of the LHC Higgs Working Group
CERN, Geneva
15th November, 2023

Introduction

The discovery of the Higgs boson 10 years ago established the theory of the SM

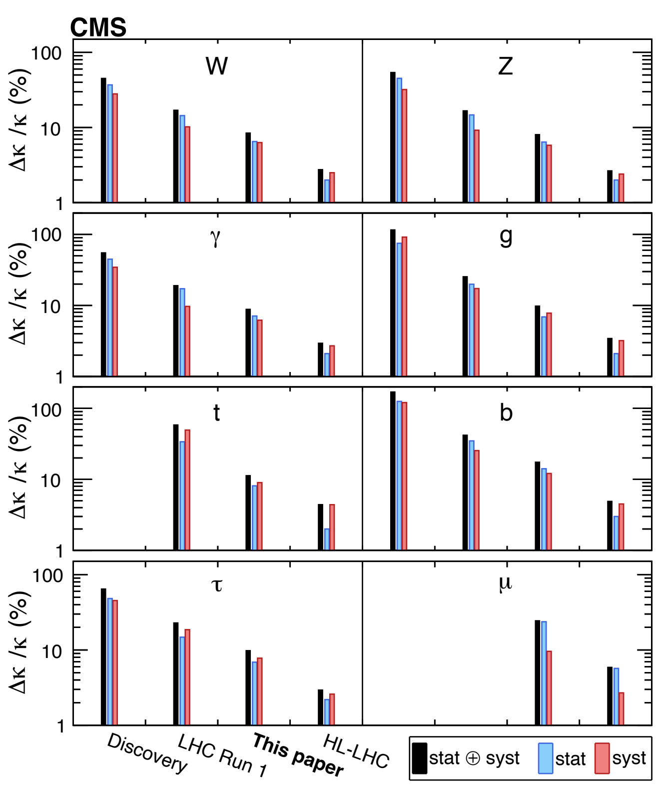
→ ***But many questions remain!***

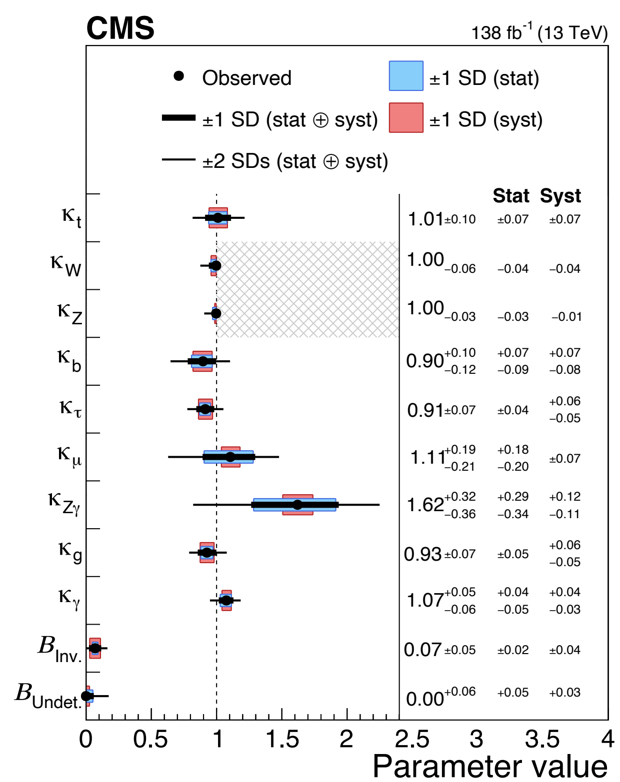
- Extended Higgs sector (MSSM, NMSSM etc.) theories can explain **Dark Matter** origin, **Hierarchy Problem**, etc. and also predict a **Higgs Resonance**
- Example: diphoton channel excess at 96 GeV can be accommodated in 2HDM, a minimal extension of SM with an extra Higgs doublet



CMS-PAS-HIG-20-002

Nature 607, 60–68 (2022)



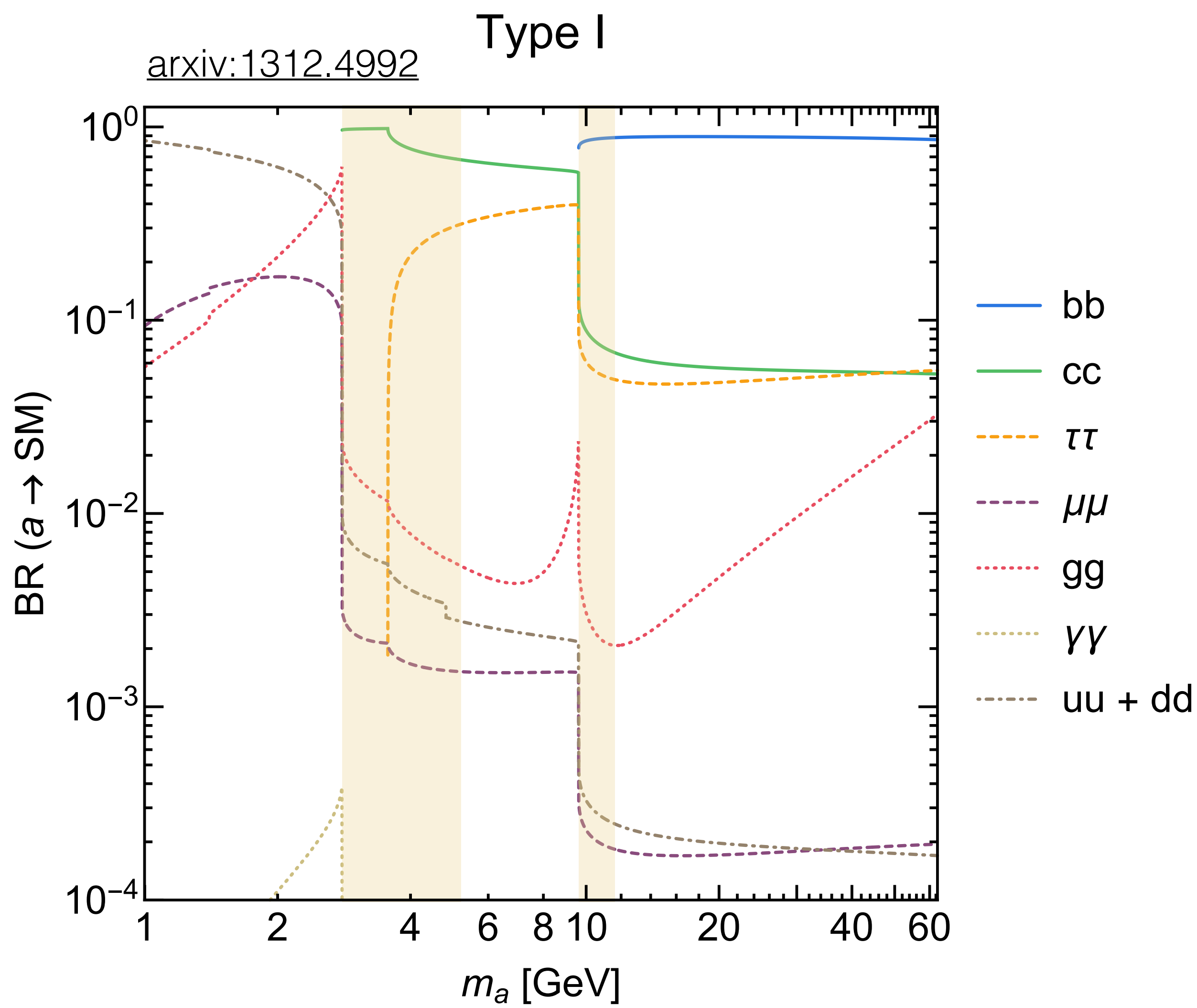


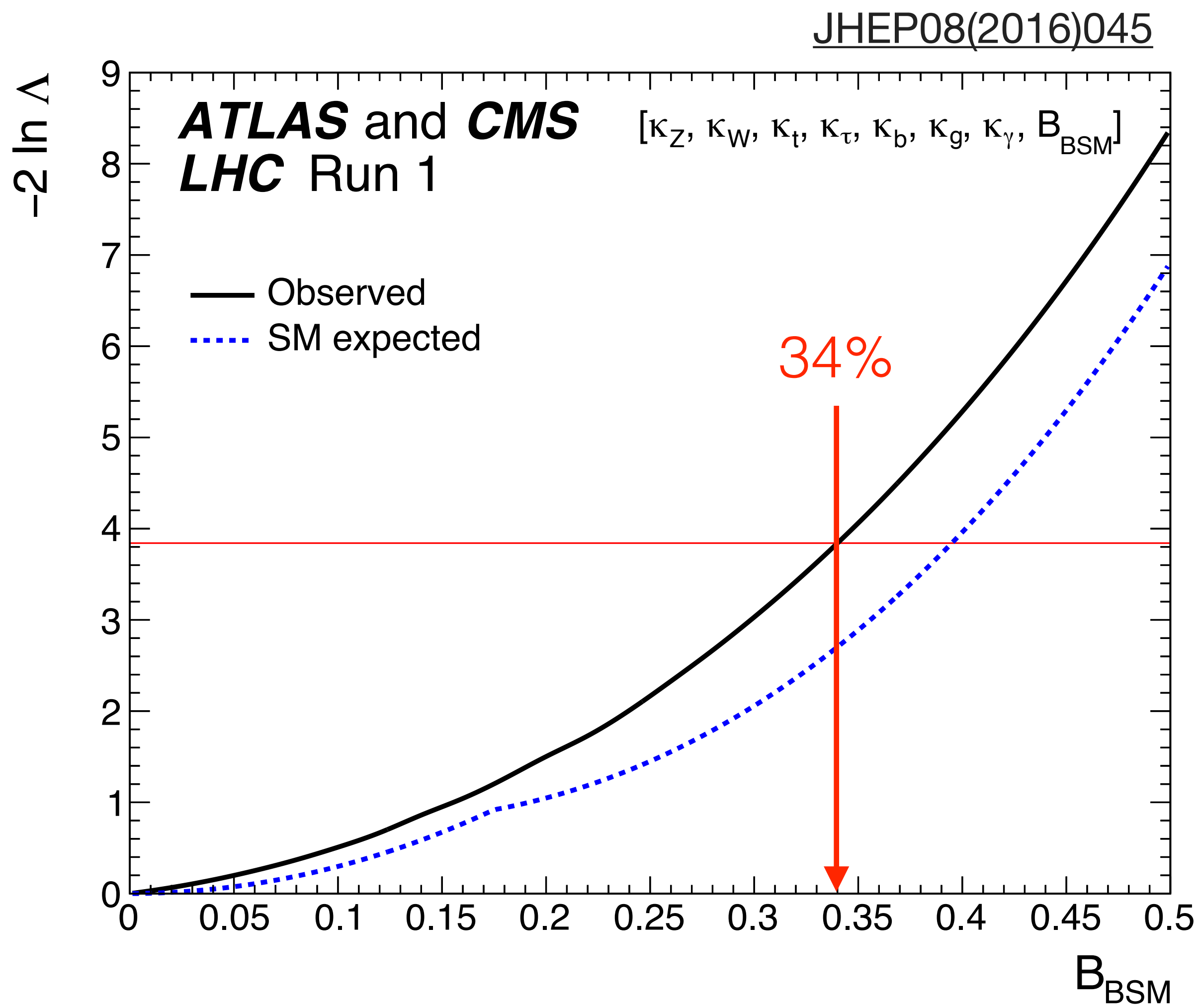
- Limited sensitivity to new physics interactions through SM Higgs coupling measurements
- Direct search for exotic particles able to probe several TeV energy scales

This talk: reviewing full Run-2 results of H→aa→2μ2b/2τ2b search from CMS-PAS-HIG-22-007

Higgs decays to pseudoscalars

- 2HDM+S theory provides wide range of possible exotic Higgs decays, while much of the parameter space of 2HDM is constrained by LHC experiments
- The additional singlet has no direct Yukawa coupling, only couples to the two Higgs fields
- Small mixing with Higgs field: H→aa→SM particles, where a is the pseudoscalar mass eigenstate mostly composed of the imaginary part of the singlet





SM compatibility: combining all Run-1 ATLAS and CMS measurements an upper limit of 34% is set on BSM Higgs decays→ loose constraint on BSM physics

2HDM+S couplings and BR

- Four types of 2HDM+S based on coupling structure of the two Higgs doublets and the SM fermions

	Type I	Type II	Type III	Type IV
Charged leptons	Φ_2	Φ_1	Φ_1	Φ_2
Up-type quarks	Φ_2	Φ_2	Φ_2	Φ_2
Down-type quarks	Φ_2	Φ_1	Φ_2	Φ_1

- Coupling of the pseudoscalar to the fermions depends on tanβ and the mixing angle between the S and Higgs doublets, θ
- Coupling ratio $\xi_f \propto \sin\theta$, thus BR(a→fermions) is independent of θ and has the following behaviour:
  - Type I: no tanβ dependence
  - Type II: decays to down quark and leptons suppressed (enhanced) for tanβ < 1 (tanβ > 1)
  - Type III: tanβ > 1 enhances all BR(a→leptons)
  - Type IV: decays to up quark and leptons suppressed (enhanced) for tanβ > 1 (tanβ < 1)

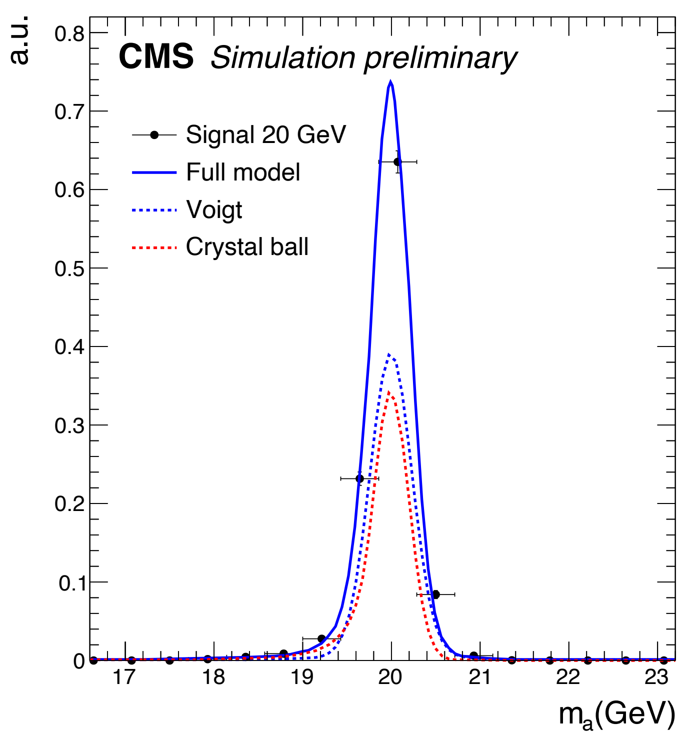
Highest production rate of H→aa→2μ2b/2τ2b is predicted by the Type III model

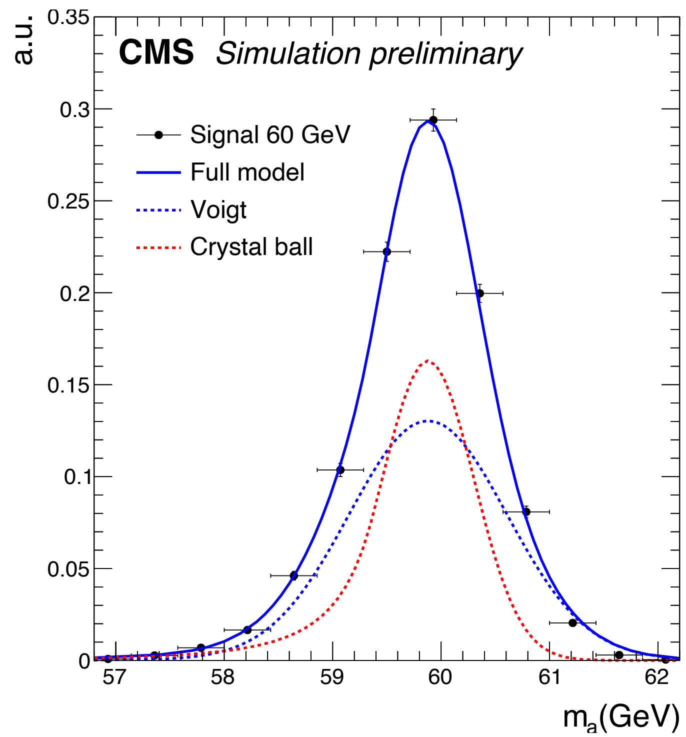
Clean signature with a precise mass resolution from $m_{\mu\mu}$ and large BR from bb

- Search for a masses within $15 < m_a < 62.5$ GeV
- Bump hunt analysis using the dimuon invariant mass $m_{\mu\mu}$

Most stringent observed upper limit till date in this final state, slightly better than ATLAS results

- Difference in analysis strategy with ATLAS: unbinned maximum likelihood fit, completely data-driven background
- Parameters of the signal model (Voight profile+Crystal Ball) are independent of m_a, only the resolution of the model varies linearly with m_a





2016-only result: BR($H \to aa \to 2\mu 2b$) values constrained at 95% CL between (1-7)x10^{-4} depending on m_a

2μ2b: event selection

- Signal events completely reconstructed from final state particles → not expected to produce any p_T^{miss}
  - Events should have at least two muons with opposite charge and at least two b-tagged jets
  - Both single and dimuon triggers are considered, p_T thresholds for muons are 17 and 15 GeV for the leading and subleading
  - Signal does not produce any genuine neutrino: $p_T^{miss} < 60$ GeV

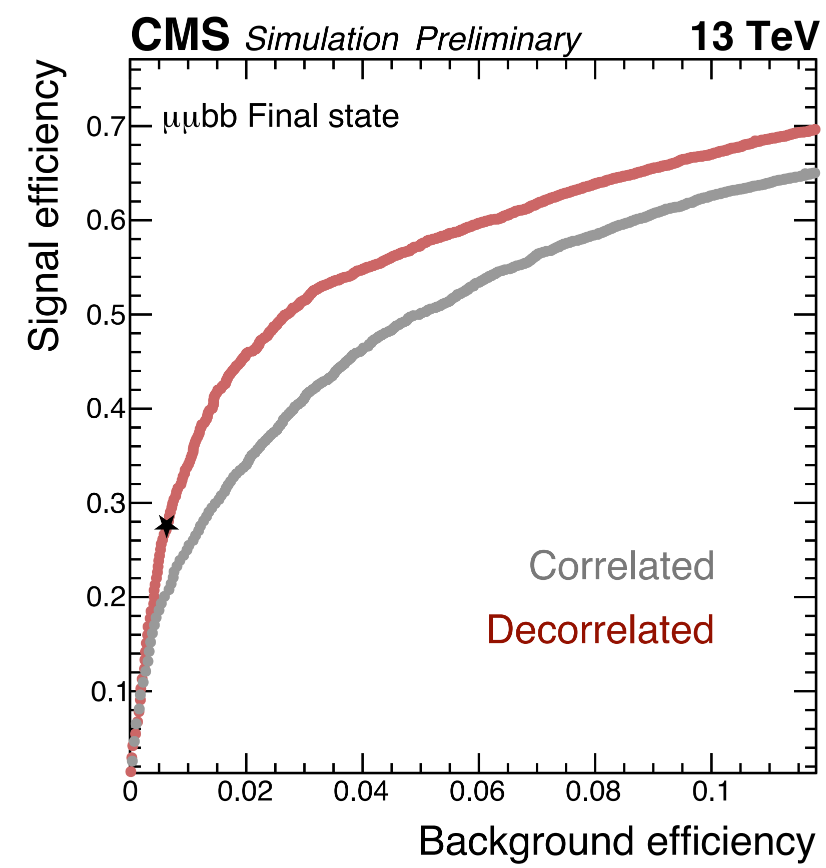
- Use of a single discriminating variable to suppress background:
  - Exploit the mass constraint: $m_{bb} \sim m_{\mu\mu} = m_a$ and $m_{bb\mu\mu} \sim m_H$

$$\chi^2_{\rm tot} \equiv \chi^2_{\rm bb} + \chi^2_{\rm H}$$

$$\chi_{\rm bb} \equiv \frac{(m_{\rm bb} - m_{\mu\mu})}{\sigma_{\rm bb}}$$

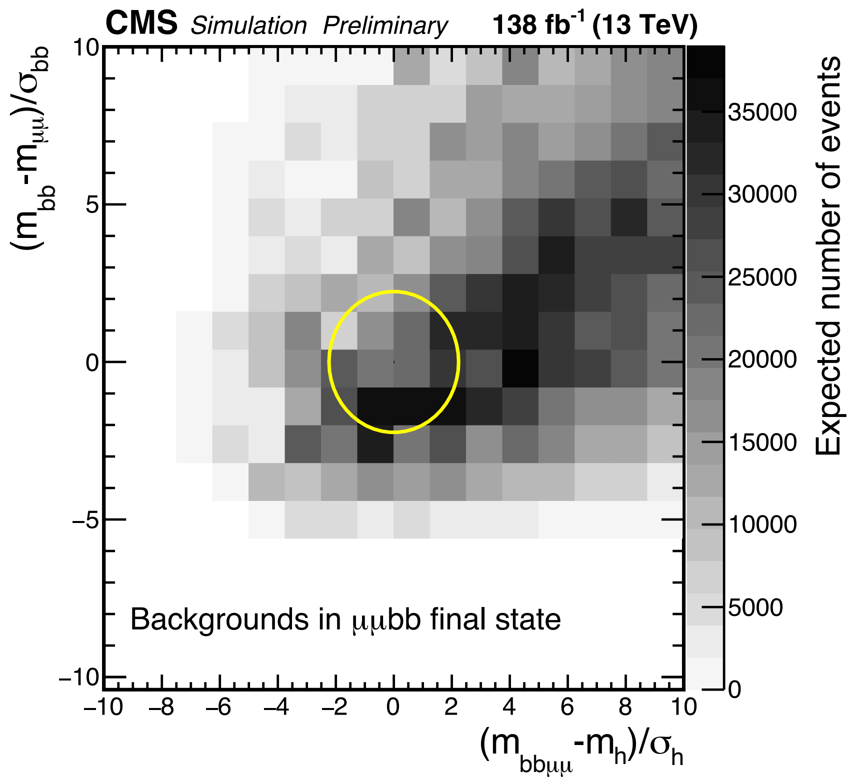
$$\chi_{\rm H} \equiv \frac{(m_{\mu\mu \rm bb} - 125)}{\sigma_{\rm H}}$$

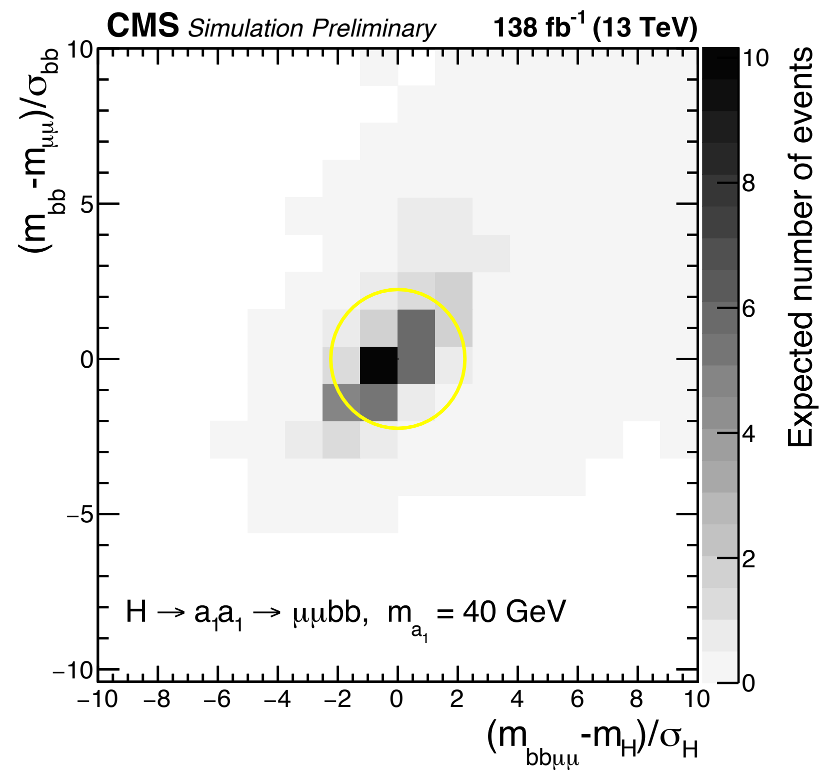
  - Single cut on $\chi_{\rm tot}$, but $\chi_{\rm H}$ and $\chi_{\rm bb}$ are correlated
  - Decorrelate using principal component analysis method



2μ2b: signal categorization

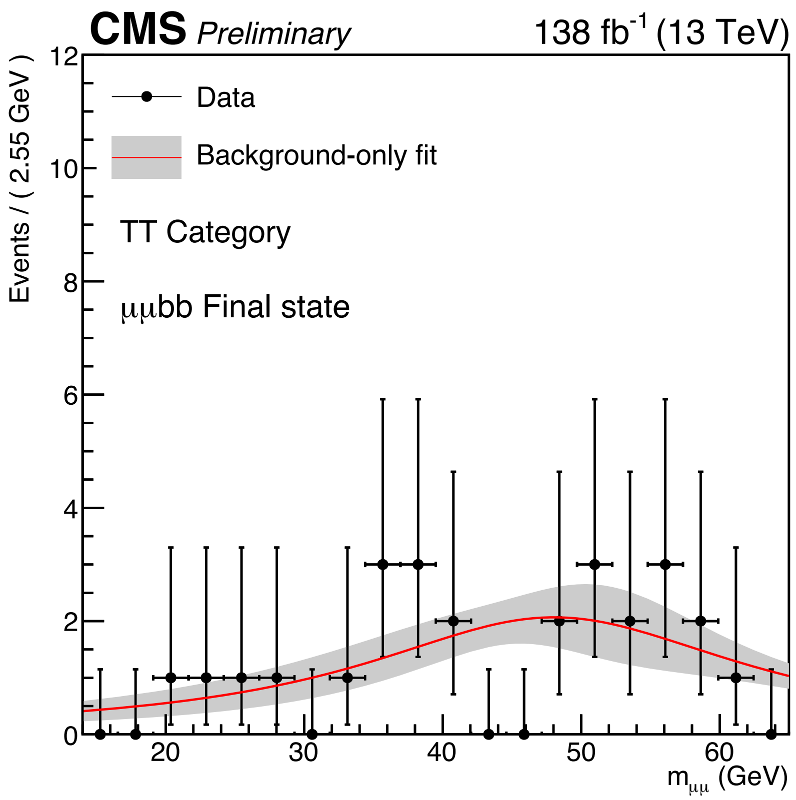
- Select $\chi^2_d < 1.5$ following optimisation studies using simulated events

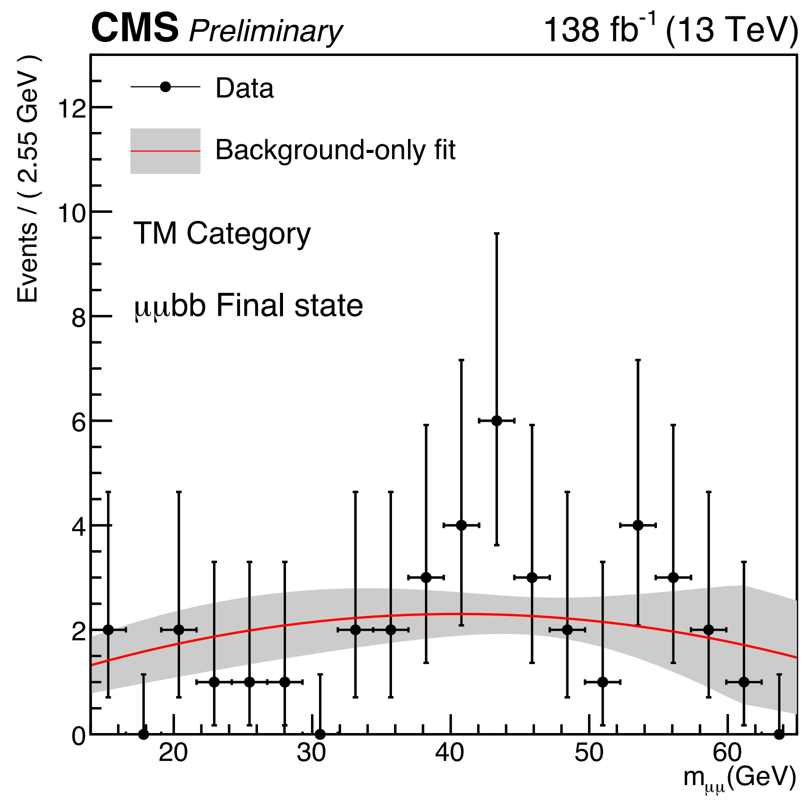


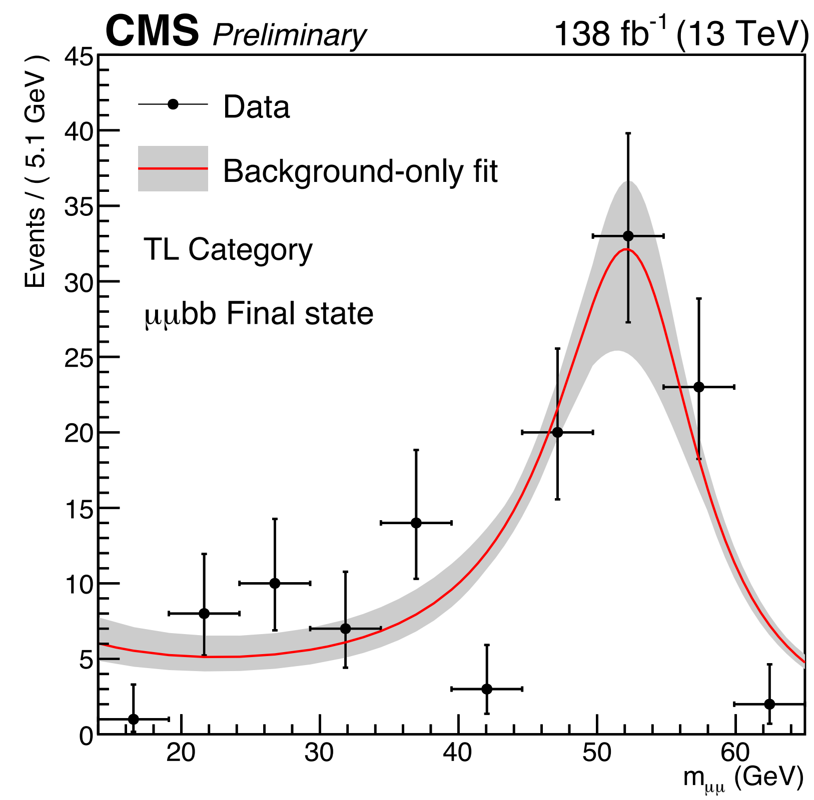


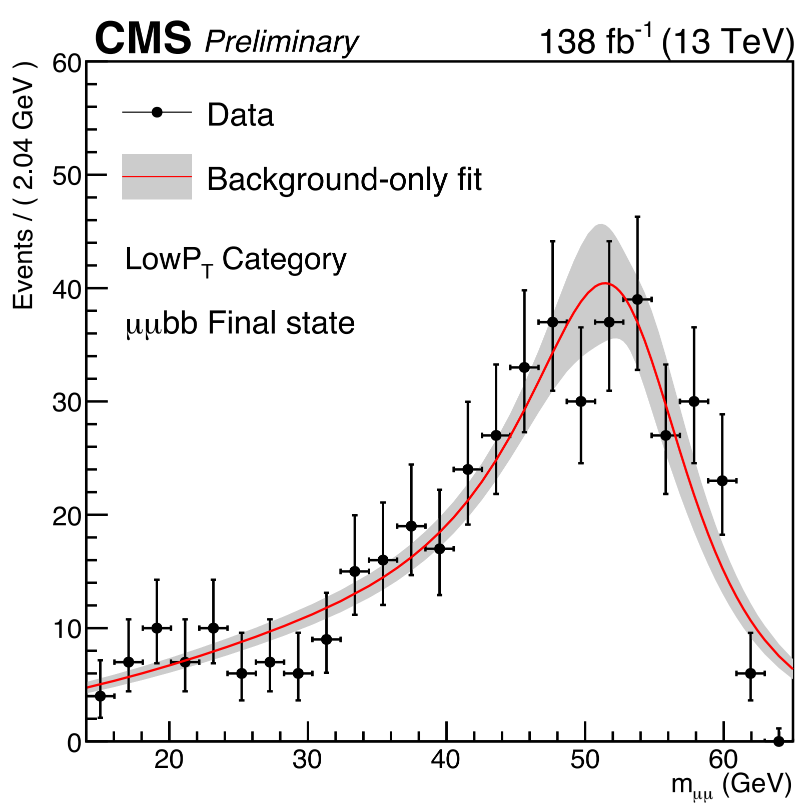
- Events are further categorised based on jet p_T and b-tag score
  - Low p_T: at least one b-jet with $p_T < 20$ GeV
  - VBF: two additional jets with $p_T > 30$ GeV, $|\eta| < 4.7$ and $m_{jj} > 250$ GeV
  - TL: looser b-jet passes L but fails M
  - TM: looser b-jet passes M but fails T
  - TT: looser b-jet passes T

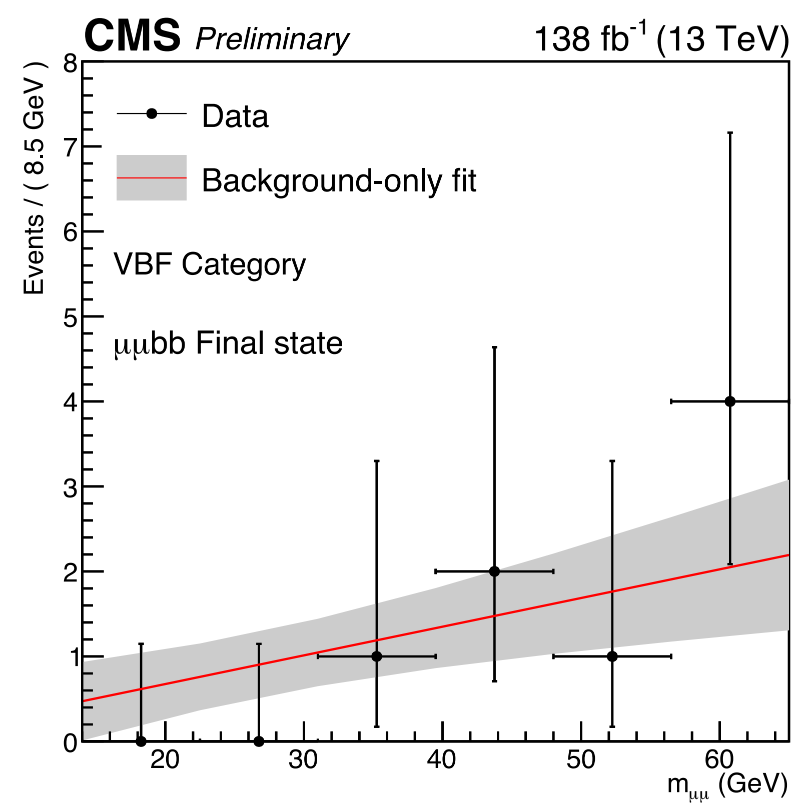
2μ2b: signal region $m_{\mu\mu}$ distributions











$H \to aa \to 2\tau 2b$

Relatively larger BR to bb and ττ, improved τ lepton reconstruction techniques

- Search for a masses within $12 < m_a < 60$ GeV
- Three final states explored: $e\mu$, $e\tau_h$, $\mu\tau_h$

Improved results compared to the previous analysis using partial Run-2 data (2016)

- Addition of > 1 b-jet category made possible due to increased statistics
- DNN categorisation vs. cut based event selection strategy
- SVfit algorithm to reconstruct di-tau invariant mass $m_{\tau\tau}$ including neutrino energies instead of only visible components of $m_{\tau\tau}$ distribution
- Better object reconstruction techniques based on DNN developed within CMS experiment in the recent years: DeepJet, DeepTau tagging
- More precise estimation of $Z \to \tau\tau$ using the embedding technique

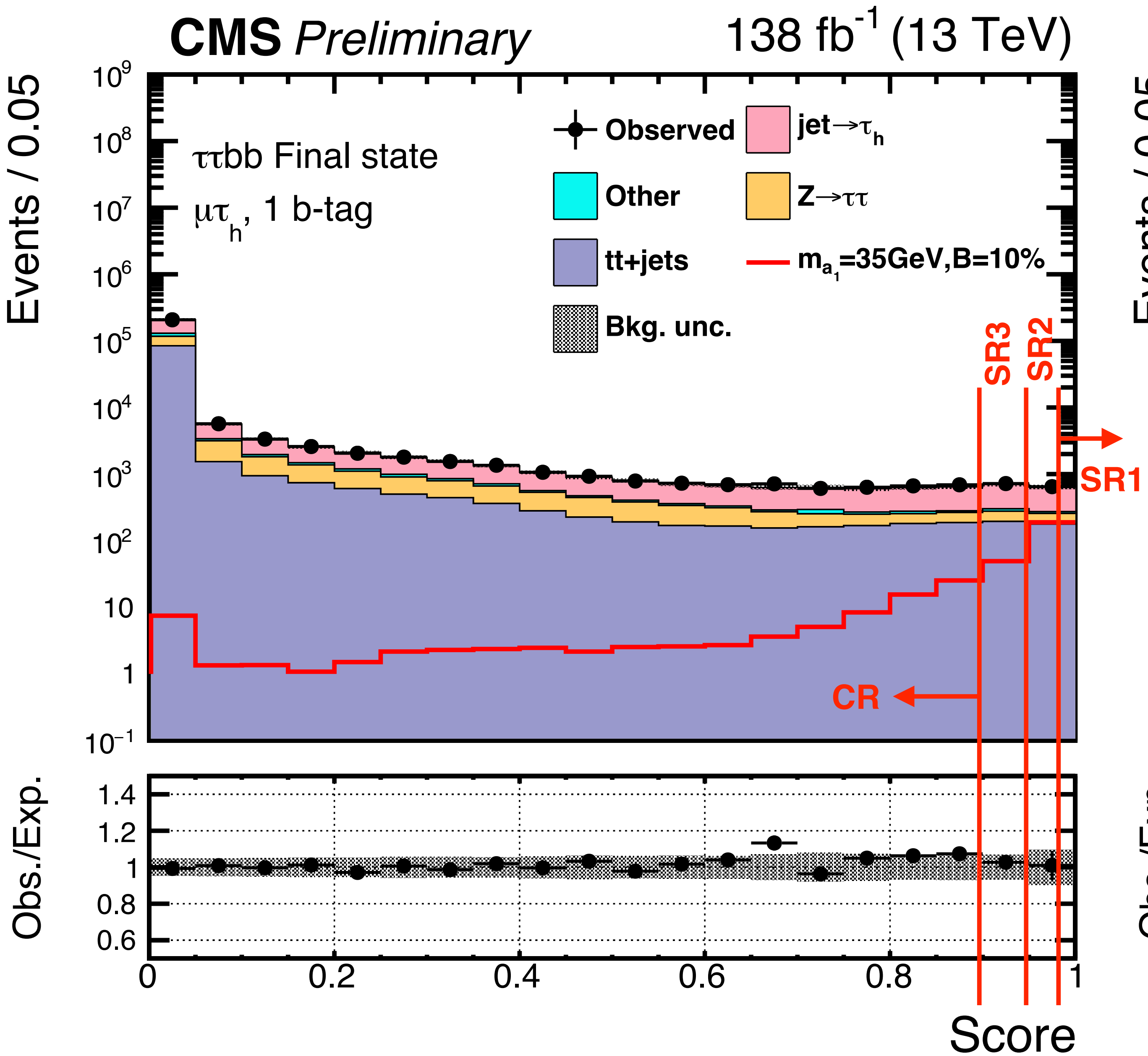
2016-only result: BR($H \to aa \to 2\tau 2b$) values constrained at 95% CL below (3-12)x10^{-2} depending on m_a

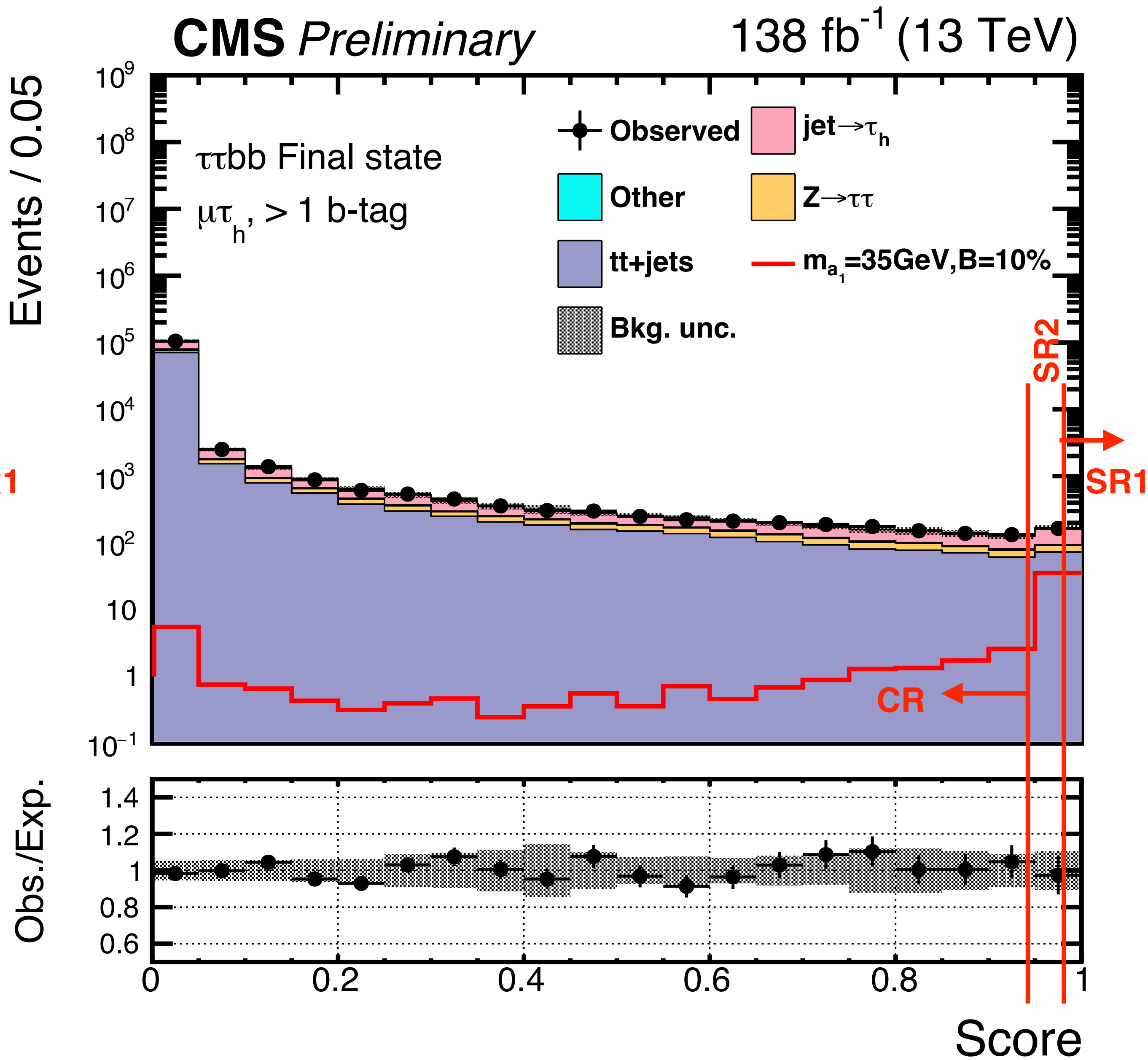
$2\tau 2b$: event selection

- Only three di-tau final states considered:
  - ee and $\mu\mu$ have low BR and large background from Drell-Yan process (DY+jets)
  - $\tau_h\tau_h$ has high trigger threshold
  - Extra lepton veto applied for each of the three final states to ensure mutually exclusive selection
  - Both single and cross-triggers used for each final state, with p_T thresholds being 1 GeV larger than the online threshold for e, μ; offline p_T threshold for τ_h is 35 GeV
- Events should have at least one loosely tagged b-jet with $p_T > 20$ GeV
  - Two broad categories based on b-jet multiplicity: = 1 and > 1 b-jet
- DNN categorisation:
  - Discriminate signal against a combination of major backgrounds ($t\bar{t}$+jets and DY+jets)
  - Train one DNN for each of the three channels and two b-jet categories: six in total
  - Training variables are based on kinematics of reconstructed final state particles
  - Split the selected events further into smaller categories based on the DNN scores: events with high DNN scores constitute signal regions (SR); background rich categories are taken as control regions (CR)

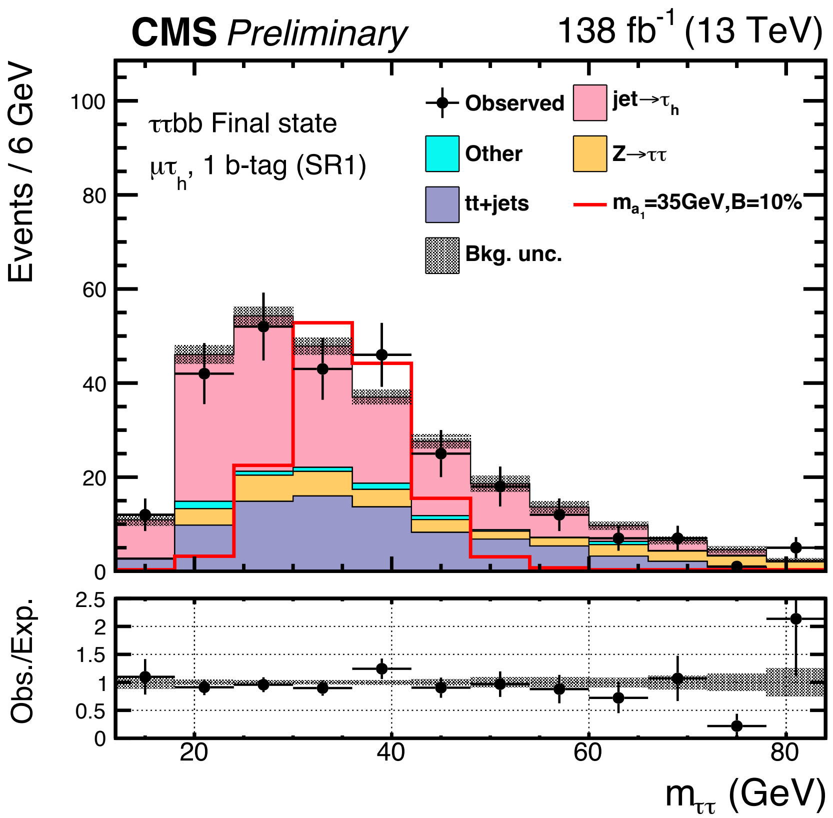
2τ2b: signal categorization

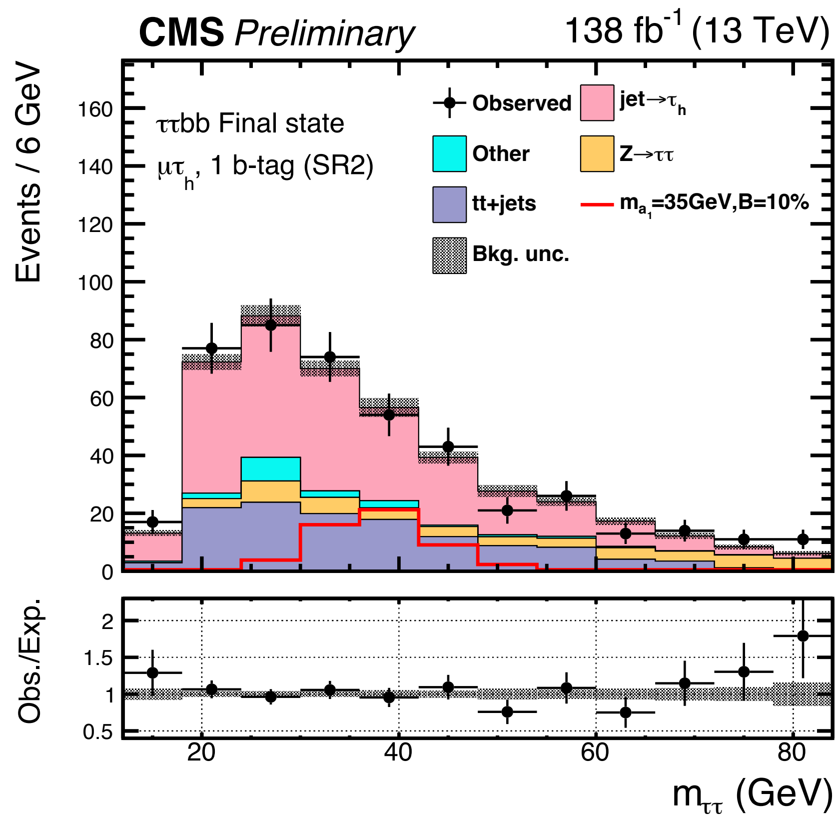
- Discriminating observables include invariant mass of visible decay products, transverse mass between an object and p_T^{miss}, m_{bb}-$m_{\tau\tau}$ etc.
- Final observable used in maximum likelihood fit: $m_{\tau\tau}$, not used as an input to DNN

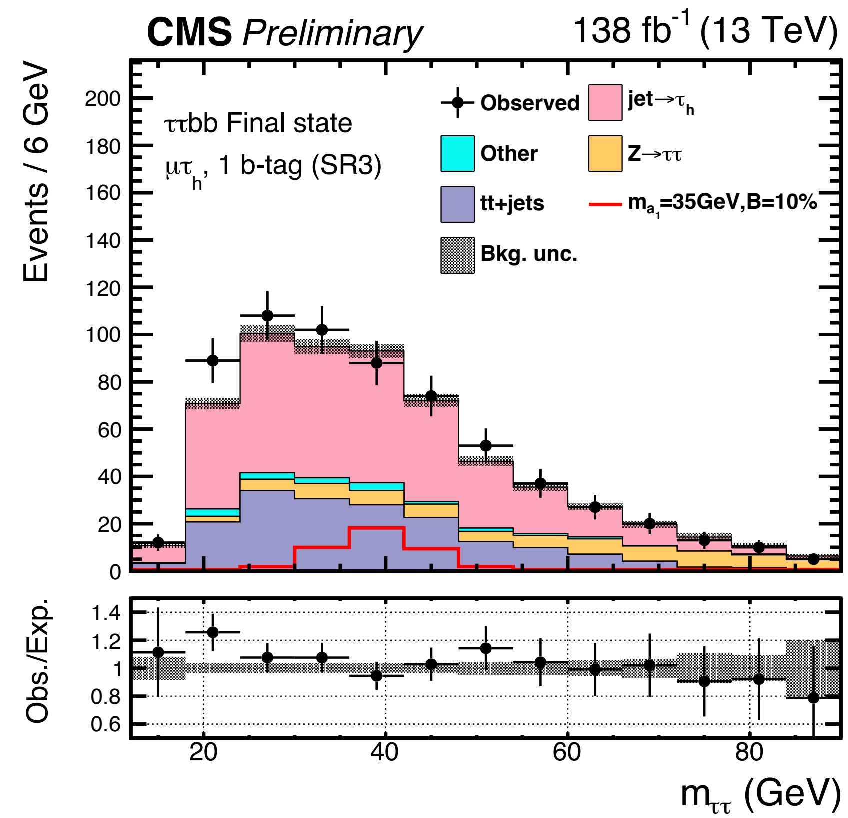


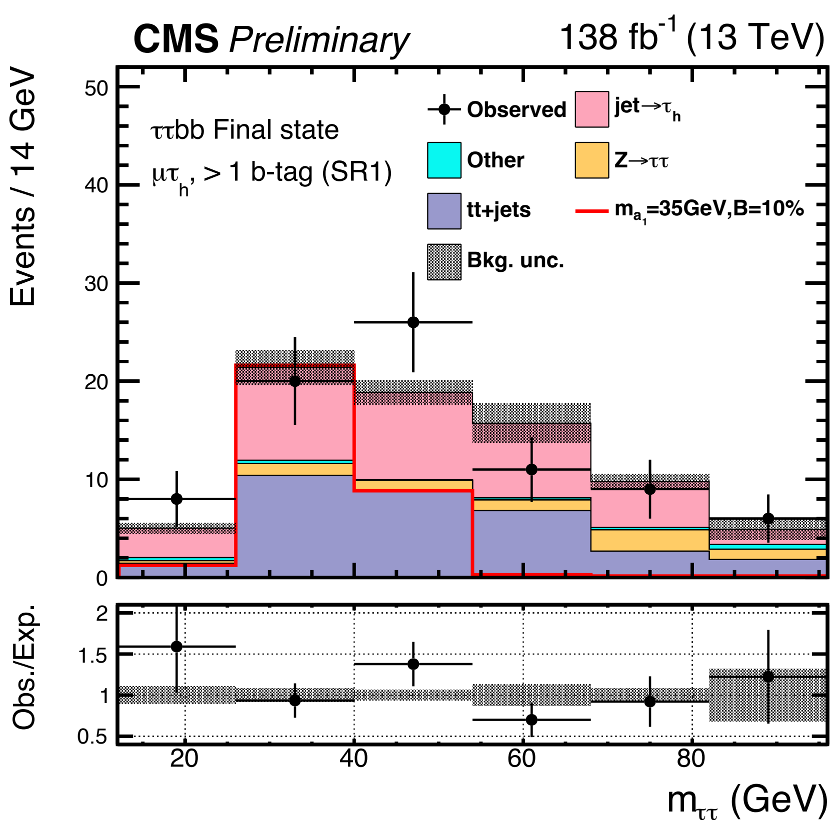


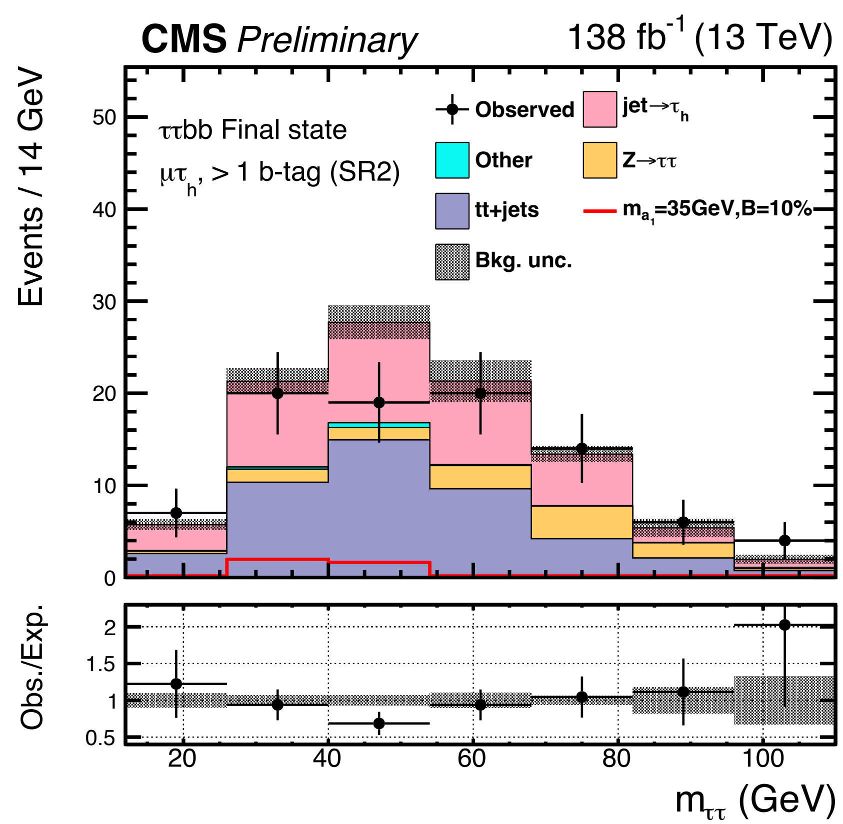
2τ2b: signal region $m_{\tau\tau}$ distributions











most sensitive channel: $\mu\tau_h$

Background estimation methods

For the $H \to aa \to 2\mu 2b$ analysis background evaluated through the ML fit, without any reference to simulation

- Discrete profiling method is used choose best-fit from a pool of background models

In the $H \to aa \to 2\tau 2b$ analysis:

- Irreducible physics backgrounds: genuine particles forming the $\tau\tau$ final state from other physics processes
  - $t\bar{t}$+jets
  - Diboson, single top, SM Higgs$\to\tau\tau$/WW
  - $Z\to\tau\tau$: the limitations in reconstructing taus is overcome by selecting well reconstructed $Z\to\mu\mu$ events from data and replacing the muon candidates with simulated taus
- Reducible backgrounds: mis-identified or *fake* particles forming the final state are estimated from data
  - Jets faking τ_h: W+jets and QCD processes have large jet multiplicity, leading to fake τ_h
  - QCD process in $e\mu$ channel: jets can also be mis-identified as e/μ and are most significant in QCD process

Systematic uncertainties

- Two broad categories: experimental and theoretical, most of which are common to both analyses

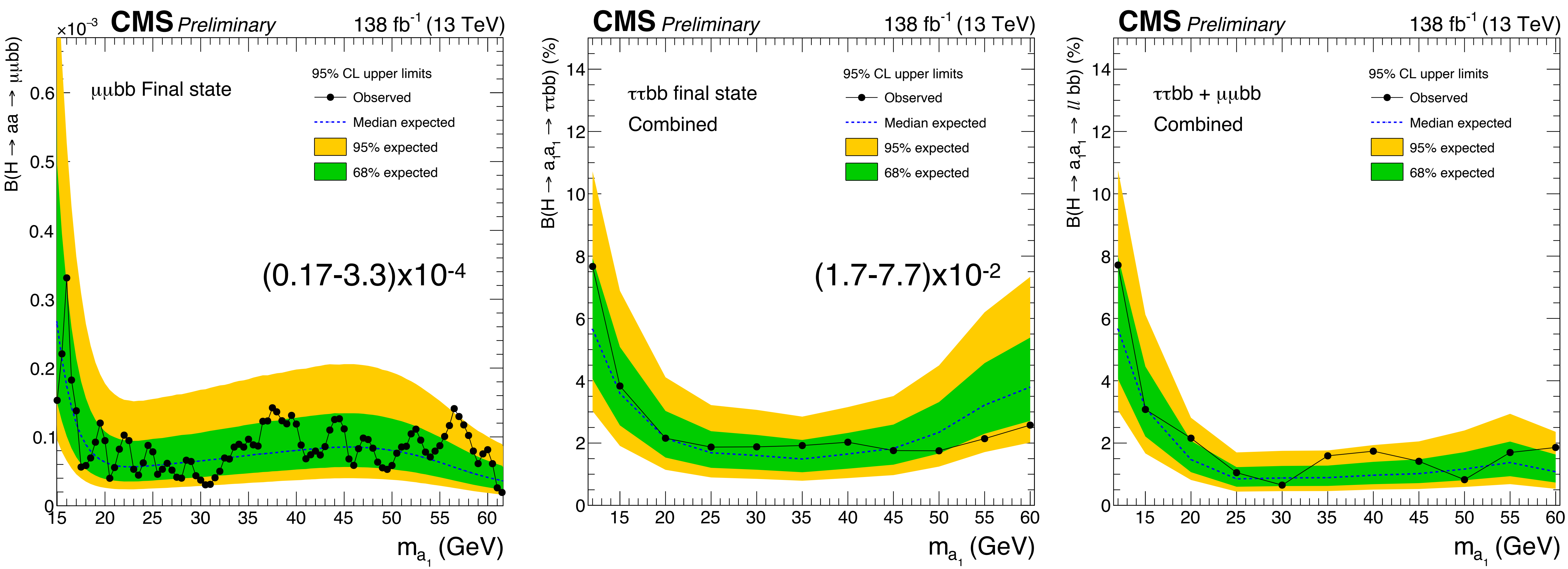
- Experimental:
  - Luminosity measurement
  - Uncertainty in measuring efficiency scale factors for e/μ/τ_h selection and trigger
  - Jet energy correction and b-tagging efficiencies
  - ECAL timing shift due to misalignment
  - Background estimations:
    - Normalisation of various SM process
    - Uncertainty in measuring different fake rates/scale factors for data-driven backgrounds
    - Uncertainty in estimating the embedded background
    - Uncertainty for imprecise background modelling from discrete profile method

2τ2b
2μ2b

- Theoretical:
  - Uncertainty in the ggF and VBF production cross sections of the Higgs boson
  - Scale variations in $t\bar{t}$+jets, single top and diboson simulations
  - Parton-shower uncertainties in $t\bar{t}$+jets

Upper limits on exotic Higgs BR

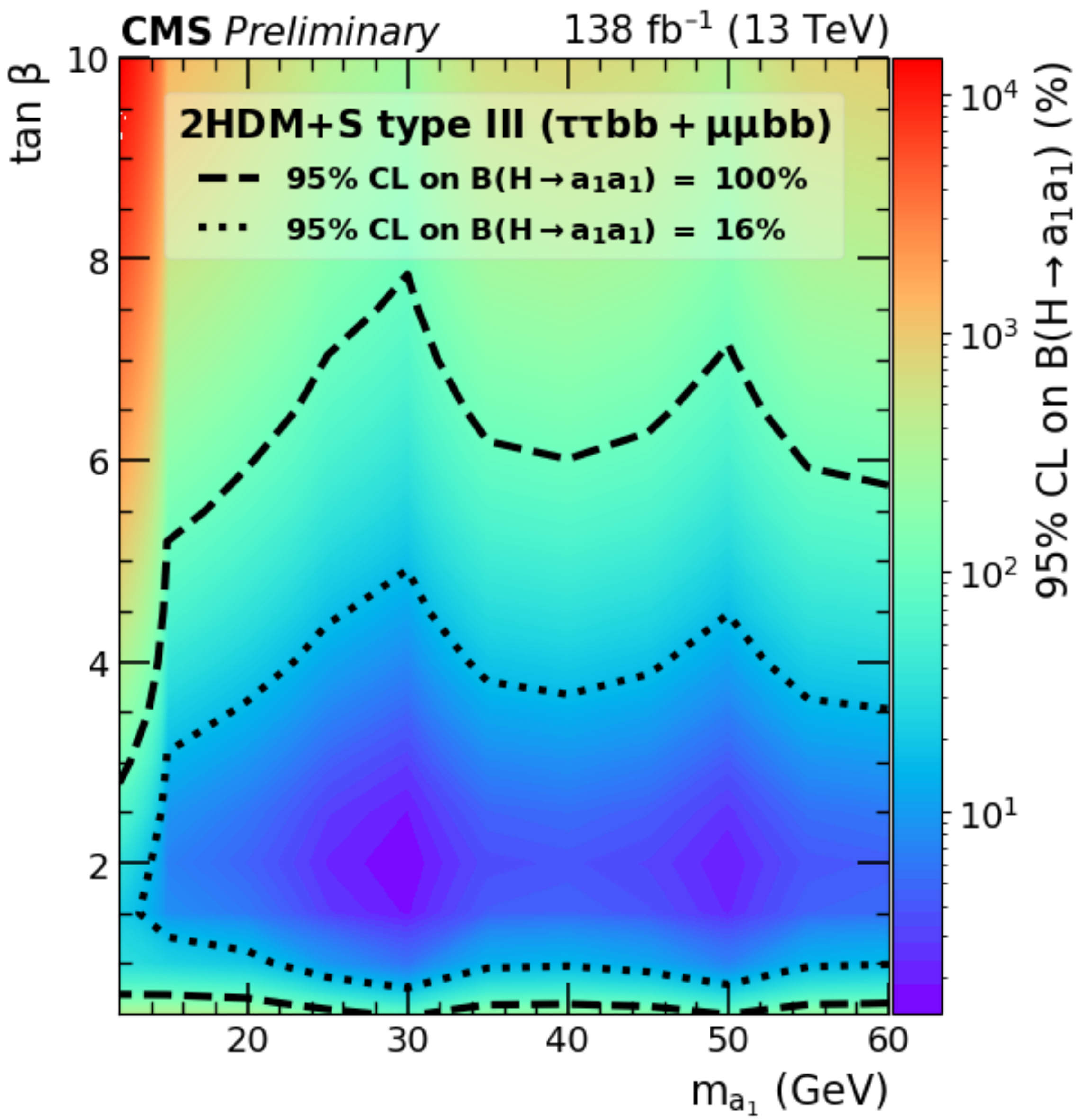
- Straightforward statistical combination: analyses utilise orthogonal data samples
- Some common uncertainties are treated as correlated, such as luminosity measurement, jet energy scale, variations in signal cross section etc
- Type-independent upper limits on BR(H→aa→llbb) in the context of 2HDM+S are derived as a function of m_a where l is a μ or τ

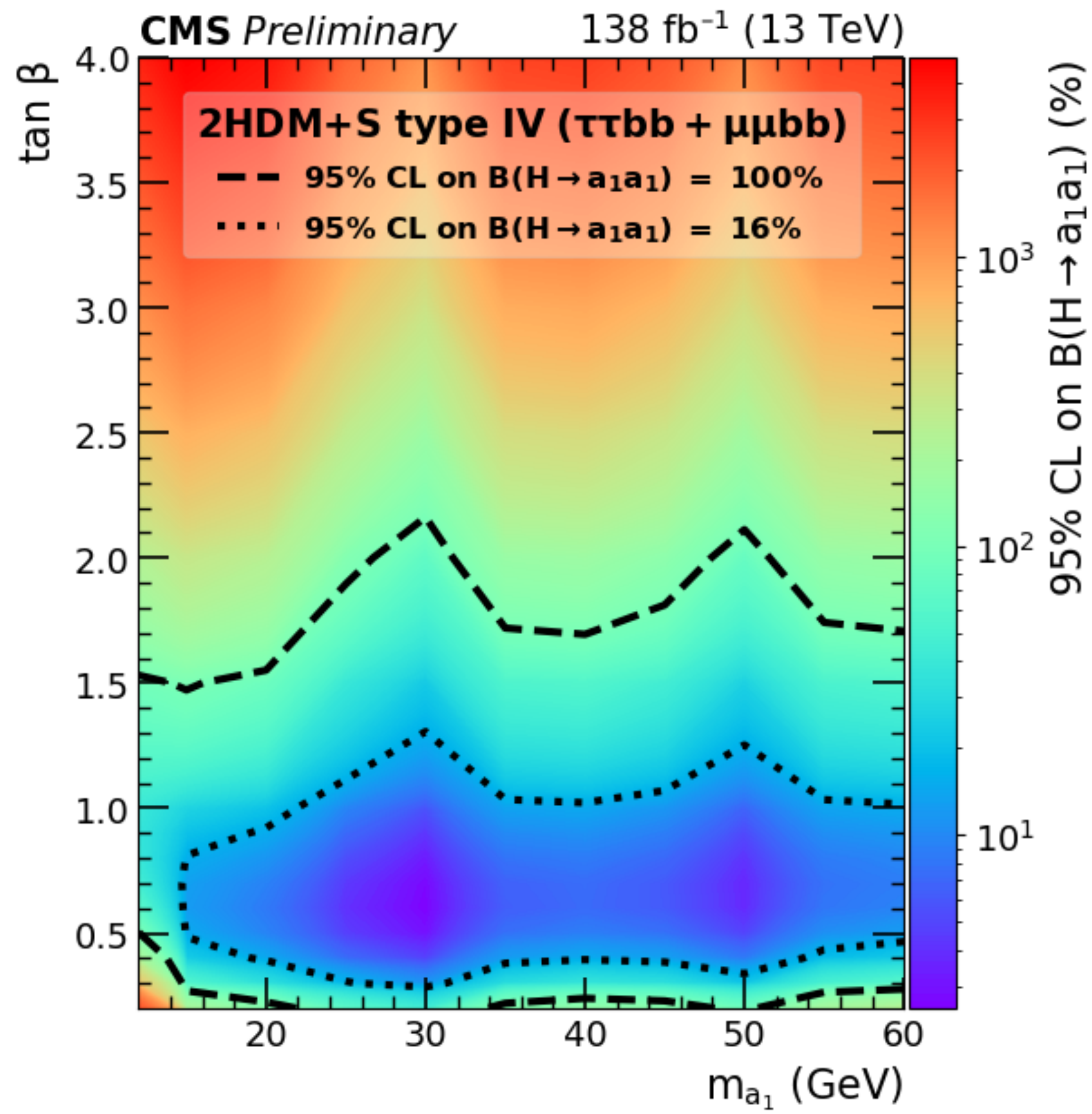


Interpreting in terms of different 2HDM+S: BR(H→aa) values excluded above 23% (Type II tanβ > 1), 7% (Type III tanβ = 2.0) and 15% (Type IV tanβ = 0.5)

Implications for different models

Stringent upper limits are set for most Type III and Type IV 2HDM+S scenarios





16% contour corresponds to combined upper limit on Higgs to invisible decays obtained from previous Run 2 results

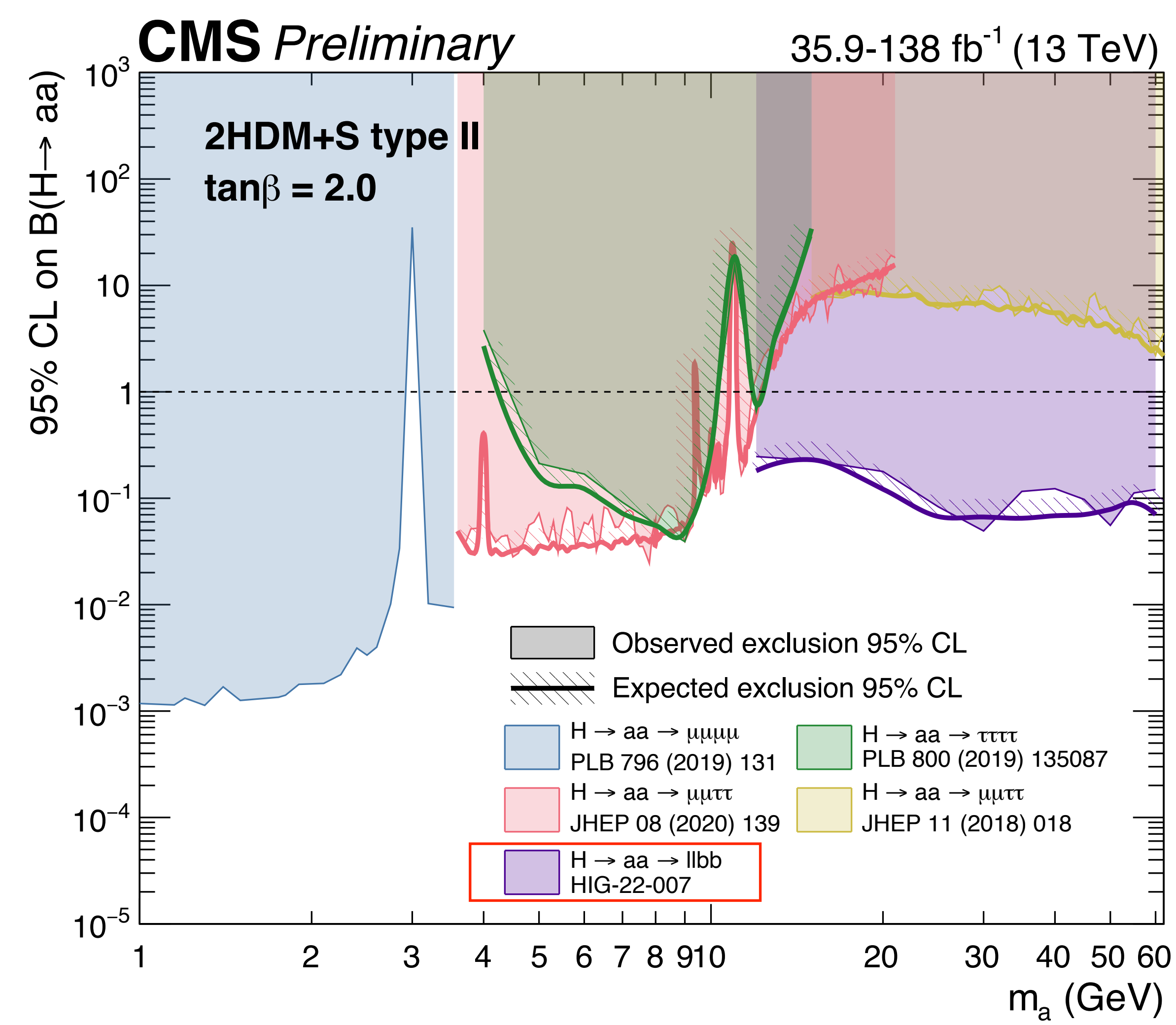
Summary

Higgs portal to hidden BSM sector being explored by CMS analyses in different final states

→ **Many full Run-2 results are public, some are work in progress**

- Improved sensitivity compared to previous searches due to changes in analysis strategy rather than the increase in data statistics alone
- No significant excess over SM prediction *just yet*, many other possibilities remain to be explored
  - Asymmetric pseudoscalar masses
  - Signals with low pseudoscalar mass to be analysed using boosted reconstruction techniques

Direct searches benefit the most with increase in luminosity: look forward to Run-3!



Thank You

Backup

$2\tau 2b$: triggers and objects

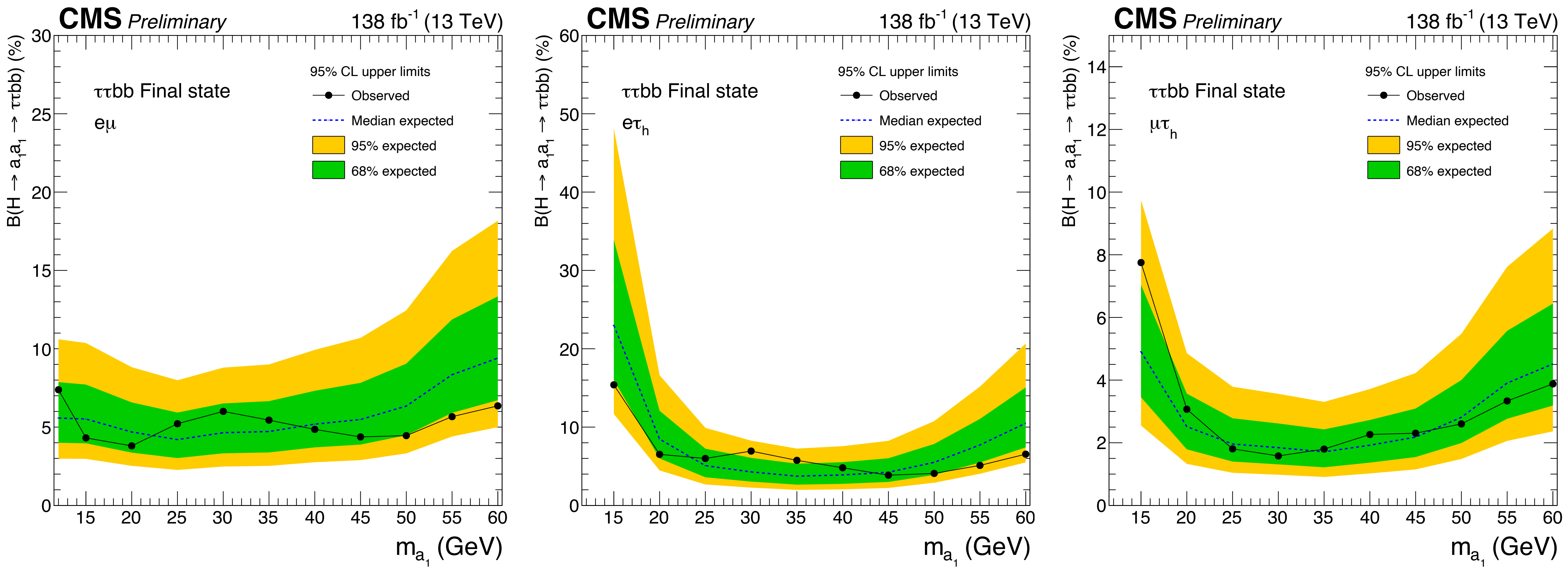
	Type	$e\mu$		$e\tau_h$		$\mu\tau_h$	
		e	μ	e	τ_h	μ	τ_h
2016	single	-	-	25	-	22	-
	leading	23	23	-	-	-	20
	sub-leading	12	8	-	-	19	-
2017	single	-	-	27, 32	-	24, 27	-
	leading	23	23	-	30	-	27
	sub-leading	12	8	24	-	20	-
2018	single	-	-	32, 35	-	24, 27	-
	leading	23	23	-	30	-	27
	sub-leading	12	8	24	-	20	-

- Electrons and muons are reconstructed within $|\eta| < 2.4$ and τ_h within $|\eta| < 2.1$
- Offline e, μ and τ_h are matched to the trigger objects, with p_T thresholds being 1 GeV larger than the online threshold for e, μ; offline p_T threshold for τ_h is 35 GeV
- In case both single and cross-triggers are present in the event, use lowest threshold
- Additional identification/isolation requirements on e/μ/τ_h
- Anti-k_T jets are reconstructed within $|\eta| < 2.4$ using a cone size of 0.4

Upper limits on exotic Higgs BR

Limit is set on SM like Higgs→aa→2τ2b:

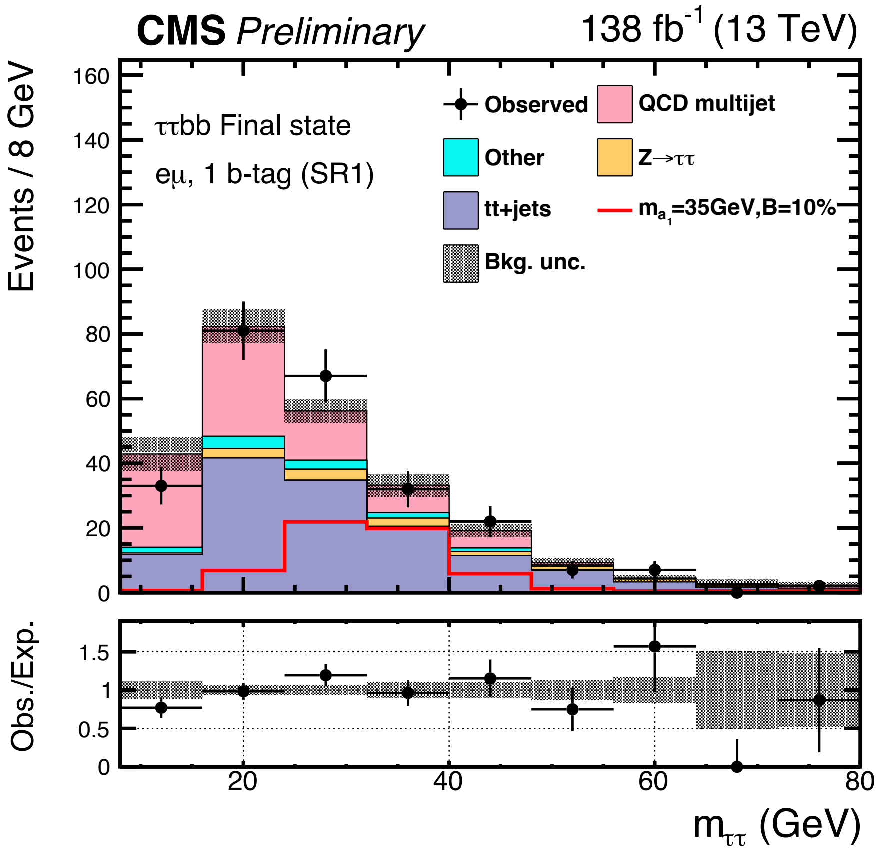
- Most sensitive channel: $\mu\tau_h$, dominant background is Z→ττ and τ_h fakes from QCD multijet
- Dominant systematic uncertainty from fake τ_h background estimation
- Analysis is still statistically limited

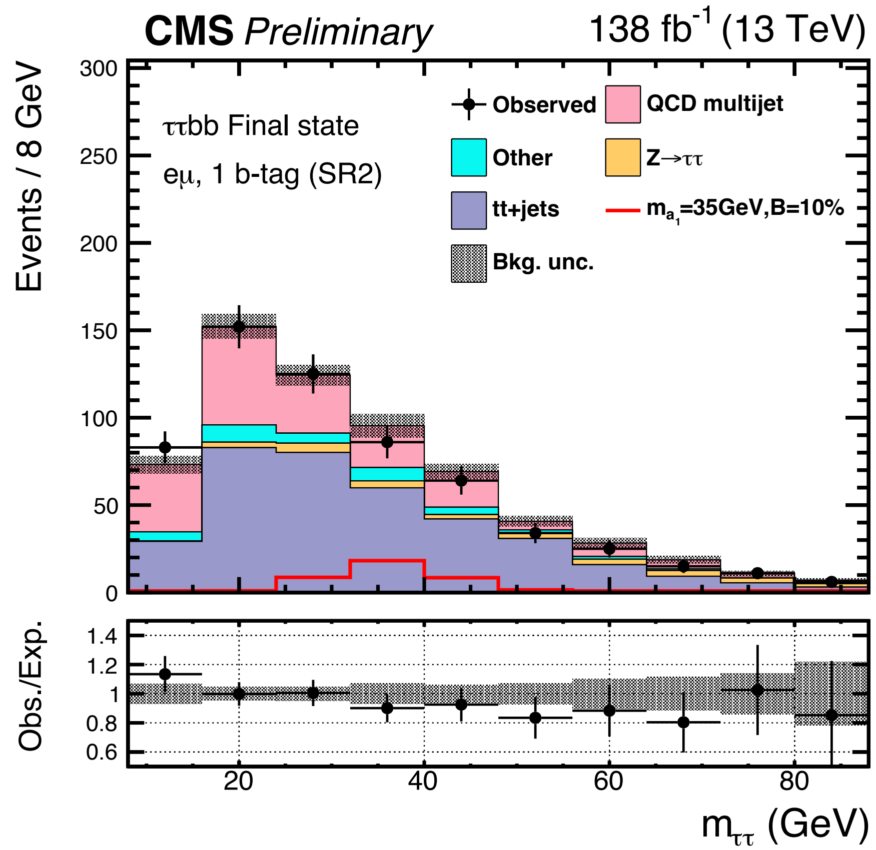


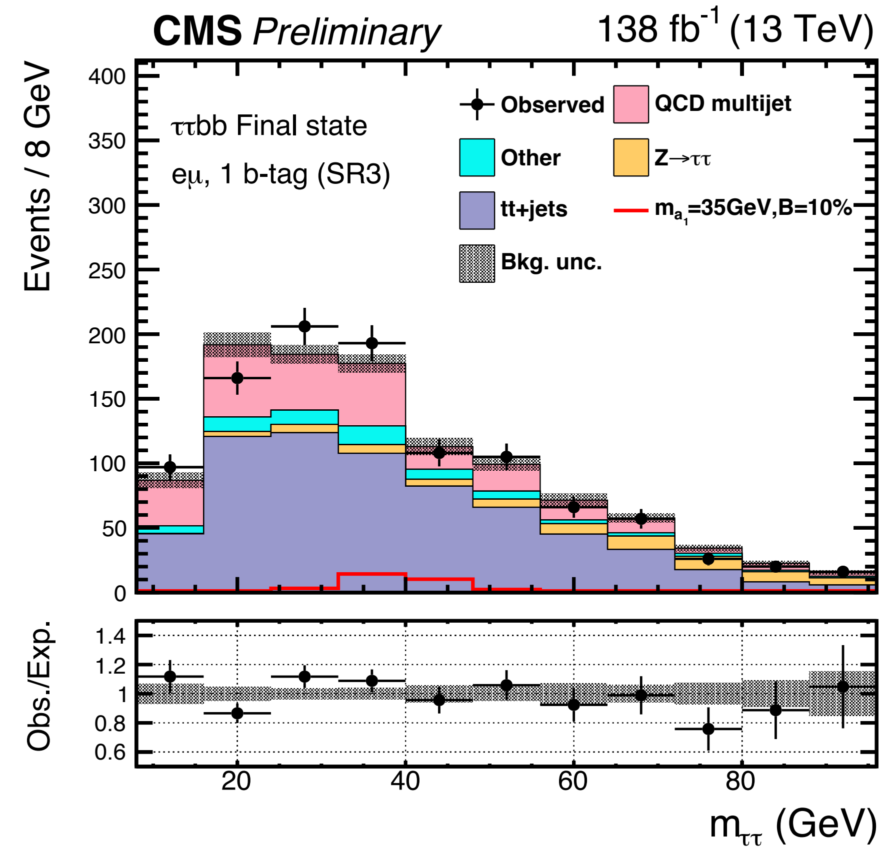
Only the $e\mu$ channel is sensitive to the 12 GeV mass point

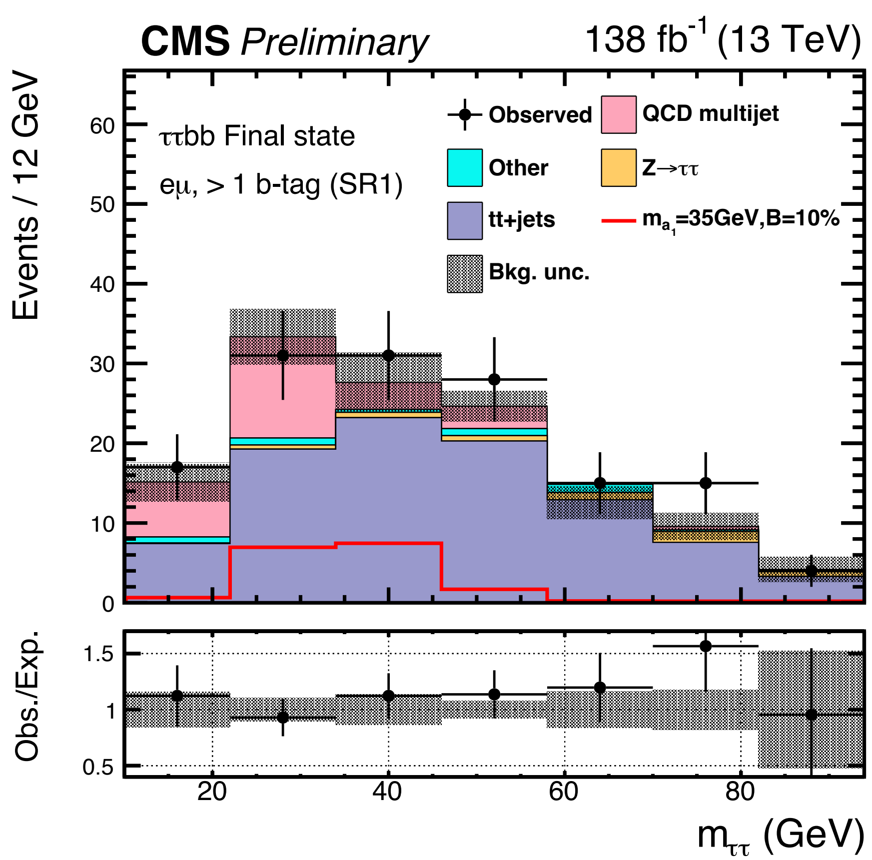
- For low m_a the decay products are boosted, need dedicated reconstruction
- In this analysis, a ΔR requirement is applied between the final state particles, which has a lower threshold in $e\mu$ channel

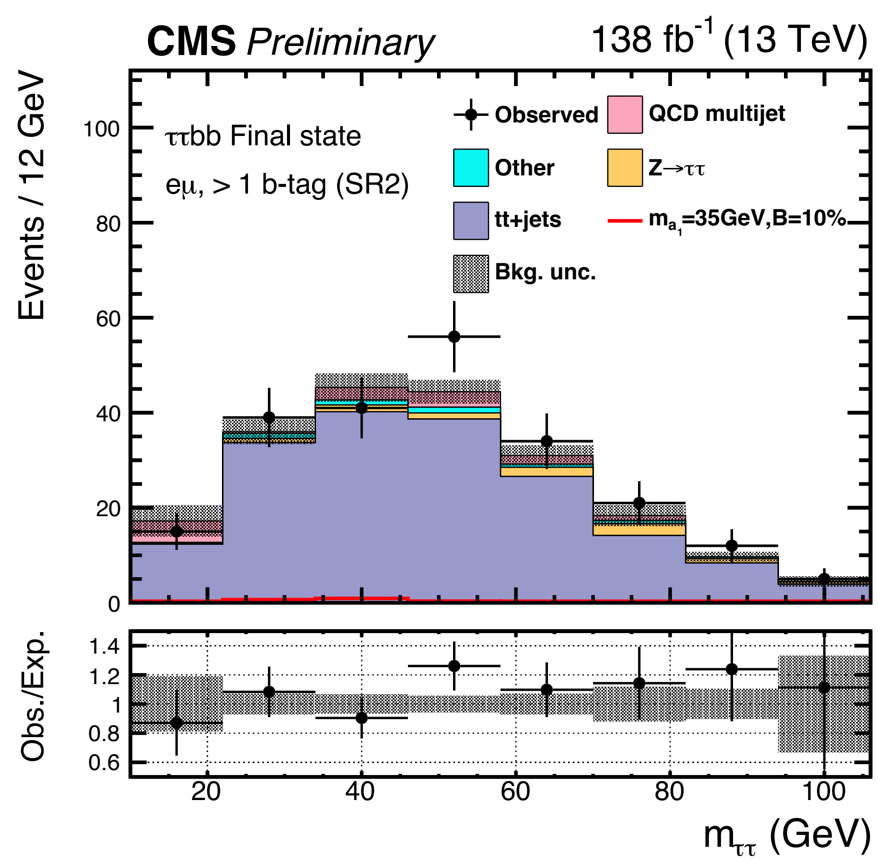
Single region distributions: eμ

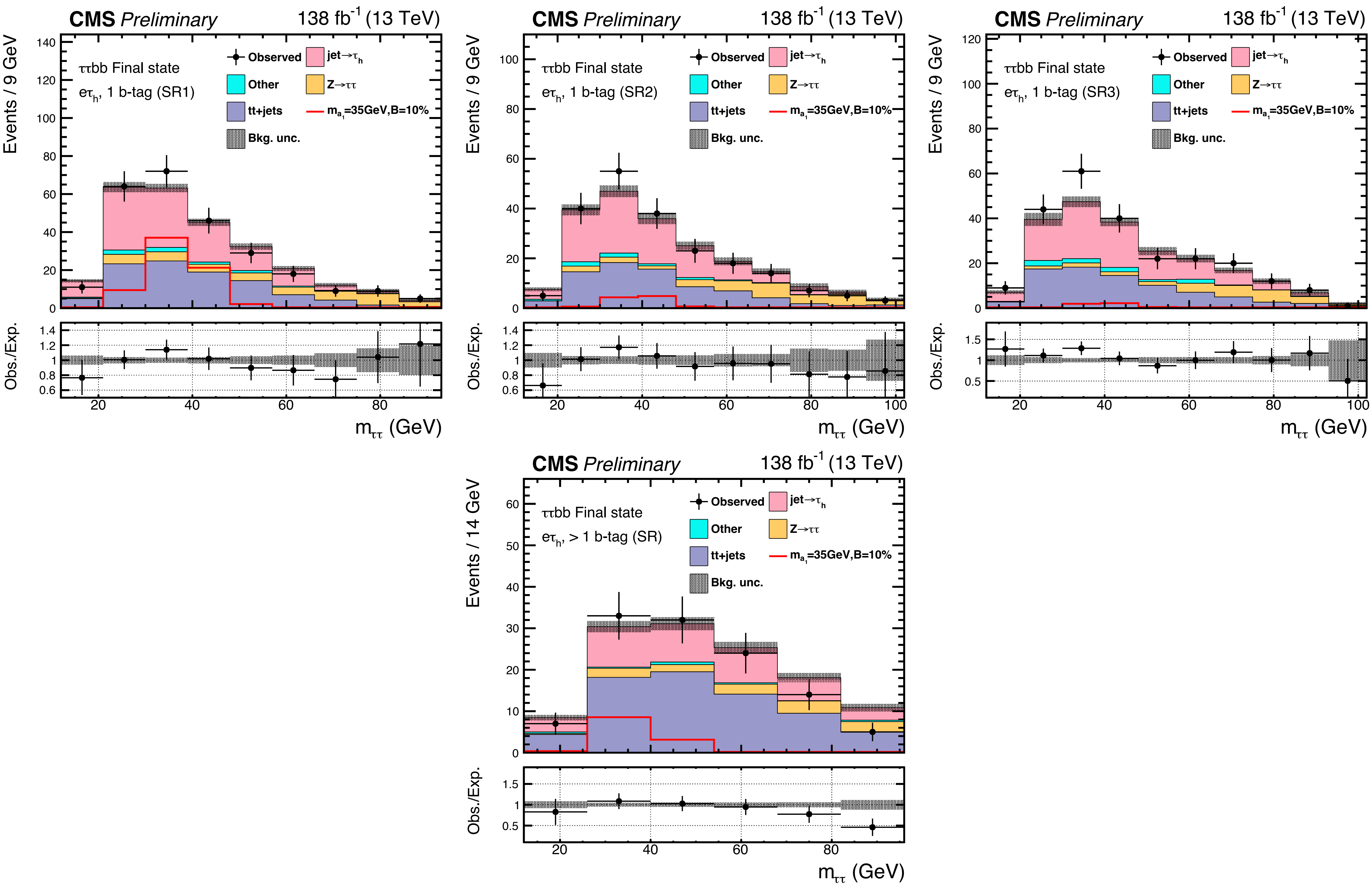











CMS Preliminary
138 fb-1 (13 TeV)
ττbb Final state
eτh, 1 b-tag (SR1)
eτh, 1 b-tag (SR2)
eτh, 1 b-tag (SR3)
eτh, > 1 b-tag (SR)
Observed
jet→τh
Other
Z→ττ
tt+jets
ma1=35GeV,B=10%
Bkg. unc.
Events / 9 GeV
Events / 14 GeV
Obs./Exp.
mττ (GeV)

Single region distributions: μτh