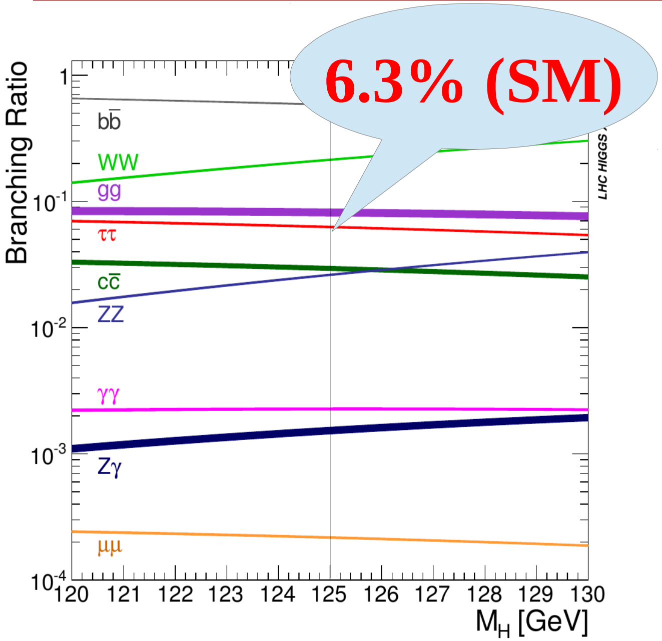
Observation of H → ττ Decays at CMS

Alexei Raspereza

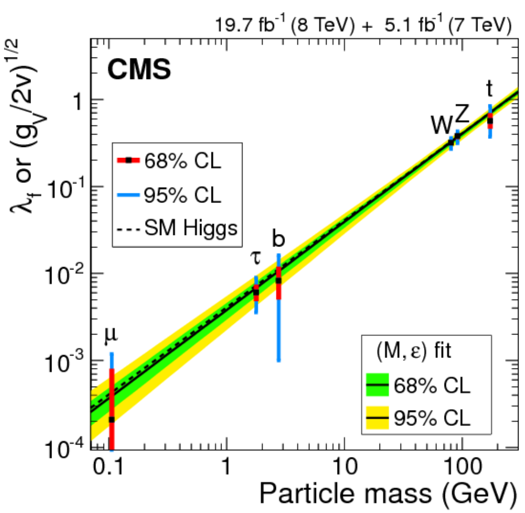
LHC Physics Discussion, 2017/09/11

Why studying H → ττ decays ?



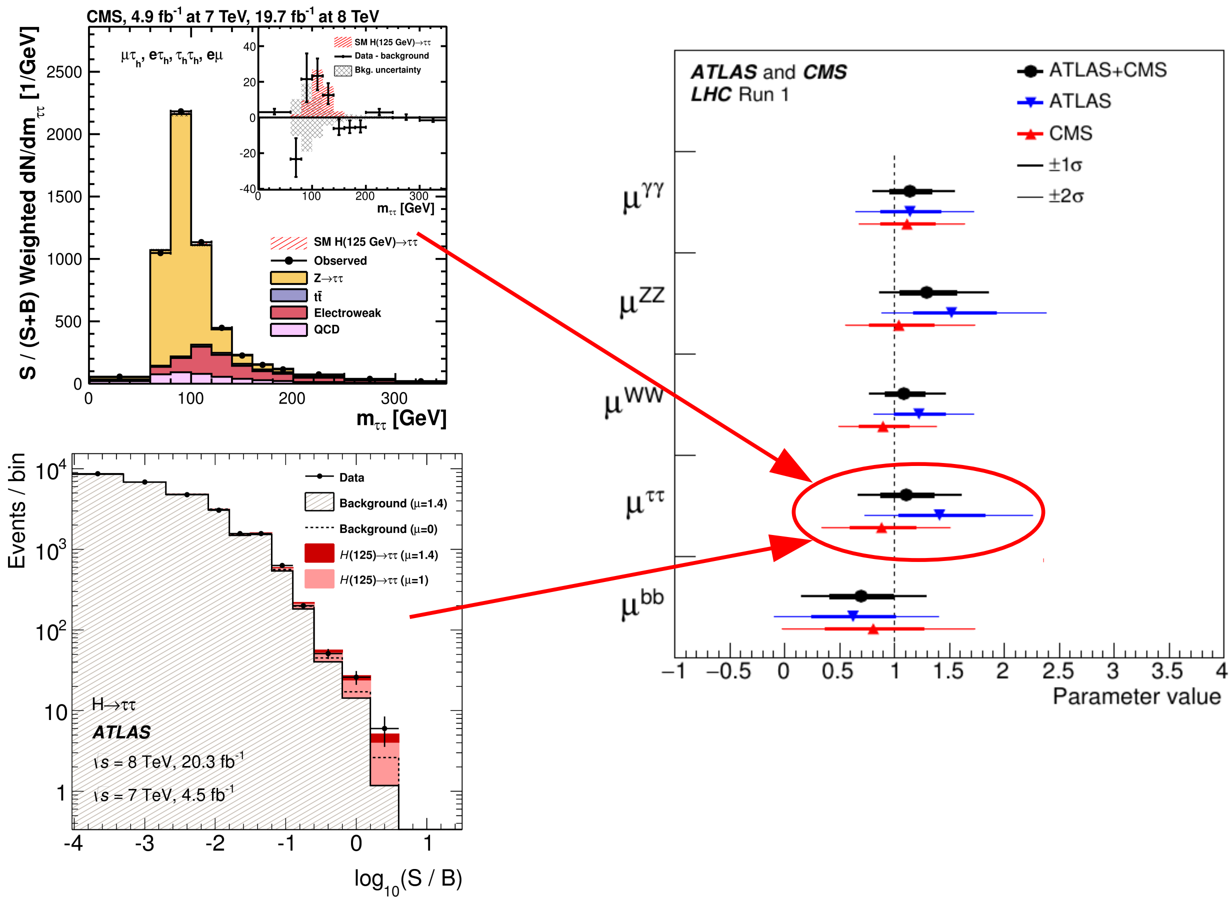
- **sizable BR → high signal rate is expected**
- **probe Yukawa coupling → testing mass-coupling relation for fermions**
- **access to CP quantum numbers (CP mixing)**
- **one of the most sensitive channels to VBF production**
- **boosted Higgs bosons in gg → H**

  → sensitivity to new physics



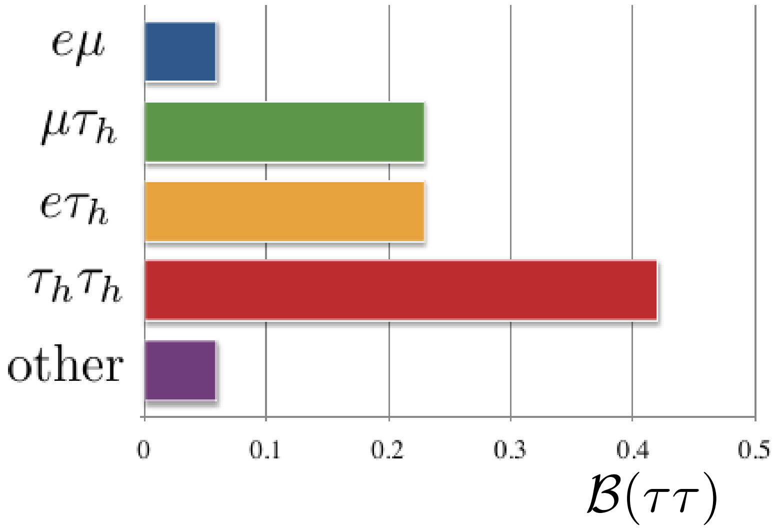
Results from Run1 and goals for Run2

- Run1
  - signal established with significance of 5.4σ in combination of ATLAS and CMS data
  - measurements consistent with SM expectations



- Run2
  - observation of signal by single experiment
  - larger dataset and higher center-of-mass energy → improve precision of coupling measurements (search for possible deviations from the SM)

Analysed Dataset and Decay Modes



- **Dataset :**

  35.9 fb^{-1} collected by CMS at c-o-m energy of 13 TeV

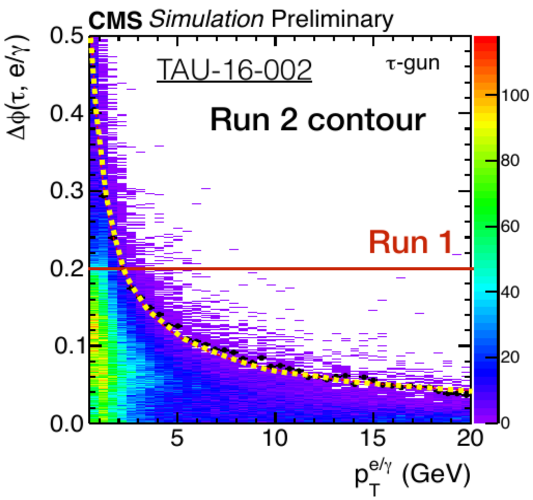
- **four decay modes of tau pairs exploited (94% of final states)**

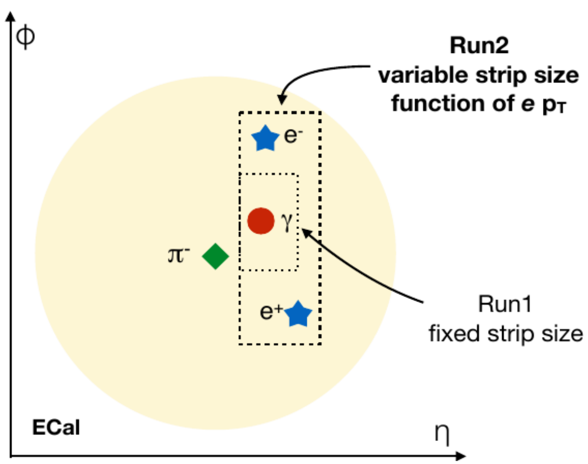
  $e\mu, \quad \mu\tau_h, \quad e\tau_h, \quad \tau_h\tau_h$

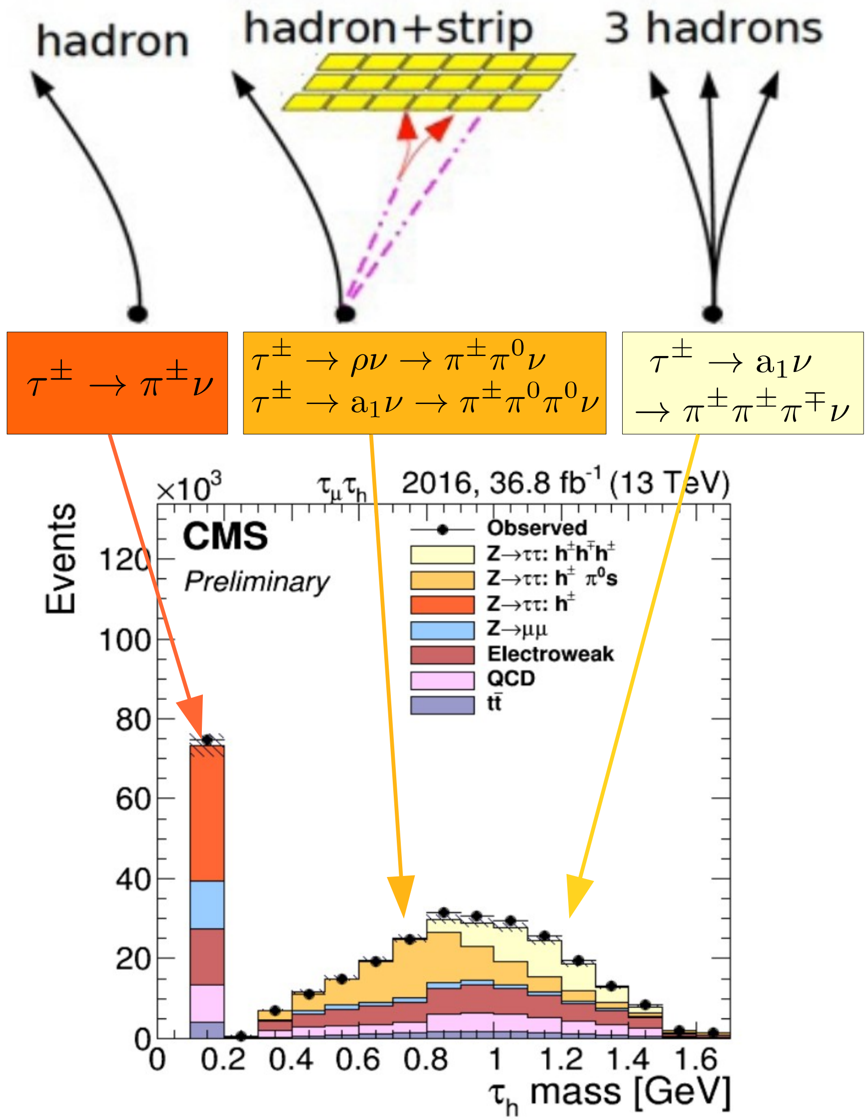
- **Final states with hadronically decaying tau leptons amount to 88% of all di-tau decays**
  - **efficient triggering and identification of τ_h in harsh pileup environment is essential**

τ_h Identification at CMS

- τ_h seeded by anti-k_T jets (ΔR_{cone}=0.4)
- uses as input particle-flow objects

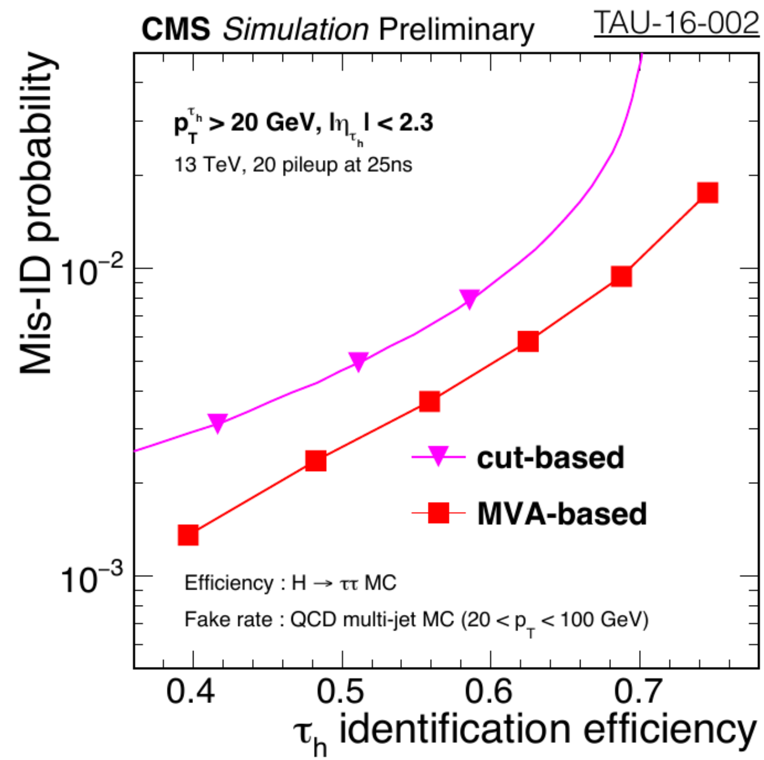


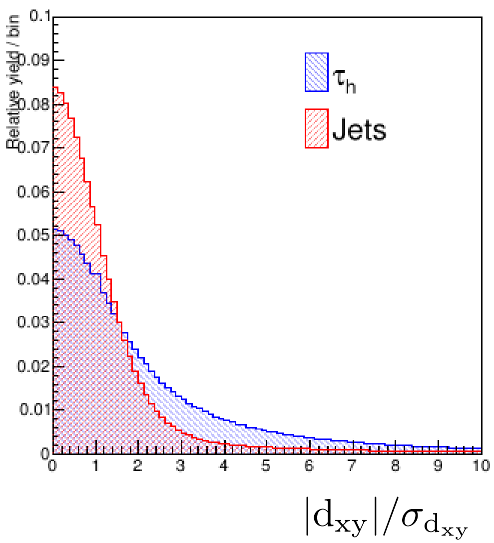


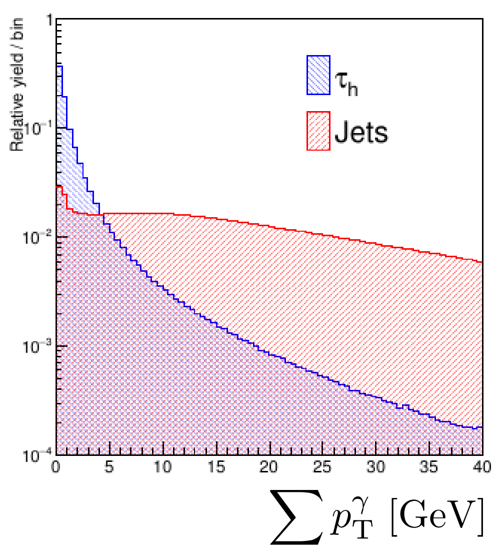


τ_h Identification

- **Run 1 approach : cut on isolation variables**
- **Run 2 approach : MVA discrimination of τ_h against hadronic jets**
  - **isolation variables (p_T sums)**
  - **τ_h decay length information (track impact parameters, SV decay length)**
  - **multiplicity of particle-flow objects in signal/isolation cones**
  - **calorimeter cluster shapes**



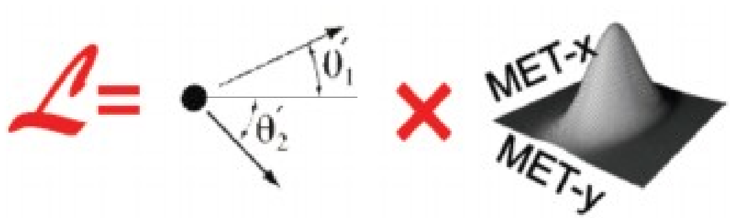




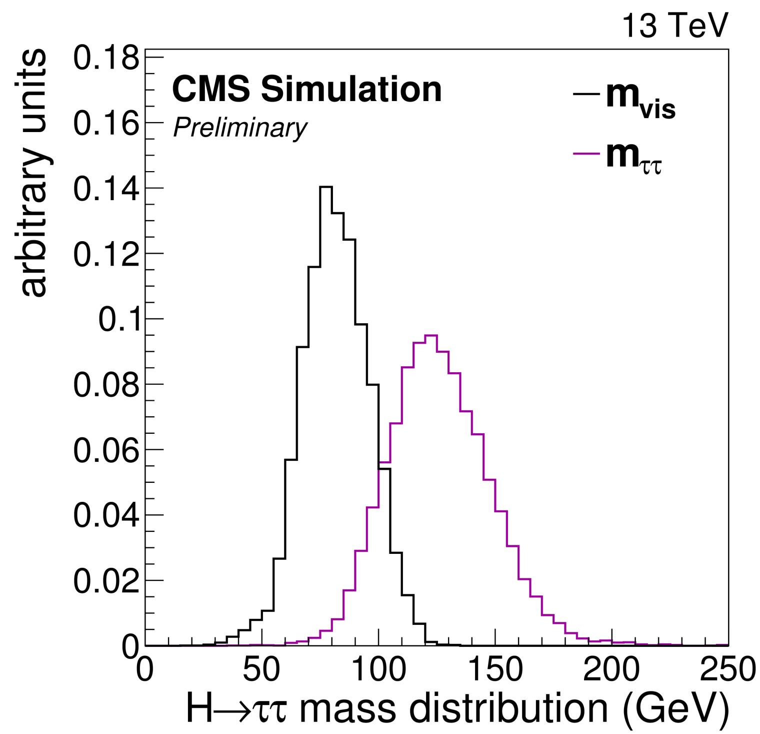
- significant improvement compared to Run 1 approach
- τ_h efficiency of ≃ 45% for mis-ID probability of ≃ $2 \cdot 10^{-3}$

Di-tau Mass Reconstruction

- **Fully reconstructed di-tau mass is key variable discriminating signal against dominant Z → ττ background**
- **Reconstruction of** $m_{\tau\tau}$ **with dynamic likelihood algorithm**
- **Inputs :** $\vec{p}_{\tau_1}, \ \vec{p}_{\tau_2}, \ \vec{p}_{\text{mis}}, \ \text{cov}(\vec{p}_{\text{mis}})$
- **Estimate of** $m_{\tau\tau}$ **is obtained by maximizing likelihood combining**
  - **matrix elements of tau decays**
  - χ^2 **of** $\vec{p}_{\text{mis}}$ **measurement**



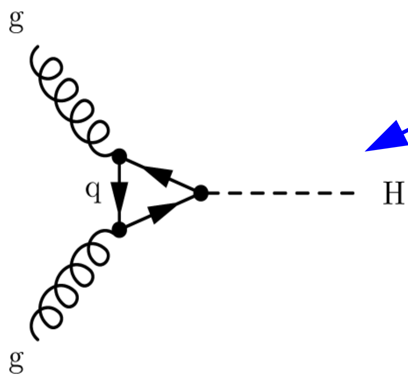
- **Better separation of H → ττ signal and Z → ττ background compared to the invariant mass of visible τ decay products**
  - **the peak position is shifted to the nominal value of resonance mass**
  - **mass resolution : 15-20%**



Event categorization

- Major production mechanisms are targeted with specific event categories

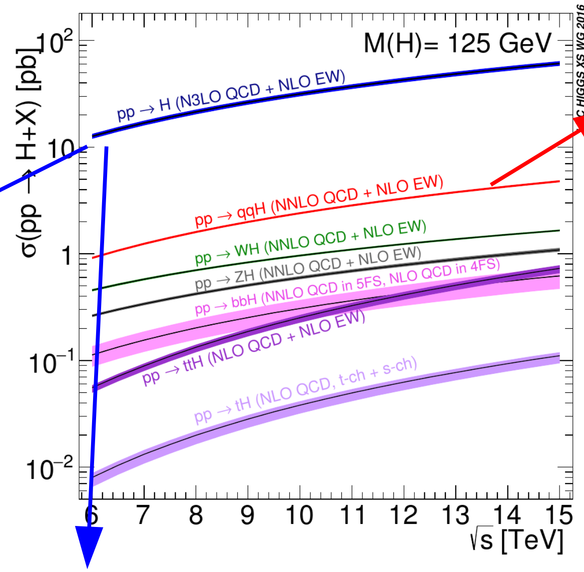
0-jet : no jets with p_T > 30 GeV, |η|<4.7



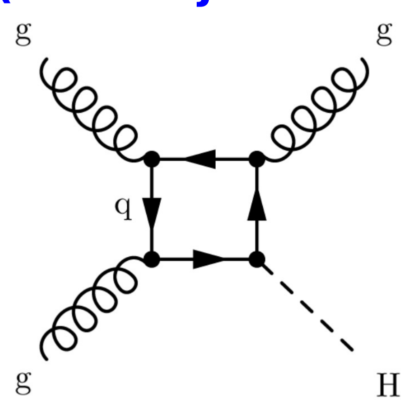
- largest signal yield
- but also largest bkgd

- calibration of tau reconstruction with Z → ττ standard candle

- constrain uncertainties related to instrumental noise



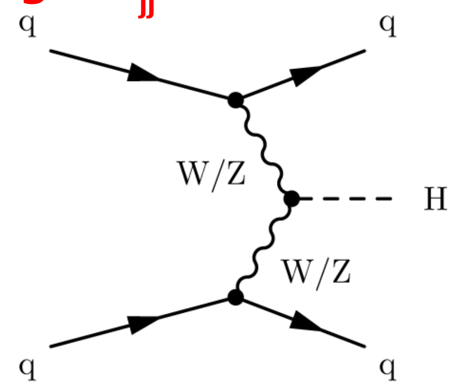
boosted : high p_T(H) (not in 0jet or VBF)



- enhanced S/B compared to 0-jet category

- improved di-tau mass resolution

VBF : two jets with high m_{jj}



- Small signal yield

- but highest S/B ratio

Overview of search channels : $\tau_h\tau_h$

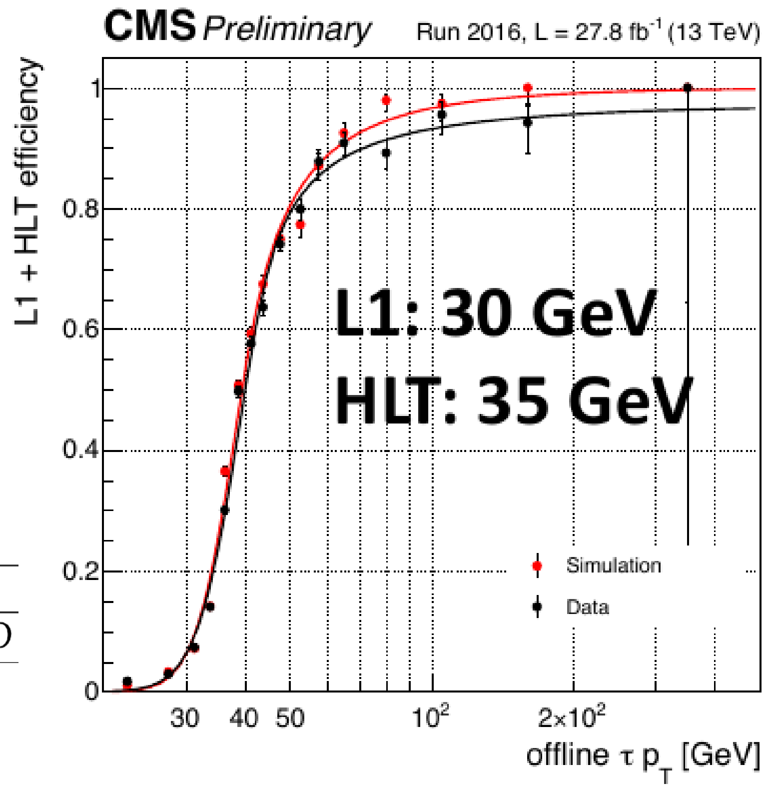
- **largest branching fraction of 42%**
- **Triggering is major a challenge**
  - **Improvements in L1 and HLT**

    – higher readout granularity

    – sophisticated ID using dynamic clustering at hardware level

    → same p_T thresholds as in Run1 but faster turn-on and higher efficiency

Channel	Trigger requirement	Lepton selection		
		p_T (GeV)	η	Isolation
$\tau_h\tau_h$	$\tau_h(35)$ & $\tau_h(35)$	$p_T^{\tau_h} > 50$ & 40	$\lvert\eta^{\tau_h}\rvert < 2.1$	MVA τ_h ID



- **Main backgrounds**
  - **Irreducible Z → ττ**

    → estimated with simulation with data-driven corrections (kinematics of Z and accompanying jets) derived from Z → μμ control region
  - **QCD multijet background (suppressed by tight tau ID)**

    → measured exclusively from data (extrapolation from sideband with loose tau ID)

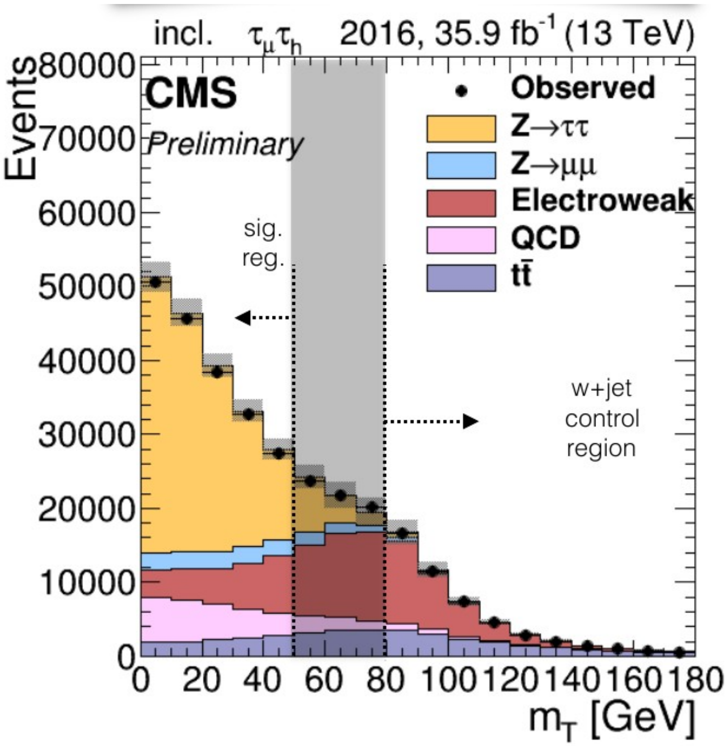
Overview of search channels : $\mu+\tau_h$ and $e+\tau_h$

- **lower branching fractions (23% +23%) but cleaner signature**

  single-lepton triggers with higher p_T thresholds compared to Run1

  $\mu+\tau_h$ channel : combination of single-muon and muon+tau cross trigger → increase in acceptance

Final state	Trigger requirement	Lepton selection		
		p_T (GeV)	η	Isolation
$\mu\tau_h$	$\mu(22)$	$p_T^\mu > 23$ $p_T^{\tau_h} > 30$	$\lvert\eta^\mu\rvert < 2.1$ $\lvert\eta^{\tau_h}\rvert < 2.3$	$I^\mu < 0.15$ MVA τ_h ID
	$\mu(19)$ & $\tau_h(21)$	$20 < p_T^\mu < 23$ $p_T^{\tau_h} > 30$	$\lvert\eta^\mu\rvert < 2.1$ $\lvert\eta^{\tau_h}\rvert < 2.3$	$I^\mu < 0.15$ MVA τ_h ID
$e\tau_h$	$e(25)$	$p_T^e > 26$ $p_T^{\tau_h} > 30$	$\lvert\eta^e\rvert < 2.1$ $\lvert\eta^{\tau_h}\rvert < 2.3$	$I^e < 0.1$ MVA τ_h ID



- **Major backgrounds :**
  - **Irreducible Z → ττ (estimated as in $\tau_h\tau_h$ channel)**
  - **W+Jets with jet → τ_h fakes**

    - suppressed by m_T cut

    - estimated from high m_T sideband
  - **QCD multijets (estimated from sideband regions with same sign lepton pairs and relaxed lepton ID)**
  - **Z → ee/μμ with e/μ → τ_h fakes (estimated from simulation corrected for e/μ → τ_h misidentification rates measured in data)**

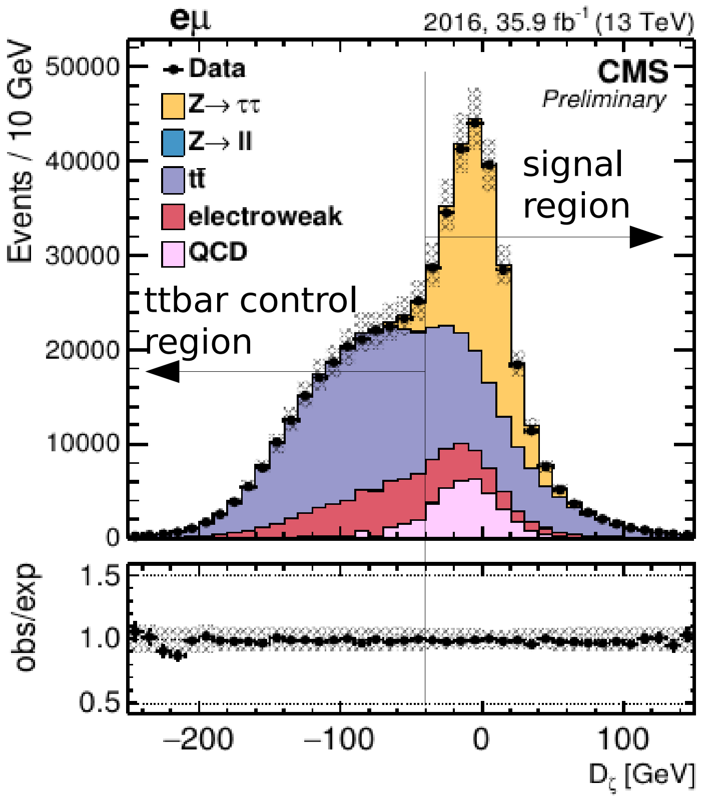
Overview of search channels : e+μ

- **Cleanest signature but small branching fraction (6%) → lowest sensitivity**

  inter-calibration of Drell-Yan background w/o τ_h

  e+μ cross triggers with asymmetric thresholds are used → high signal acceptance

Channel	Trigger requirement	Lepton selection		
		p_T (GeV)	η	Isolation
eμ	e(12) & μ(23)	$p_T^e > 13$ $p_T^\mu > 24$	$\|\eta_e\| < 2.5$ $\|\eta_\mu\| < 2.4$	$I^e < 0.15$ $I^\mu < 0.2$
	e(23) & μ(8)	$p_T^e > 24$ $p_T^\mu > 15$	$\|\eta_e\| < 2.5$ $\|\eta_\mu\| < 2.4$	$I^e < 0.15$ $I^\mu < 0.2$



- **Major backgrounds :**
  - **Irreducible Z → ττ (estimated as in other channels)**
  - **top-pairs**

    - suppressed by b-tag veto and requiring alignment of missing p_T with visible decay tau products (e, μ)

    - estimated from simulation corrected for top pT distribution

    - constrained in the sideband region
  - **QCD multijets (estimated from sideband regions with same sign lepton pairs and relaxed lepton isolation)**

Signal Extraction

- **Signal is extracted by simultaneous maximum-likelihood fit in 12 signal channels**

  4 final states (eμ , eτ_h $\mu\tau_h$, $\tau_h\tau_h$) x 3 event category (0-jet, VBF, Boosted)

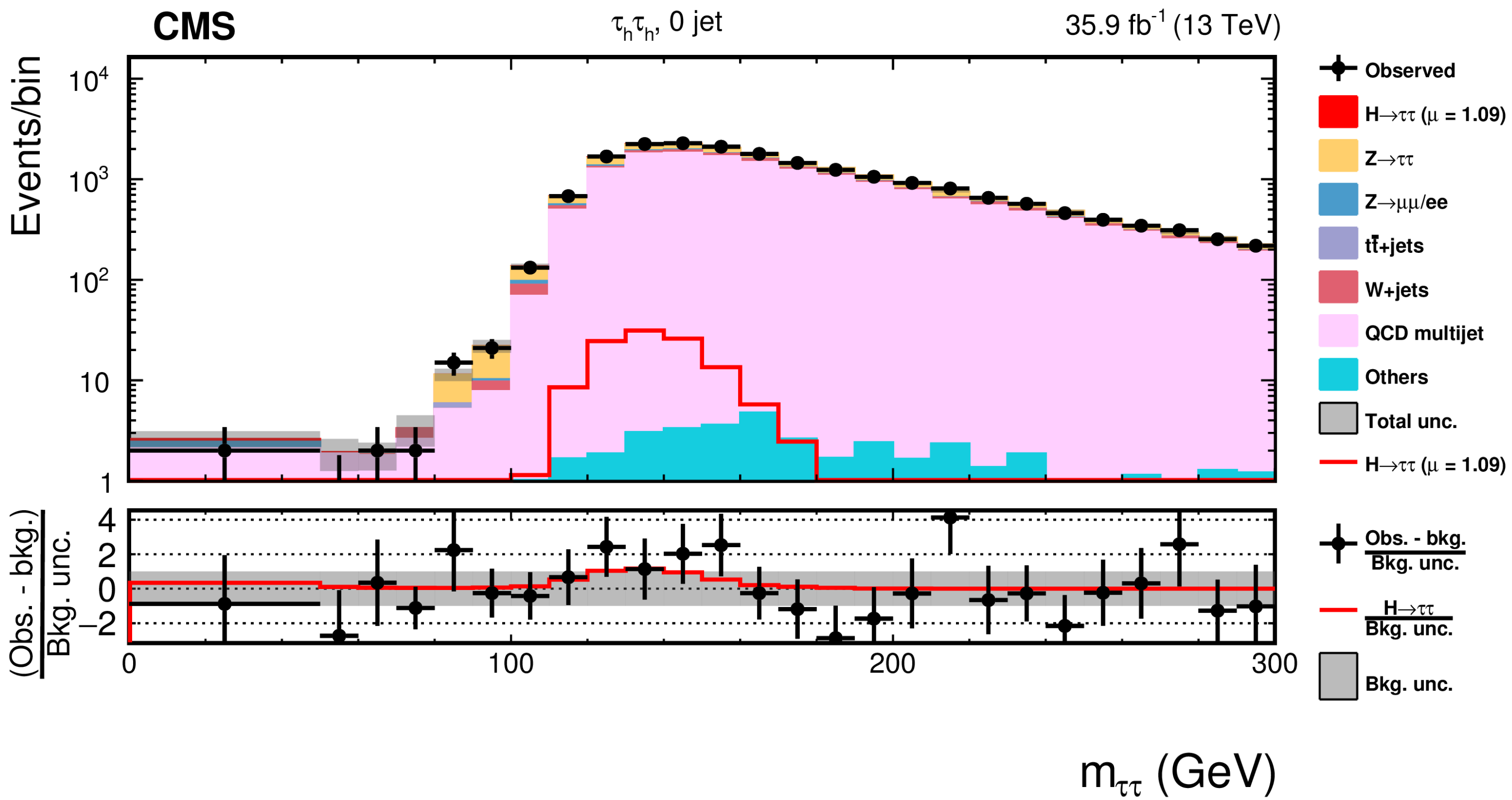
  and 12 control regions, constraining W+Jets, QCD and ttbar backgrounds
- **In all channels but one fit is performed with 2D distributions**

	0-jet	VBF	Boosted
		Selection	
$\tau_h\tau_h$	No jet	$\geq$2 jets, $p_T^{\tau\tau} > 100\,\text{GeV}$, $\Delta\eta_{jj} > 2.5$	Others
$\mu\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300\,\text{GeV}$, $p_T^{\tau\tau} > 50\,\text{GeV}$, $p_T^{\tau_h} > 40\,\text{GeV}$	Others
eτ_h	No jet	$\geq$2 jets, $m_{jj} > 300\,\text{GeV}$, $p_T^{\tau\tau} > 50\,\text{GeV}$	Others
eμ	No jet	2 jets, $m_{jj} > 300\,\text{GeV}$	Others
		Observables used in fit	
$\tau_h\tau_h$	$m_{\tau\tau}$	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$\mu\tau_h$	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
eτ_h	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
eμ	p_T^{μ}, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$

- **Di-tau mass or visible mass is used as 1st observable in 2D fit**
- **Choice of 2nd observable is motivated by**
  - **production signatures probed in a given category (boosted, VBF)**
  - **calibration of backgrounds specific to a given decay channel (0-jet)**

Postfit final discriminants : $\tau_h\tau_h$ channel

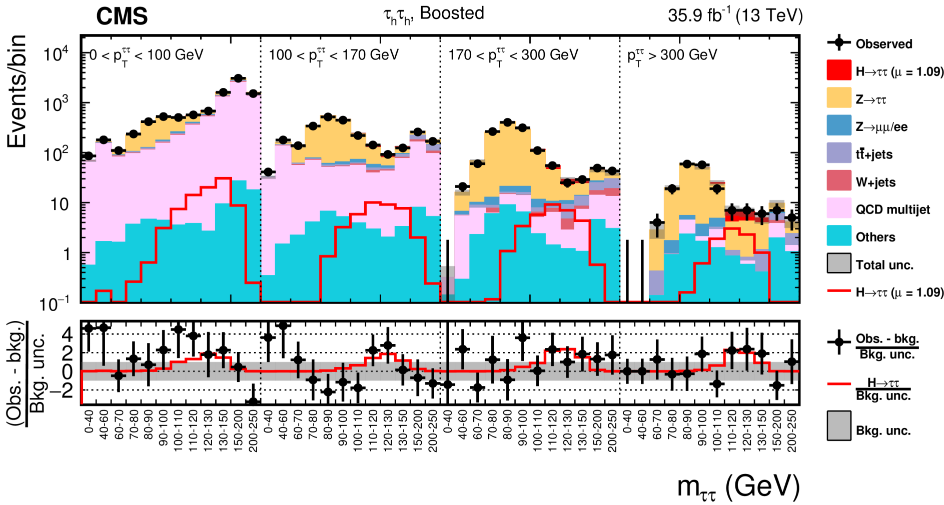
- **1D $m_{\tau\tau}$ distribution used to extract signal in 0-jet category**
- **No gain by adding second variable**



- **Moderate excess above background prediction at $m_{\tau\tau}$ ~ 125 GeV**

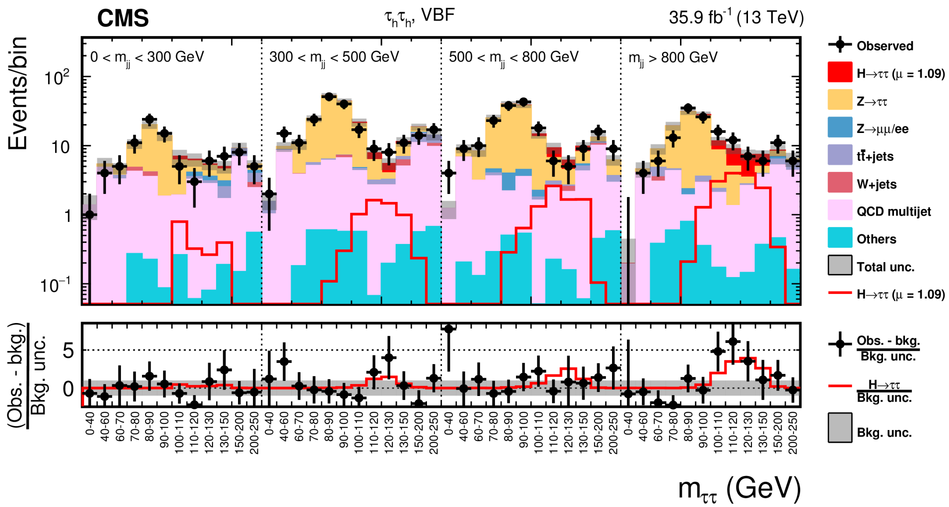
Postfit final discriminants : $\tau_h\tau_h$ channel

- **Unrolled 2D distributions in Boosted and VBF categories**



- **Boosted category**

$$(m_{\tau\tau}, p_T^{\tau\tau})$$



- **VBF category**

$$(m_{\tau\tau}, m_{jj})$$

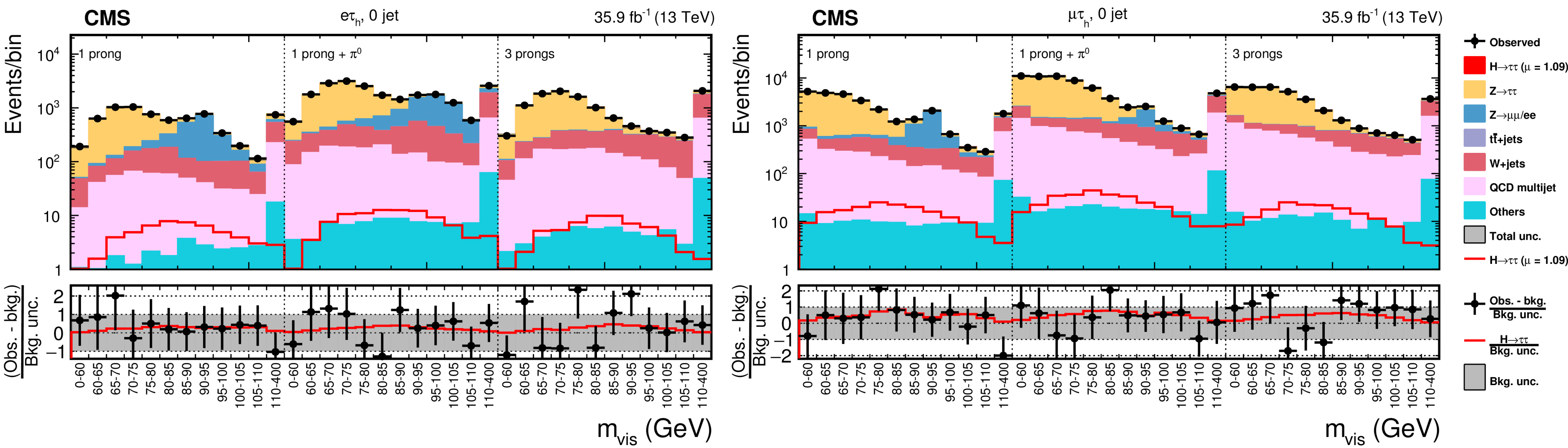
excellent S/B at high m_{jj}
(most sensitive channel)

- Signal clearly developing in bins with high S/B ratio

Postfit final discriminants : $\ell+\tau_h$ channels

- **the 2D distribution (m_{vis} vs. τ_h decay mode) used to extract signal in 0-jet category**

  → better separation between signal and Z → ee/μμ background



- **Low sensitivity to signal**
- **But category facilitates calibration of Drell-Yan background and constrains instrumental uncertainties**

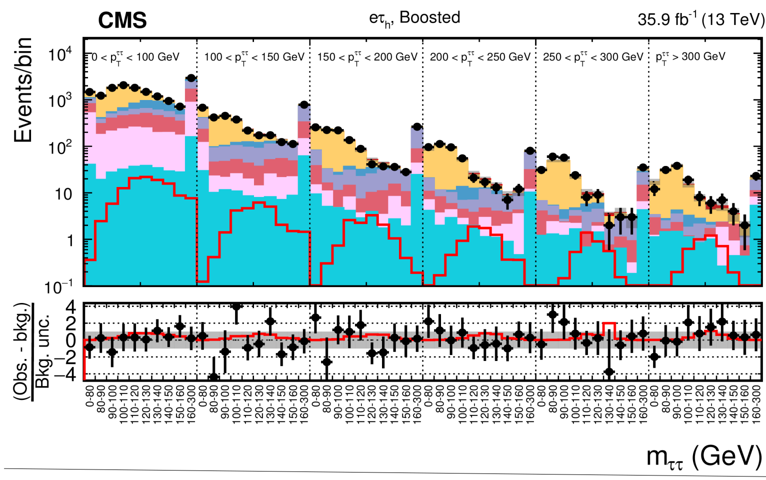
  → τ_h ID efficiency and momentum scale

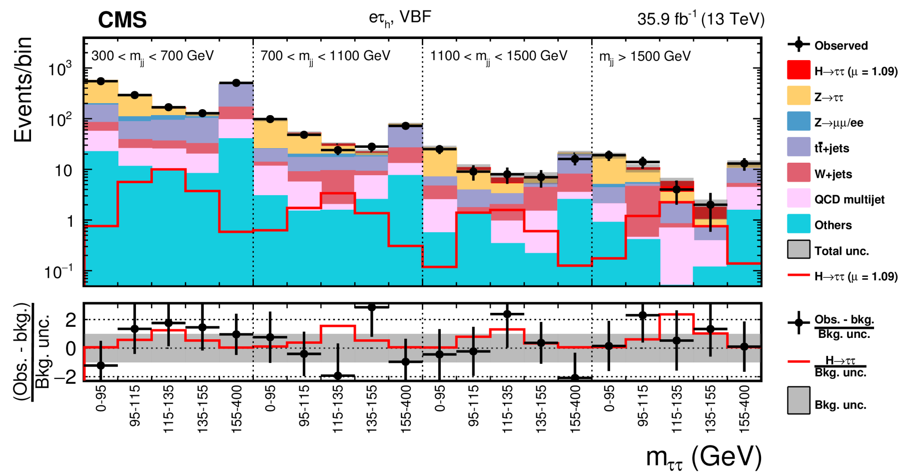
  → e/μ → τ_h fake rate and momentum scale

Postfit final discriminants : $\ell+\tau_h$ channels

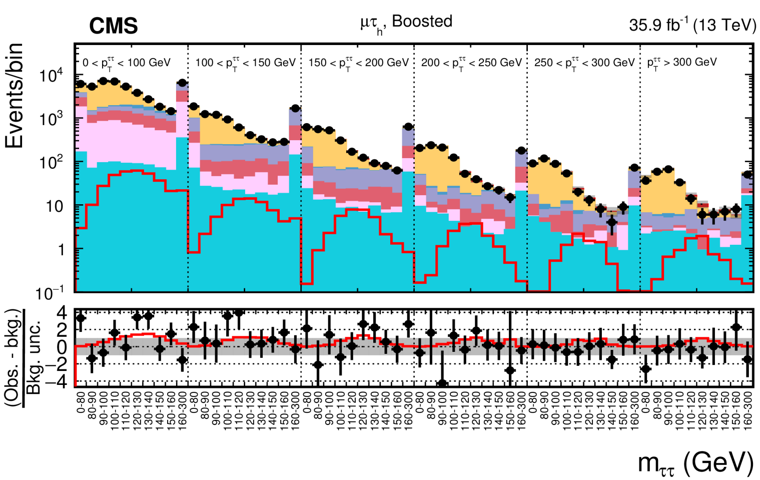
- **Unrolled 2D distributions**

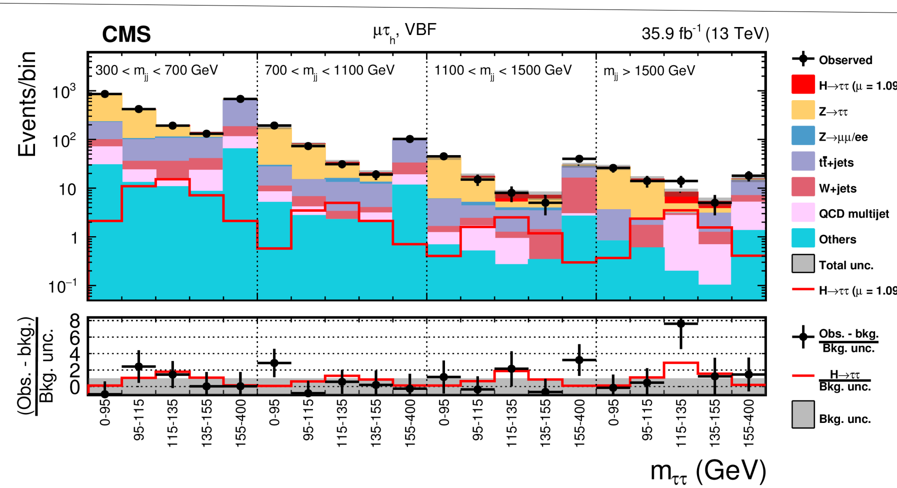
$$\text{Boosted} : (m_{\tau\tau}, p_T^{\tau\tau}) \qquad \text{VBF} : (m_{\tau\tau}, m_{jj})$$





$e\tau_h$



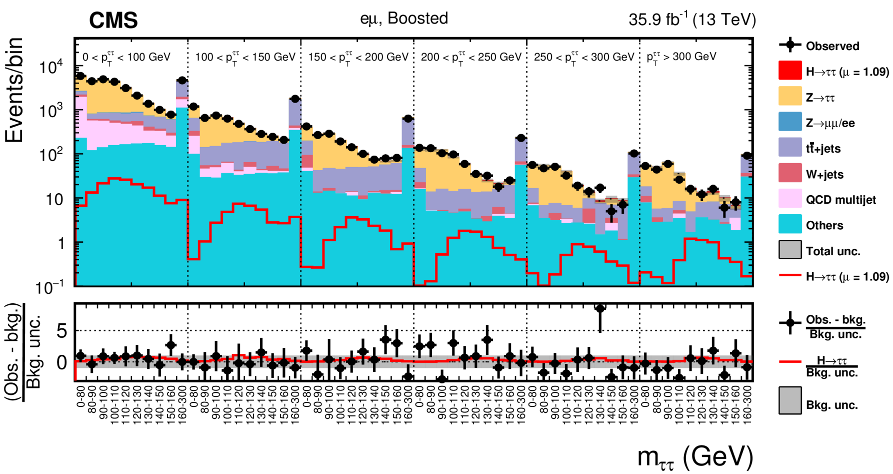


$\mu\tau_h$

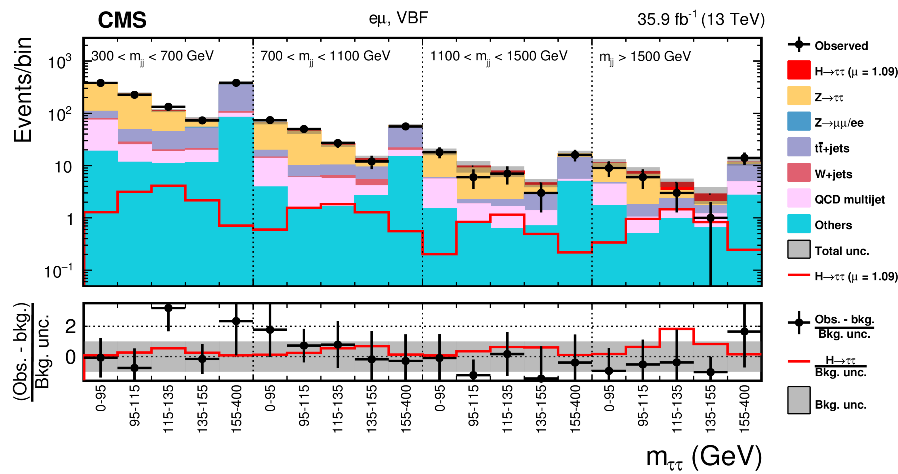
Postfit final discriminants : e+μ channel

- **Unrolled 2D distributions**

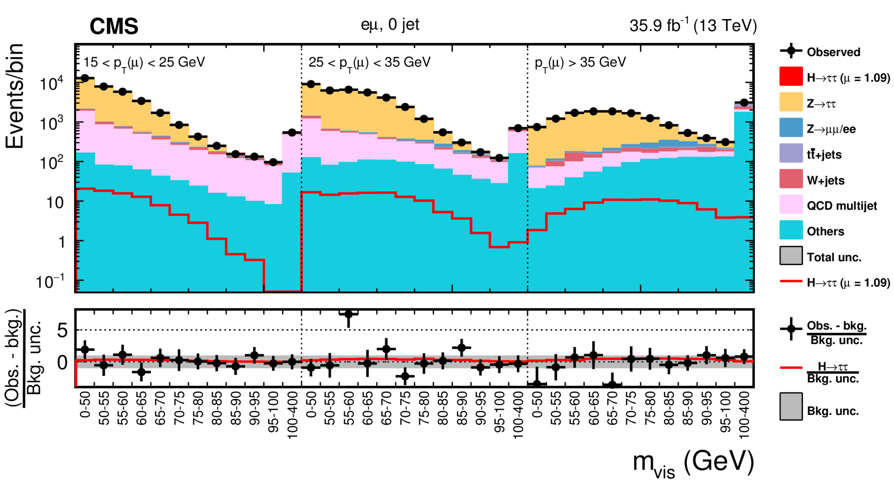
Boosted : $(m_{\tau\tau}, p_T^{\tau\tau})$



VBF : $(m_{\tau\tau}, m_{jj})$



0-jet : $(m_{\text{vis}}, p_T^{\mu})$



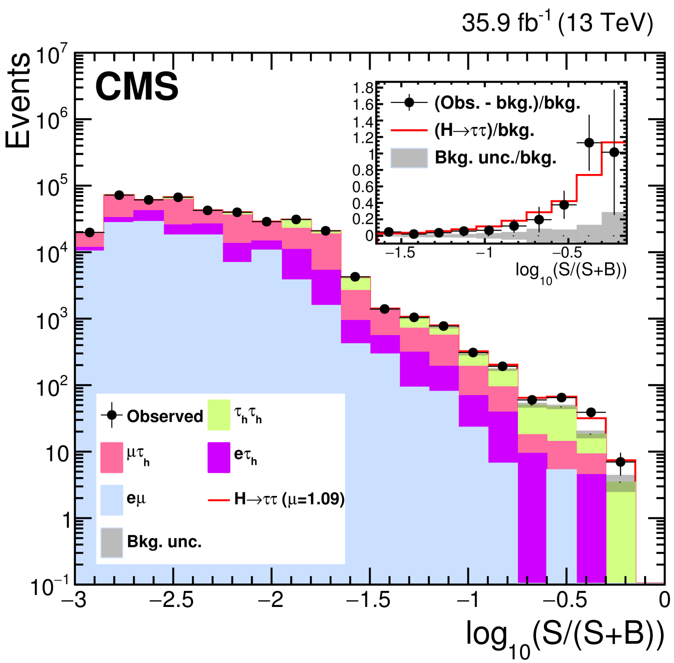
e+μ channel has lowest sensitivity
- smallest yield
- poorer mass resolution

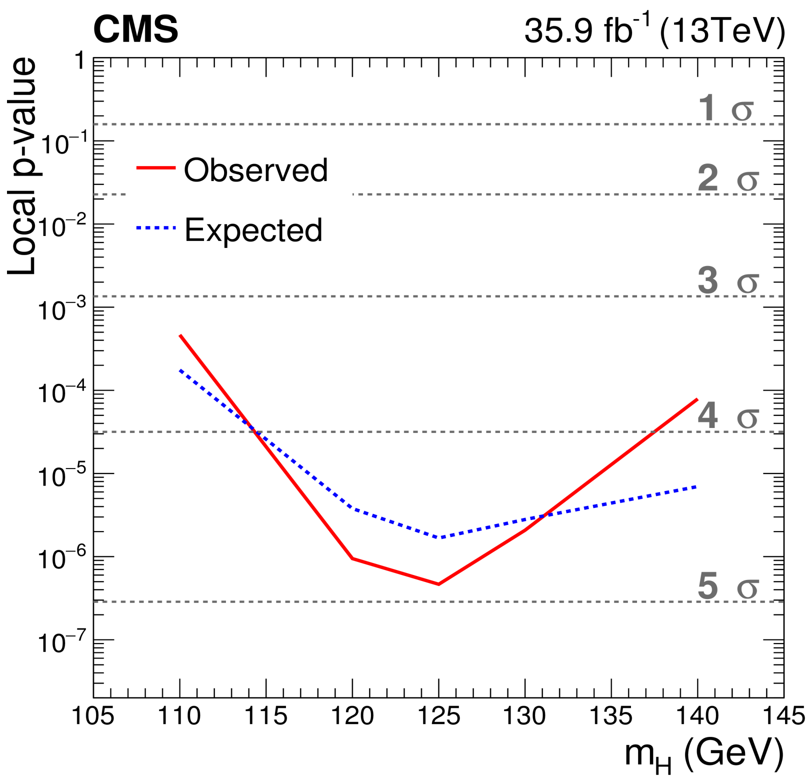
Nonetheless, non-negligible contribution to overall sensitivity

Observation of H → ττ Decays

- **Distribution of event yield in the analysis bins ordered by S/(S+B)**

→ **clearly visible excess in data w.r.t. background-only expectation**



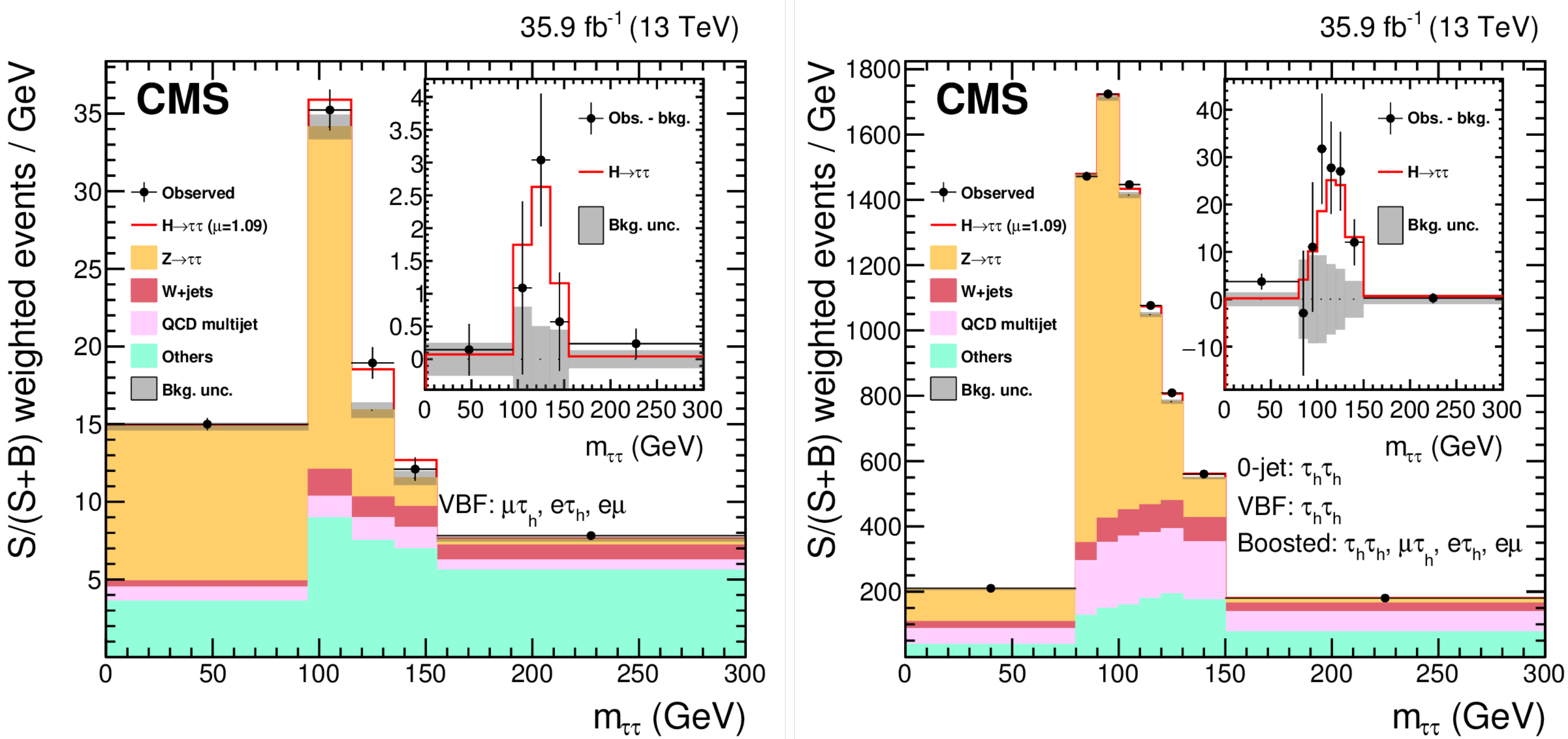


- **obs. (exp.) significance at m_H = 125 GeV**
- $4.9\sigma \;\; (4.7\sigma)$ **with Run II data only**

- **Combination with Run I CMS data yields**
  $5.9\sigma \;\; (5.9\sigma)$

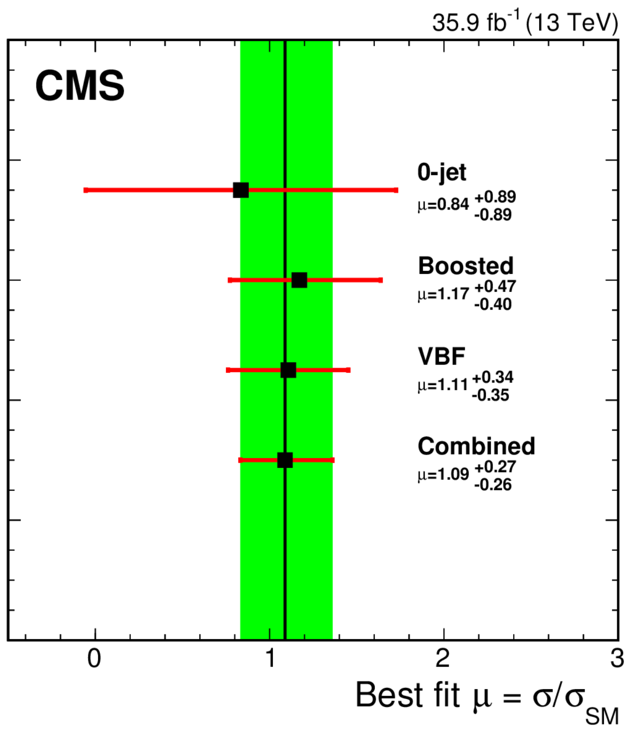
- **first observation of Yukawa coupling in single fermionic decay channel at CMS**

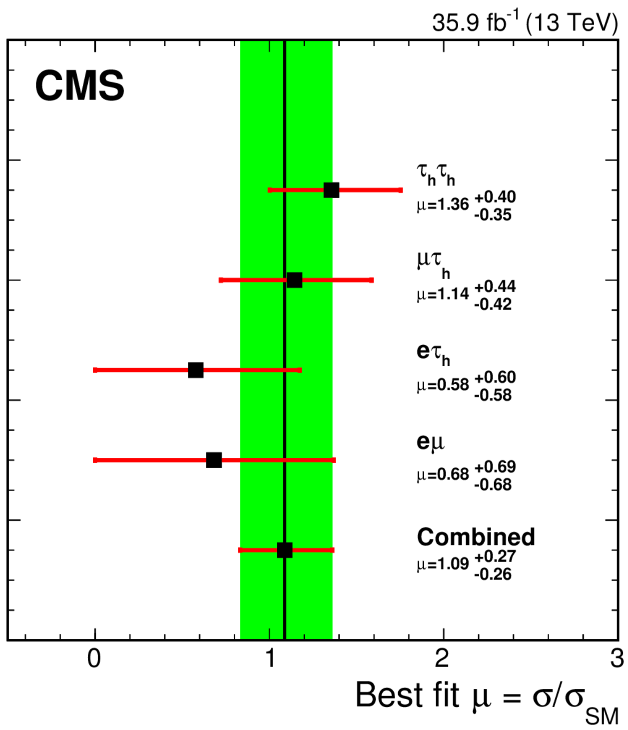
Visualization of signal : $m_{\tau\tau}$ distribution



- **Events are weighted by S/(S+B) in bins of second variable of 2D distributions → unbiased mass spectrum**
- **Signal is clearly visible in the distribution of physical observable $m_{\tau\tau}$**

Measurement of the signal strength





- Measurements are consistent across channels
- Highest sensitivity comes from
  - decay side : $\tau_h\tau_h$
  - production side : VBF

Measurement of couplings

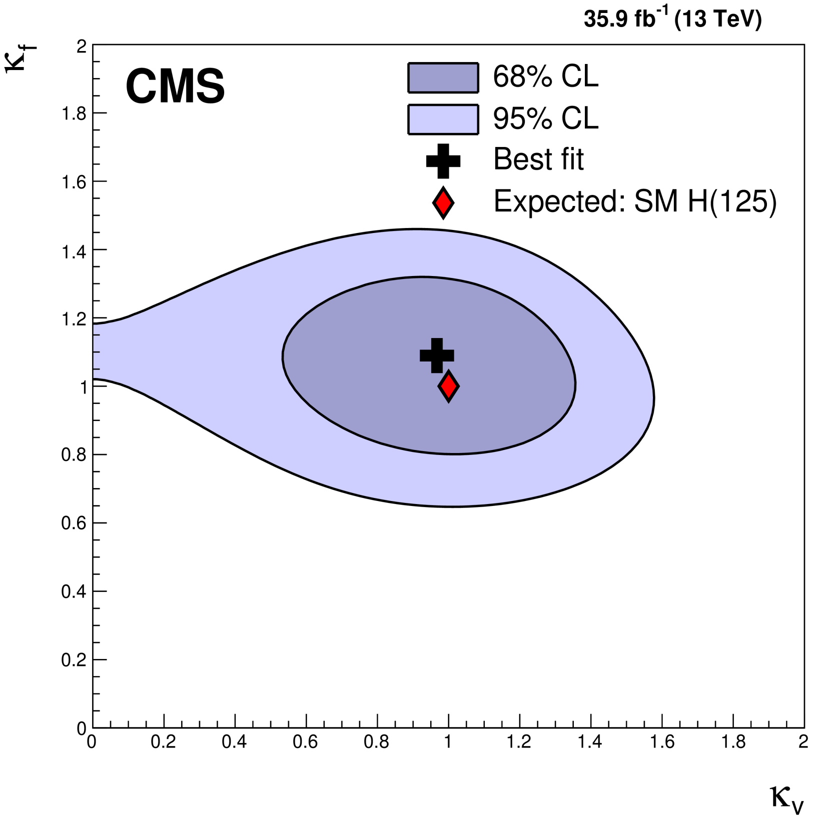
- probing universal coupling modifiers

$$\kappa_f = \frac{g_{Hf\bar{f}}}{g^{SM}_{Hf\bar{f}}}$$

affects gg → H and ttH production rates and H → bb/ττ decay rates

$$\kappa_V = \frac{g_{HVV}}{g^{SM}_{HVV}}$$

affects VBF and VH production rates and H → VV decay rates



- Contribution of VH is added but not targeted with specific category

- Contribution from H → WW (significant in eμ channel, sub-dominant in other channels) is treated as signal

- Measurement of couplings is compatible with SM expectation

Summary

- **H(125) decays to τ-leptons are studied at CMS using Run II data**

- **H(125) → ττ decay is observed with statistical significance of ~5.9σ combining data collected at 7, 8 and 13 TeV**
  - **first observation of Yukawa coupling in a single experiment and single decay channel**

- **measured H(125) properties in the H → ττ decay channel are consistent at current precision level with SM expectations**

- **Paper is submitted to PLB and made available in hep archive**

  arXiv:1708.00373 [hep-ex]

- **larger dataset is expected by end of 2017**

  → better measurement precision

  → new measurements possible, e.g. probing CP properties

Backup : Uncertainty model

- Refined uncertainty model
  - split jet energy uncertainties into 27 various sources
  - fine-grane instrumental uncertainties (τ_h Id, momentum scale, fake rate)
- Background yields and shape constrained in a dedicated control regions
- instrumental corrections constrained in 0-jet category and propagated to VBF and Boosted categories
- Most of uncertainties are constrained in global fit
- Largest impact on precision comes from
  → instrumental uncertainties
  → theoretical shape uncertainties
  → limited MC statistics

Source of uncertainty	Prefit	Postfit (%)
τ_h energy scale	1.2% in energy scale	0.2–0.3
e energy scale	1–2.5% in energy scale	0.2–0.5
e misidentified as τ_h energy scale	3% in energy scale	0.6–0.8
μ misidentified as τ_h energy scale	1.5% in energy scale	0.3–1.0
Jet energy scale	Dependent upon p_T and η	—
$\vec{p}_T^{miss}$ energy scale	Dependent upon p_T and η	—
τ_h ID & isolation	5% per τ_h	3.5
τ_h trigger	5% per τ_h	3
τ_h reconstruction per decay mode	3% migration between decay modes	2
e ID & isolation & trigger	2%	—
μ ID & isolation & trigger	2%	—
e misidentified as τ_h rate	12%	5
μ misidentified as τ_h rate	25%	3–8
Jet misidentified as τ_h rate	20% per 100 GeV τ_h p_T	15
$Z \to \tau\tau/\ell\ell$ estimation	Normalization: 7–15%	3–15
	Uncertainty in $m_{\ell\ell/\tau\tau}$, $p_T(\ell\ell/\tau\tau)$, and m_{jj} corrections	—
W + jets estimation	Normalization (eμ, $\tau_h\tau_h$): 4–20%	—
	Unc. from CR (eτ_h, $\mu\tau_h$): $\simeq$5–15	—
	Extrap. from high-m_T CR (eτ_h, $\mu\tau_h$): 5–10%	—
QCD multijet estimation	Normalization (eμ): 10–20%	5–20%
	Unc. from CR (eτ_h, $\tau_h\tau_h$, $\mu\tau_h$): $\simeq$5–15%	—
	Extrap. from anti-iso. CR (eτ_h, $\mu\tau_h$): 20%	7–10
	Extrap. from anti-iso. CR ($\tau_h\tau_h$): 3–15%	3–10
Diboson normalization	5%	—
Single top quark normalization	5%	—
$t\bar{t}$ estimation	Normalization from CR: $\simeq$5%	—
	Uncertainty on top quark p_T reweighting	—
Integrated luminosity	2.5%	—
b-tagged jet rejection (eμ)	3.5–5.0%	—
Limited number of events	Statistical uncertainty in individual bins	—
Signal theoretical uncertainty	Up to 20%	—