

Search for non-resonant Higgs bosons pairs production in the $bb\tau\tau$ final state at CMS

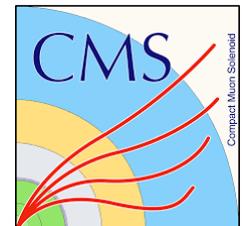
Jaime León Holgado (CIEMAT)

on behalf of the CMS Collaboration

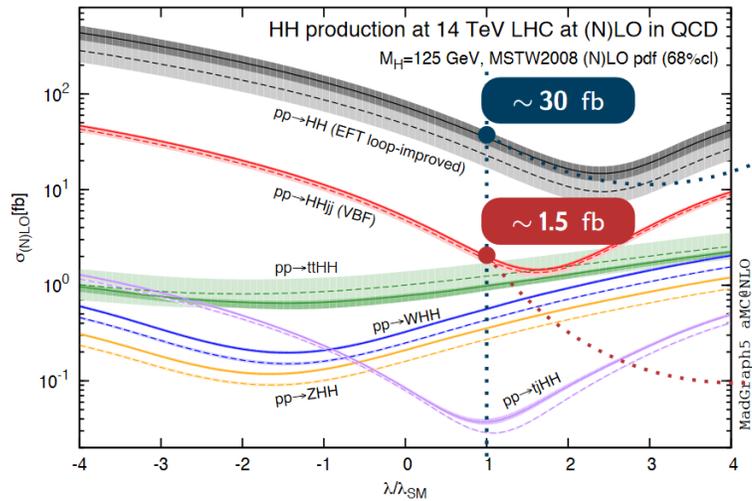
SM@LHC2022, April 12th, 2022



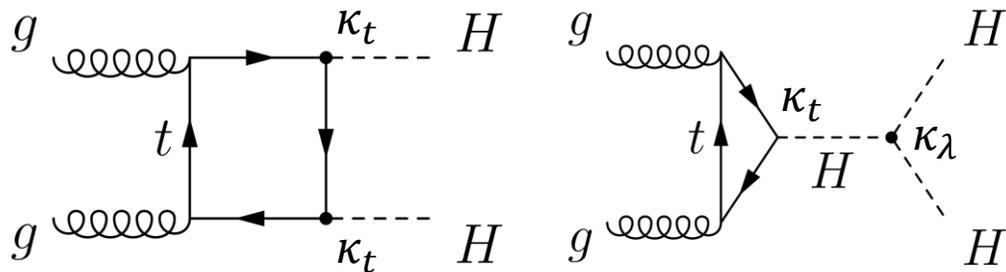
Ciemat
Centro de Investigaciones
Energéticas, Medioambientales
y Tecnológicas



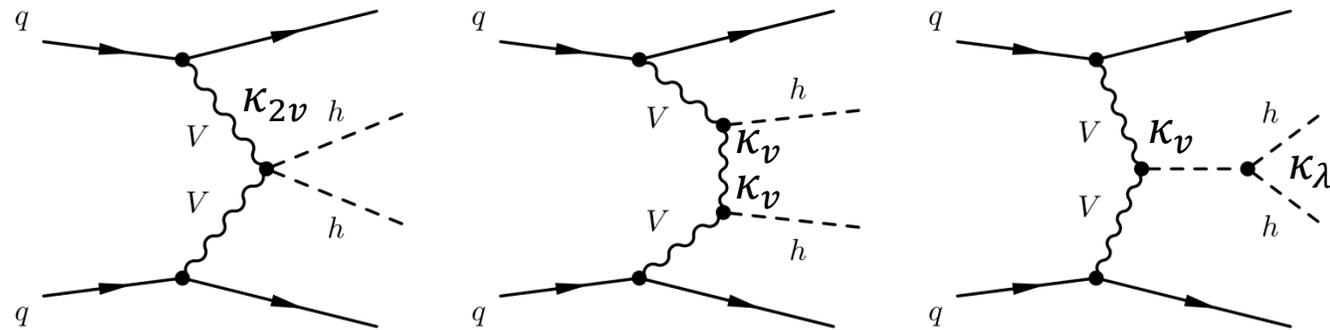
Search for Higgs Boson Pair Production at the LHC



- Several CMS analysis search for HH production using all Run 2 data (138 fb^{-1}).
- Studying main production modes: Gluon and Vector Boson Fusion.
 - $\sigma_{ggF} \sim 20 \times \sigma_{VBF}$
 - 2 additional forward jets (with high invariant mass) in the VBF production mode.
- Analysis objectives:
 - Obtain the upper limit on $r = \sigma/\sigma_{SM}$.
 - Constrain the values of the Higgs couplings involved.



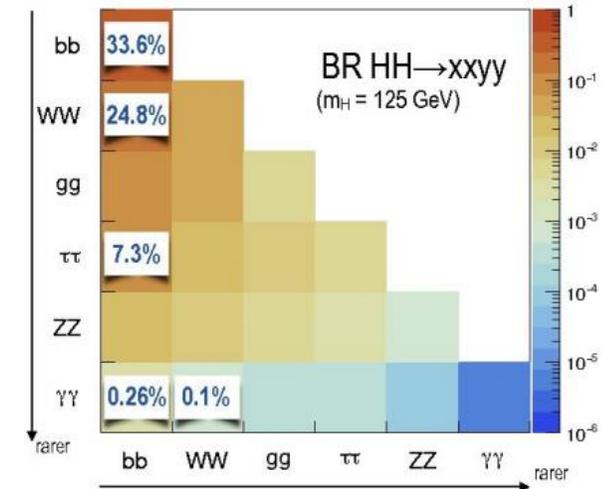
Gluon Fusion (ggF)



Vector Boson Fusion (VBF)

Search for Higgs Boson Pair Production at the LHC

- Studying decay $HH \rightarrow bb\tau\tau$: Medium branching ratio ($\sim 7.3\%$) and relatively low background.
- Considering the fully-hadronic ($\tau_h\tau_h$), and semi-leptonic ($\tau_e\tau_h$, $\tau_\mu\tau_h$) final states ($\sum BR(\tau\tau \rightarrow \tau_i\tau_h) = 88\%$).



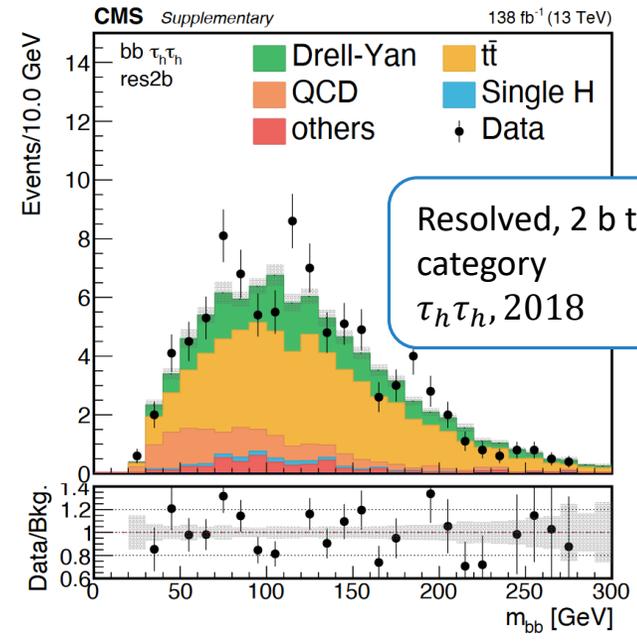
- Previous CMS $HH \rightarrow bb\tau\tau$ result: [PLB 778 \(2018\) 101](#)
- Using 2016 data: 35.0 fb^{-1}
- 95% CL upper limit on $\sigma(gg \rightarrow HH)$
obs (exp) ~ 30 (25) $\times \sigma_{SM}$

- Latest results: [CMS-PAS-HIG-20-010](#)
- Presented already at [Moriond EWK](#).

Datasets

DATA

138 fb⁻¹ of pp collision data collected by the CMS detector at the LHC during Run 2 (2016-2018).



OTHER BACKGROUNDS

Purely MC based

W + jets	Single top (tW and t channels)
Di-boson	EWK Z and W
Tri-boson	ttV and ttVV
Single SM Higgs (ggF, VBF, VH, ttH)	

MAIN BACKGROUNDS

Drell-Yan

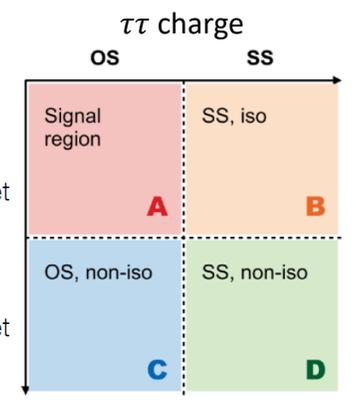
Basic estimation from MC, yield corrected by fitting in $Z \rightarrow \mu\mu$ control regions

$t\bar{t}$

Shape: MC
Yield: Data-driven from $t\bar{t}$ enriched control region

QCD

Shape and yield: Fully data-driven using ABCD method
 τ isolation using DeepTau



Analysis flow

Trigger requirements

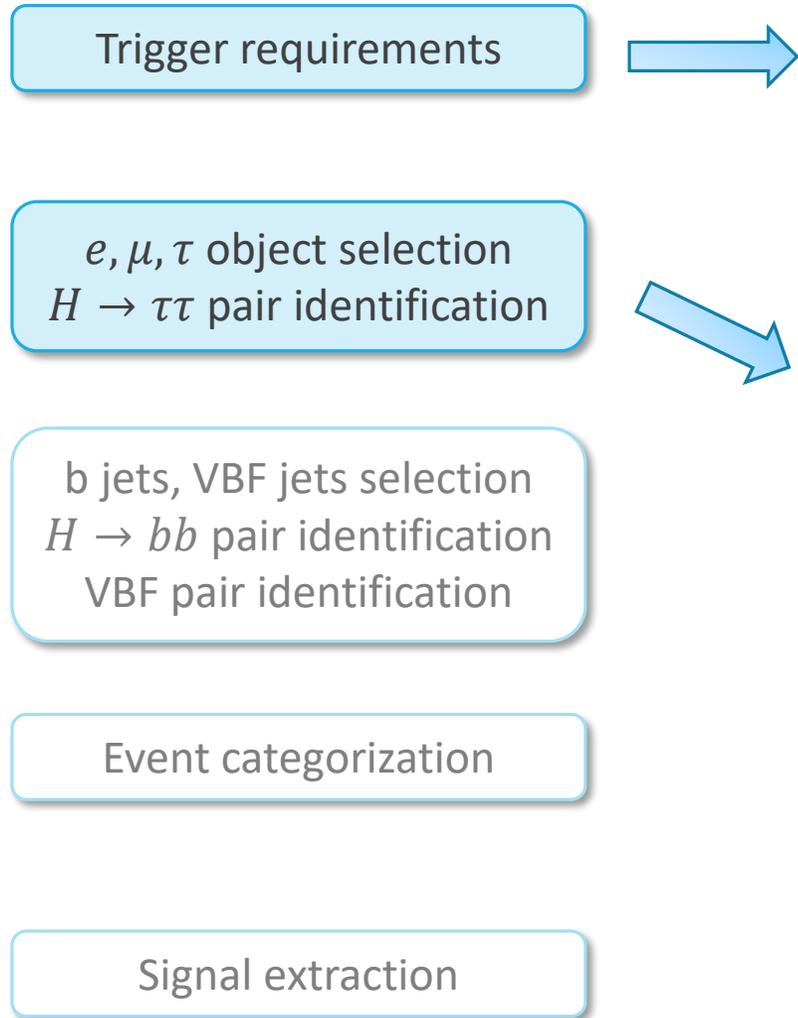
e, μ, τ object selection
 $H \rightarrow \tau\tau$ pair identification

b jets, VBF jets selection
 $H \rightarrow bb$ pair identification
VBF pair identification

Event categorization

Signal extraction

Analysis flow



Targetting $H \rightarrow \tau\tau$ decay products:

- $\tau_\mu\tau_h$ channel: single muon, muon- τ_h cross trigger
- $\tau_e\tau_h$ channel: single electron, electron- τ_h cross trigger
- $\tau_h\tau_h$ channel: di- τ_h trigger, dedicated VBF $H \rightarrow \tau_h\tau_h$ trigger.

- Match offline to trigger objects
- Apply kinematic, isolation and identification requirements.

Analysis flow

Trigger requirements

e, μ, τ object selection
 $H \rightarrow \tau\tau$ pair identification

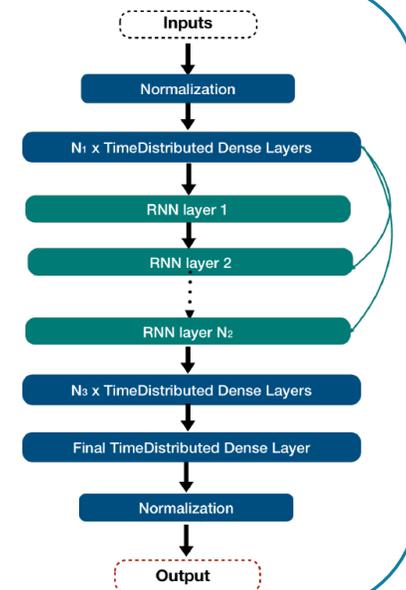
b jets, VBF jets selection
 $H \rightarrow bb$ pair identification
VBF pair identification

Event categorization

Signal extraction



- New ML-based b jet candidates' selection: **HH-btag**
 - Improving $H \rightarrow bb$ selection efficiency and mass resolution w.r.t. CMS standard b-tagging ([DeepJet](#)).
- Select VBF jet candidates (highest m_{jj} pair)



Analysis flow

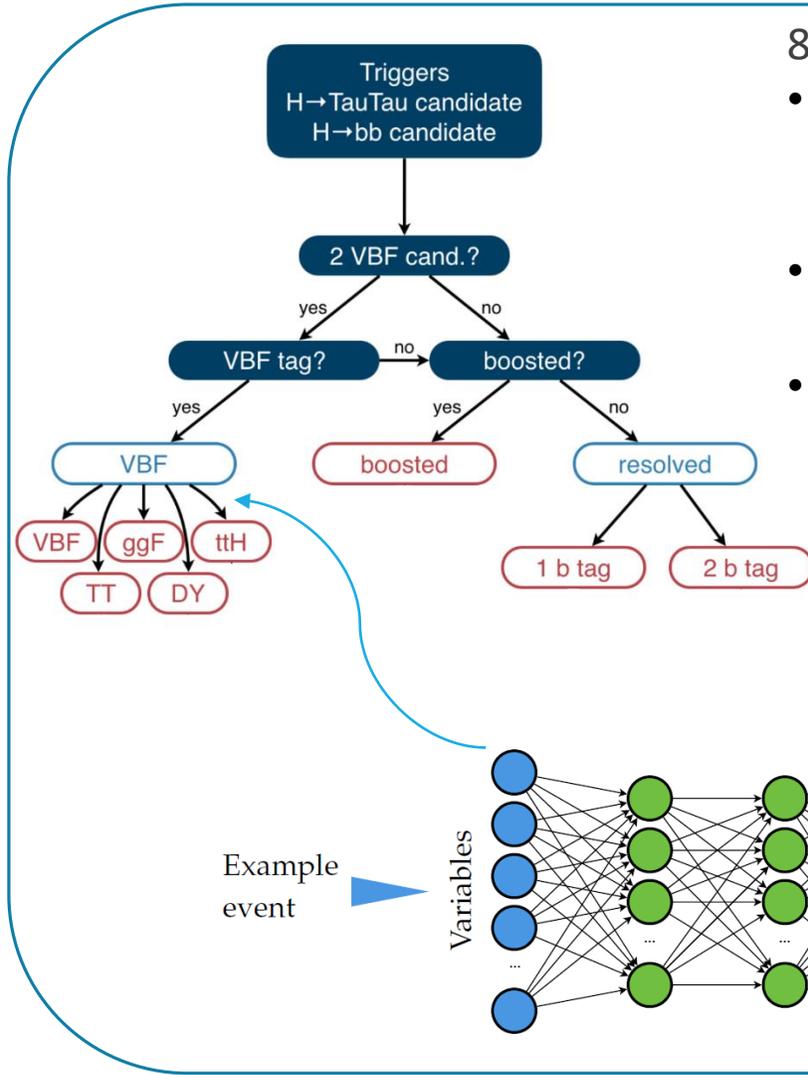
Trigger requirements

e, μ, τ object selection
 $H \rightarrow \tau\tau$ pair identification

b jets, VBF jets selection
 $H \rightarrow bb$ pair identification
 VBF pair identification

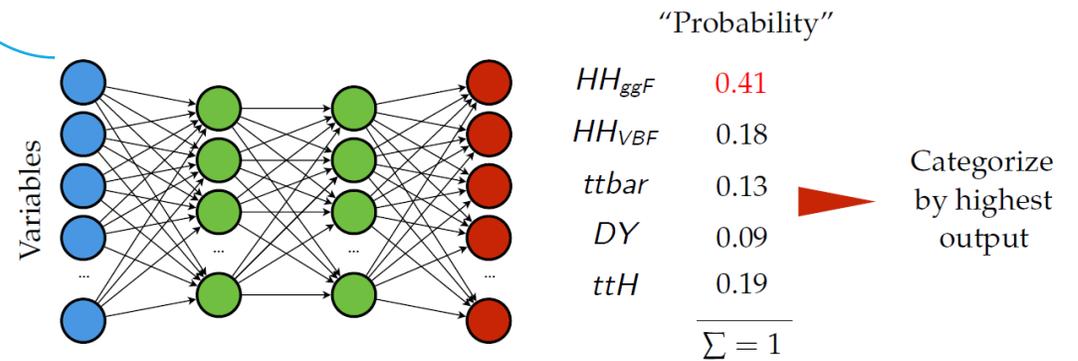
Event categorization

Signal extraction



8 final analysis categories

- **Boosted** category: Requiring a fat jet with 2 loose b-tag subjets matched to the selected b jets.
- Resolved **1 b tag** and **2 b tag**: one or two jets passing medium b-tag WP.
- VBF categories: $m_{jj} > 500$ GeV and $|\Delta\eta_{jj}| > 3$, medium b-tag requirement on at least one b jet candidate. **Five subcategories** identified using a multi-classifier.



Analysis flow

Trigger requirements

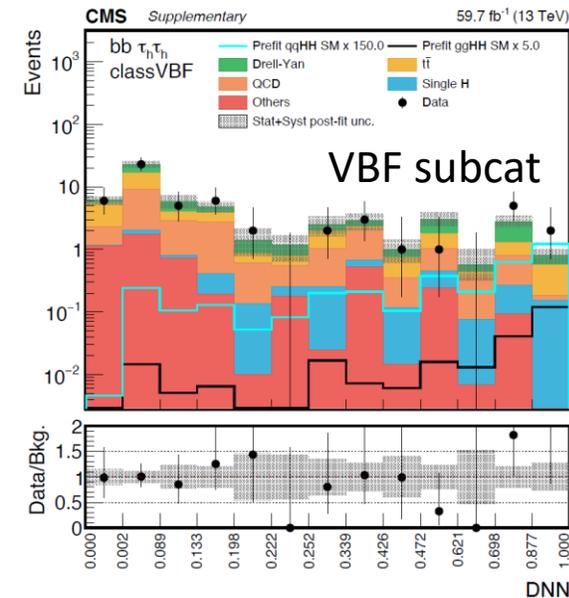
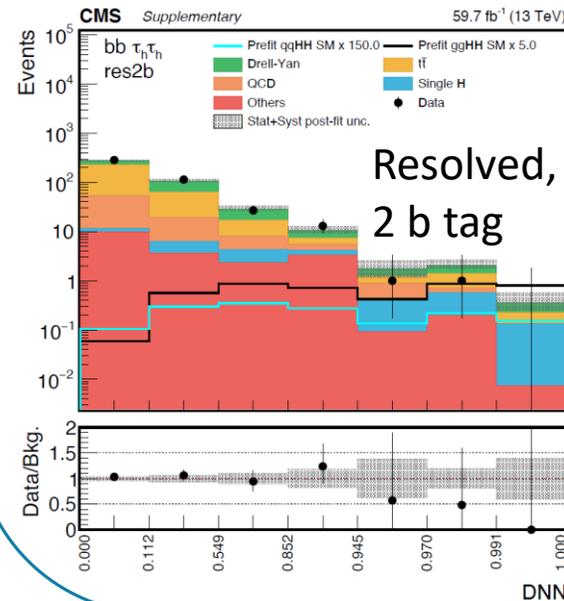
e, μ, τ object selection
 $H \rightarrow \tau\tau$ pair identification

b jets, VBF jets selection
 $H \rightarrow bb$ pair identification
VBF pair identification

Event categorization

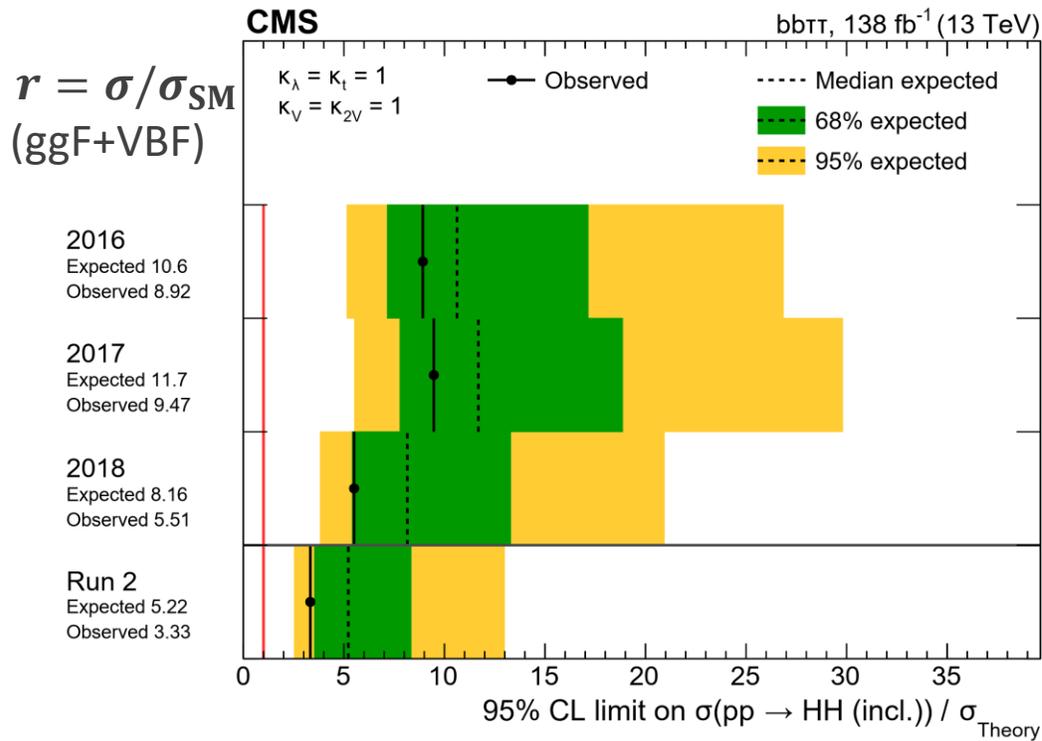
Signal extraction

- Using a dedicated ML discriminator that provides a single prediction to discriminate between signal and background.
- Trained with boosted and resolved events and considering as signal only the ggF samples.
- Binned shape of the output score used in all analysis categories.

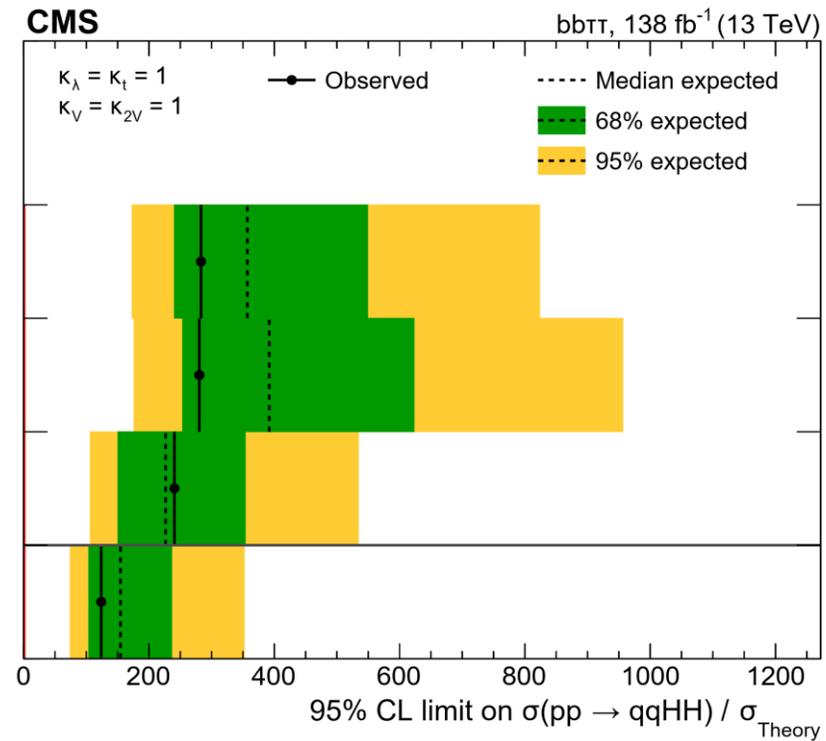


Resolved 2 b
tag and VBF
subcategory
 $\tau_h \tau_h$, 2018

Results: Upper limits @SM



Obs (exp) upper limit on inclusive HH production:
3.3 (5.2) $\times \sigma_{SM}$

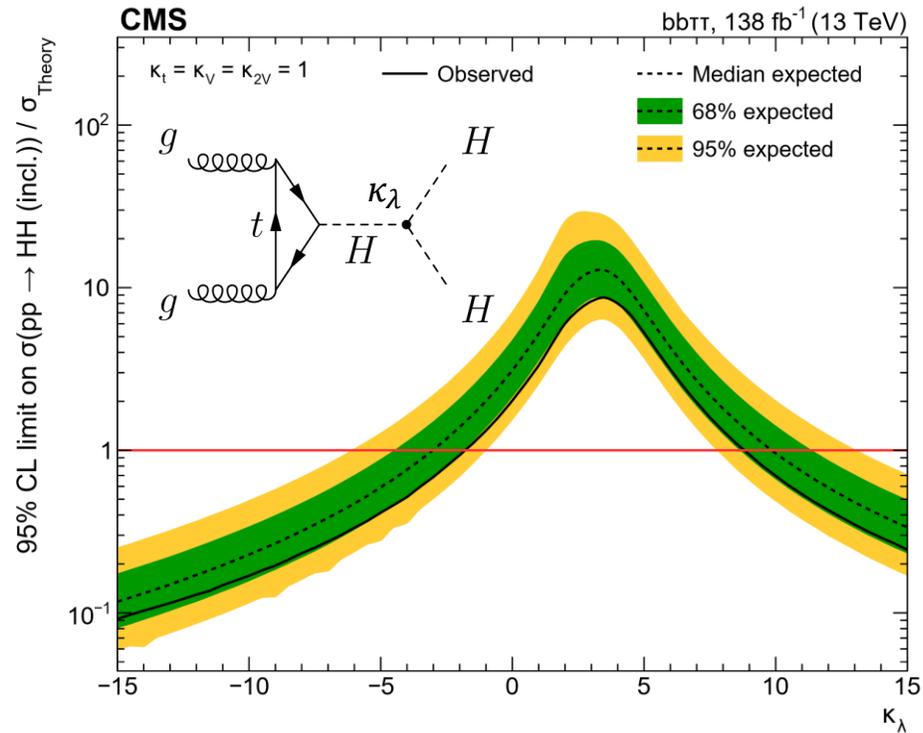


Obs (exp) upper limit on VBF HH production:
124 (154) $\times \sigma_{SM}$

- ggF + VBF: **second most sensitive** CMS HH analysis.
- VBF: **most sensitive** CMS HH analysis.

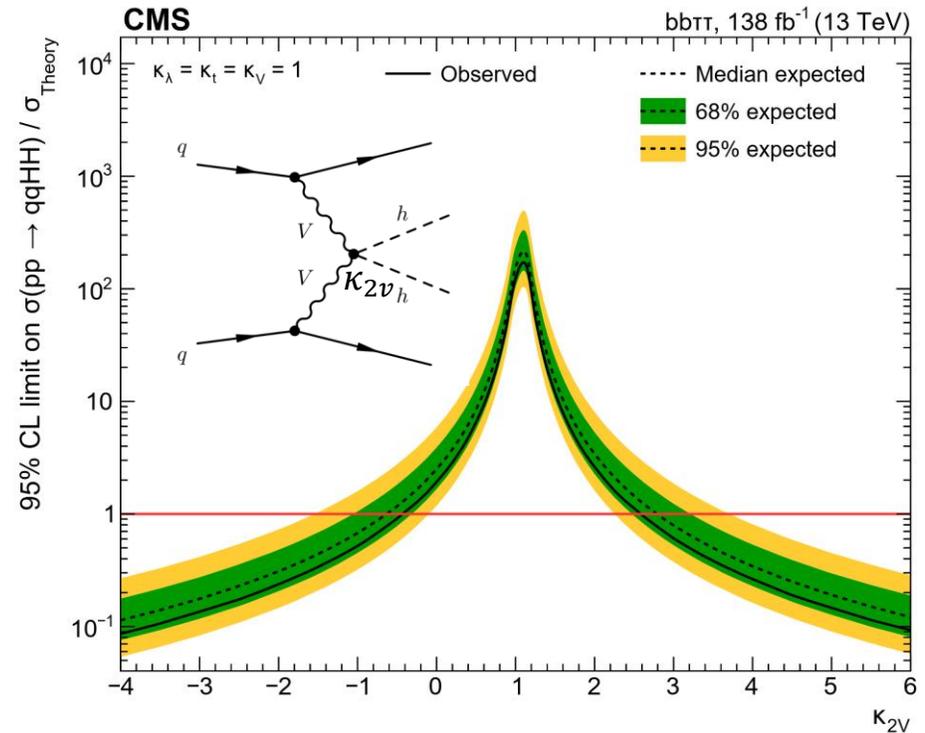
Results: Inclusive HH upper limits vs. κ_λ and $\kappa_{2\nu}$

Limits on r



- Observed: $\kappa_\lambda \in [-1.8, 8.8]$
- Expected: $\kappa_\lambda \in [-3, 9.9]$

Limits on r_{qqHH}



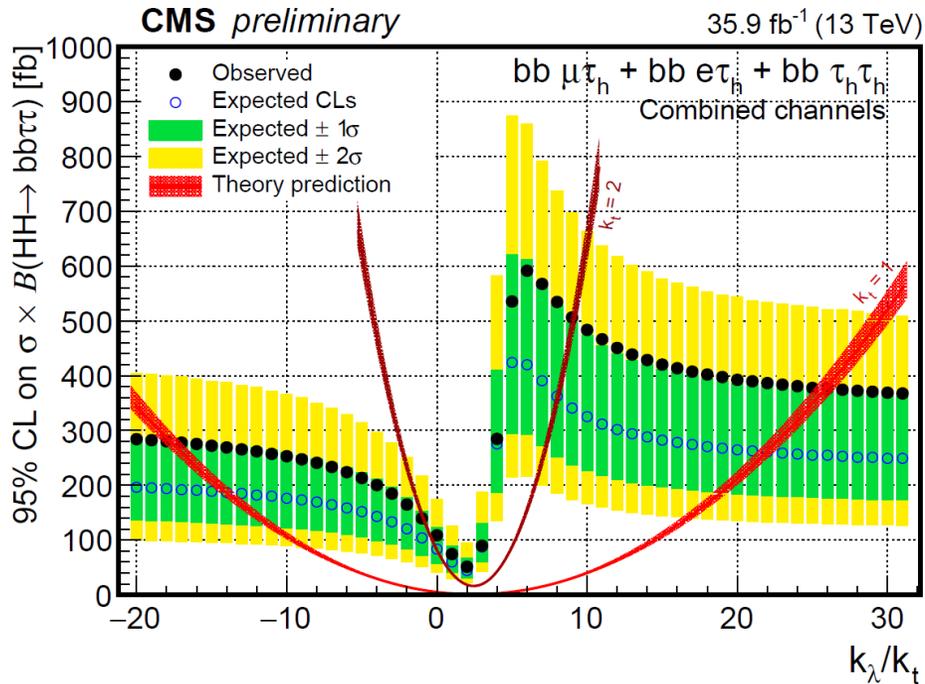
- Observed: $\kappa_{2\nu} \in [-0.4, 2.6]$
- Expected: $\kappa_{2\nu} \in [-0.6, 2.8]$

Conclusions

- Search for non-resonant HH production in the $bb\tau\tau$ final state has been presented.
 - Using the full Run 2 data set (138 fb^{-1}).
 - Studying the main production modes: ggF and VBF.
- Considerable improvement compared to the previous CMS $HH \rightarrow bb\tau\tau$ results due to increased statistics and introduction of advanced Machine Learning techniques at analysis and object identification levels.
- Observed (expected) 95% CL upper limit for the SM point:
 - $\sigma_{\text{ggF+VBF}} = 3.3 (5.2) \times \sigma_{\text{ggF+VBF}}^{\text{SM}}$
 - $\sigma_{\text{VBF}} = 124 (154) \times \sigma_{\text{VBF}}^{\text{SM}}$

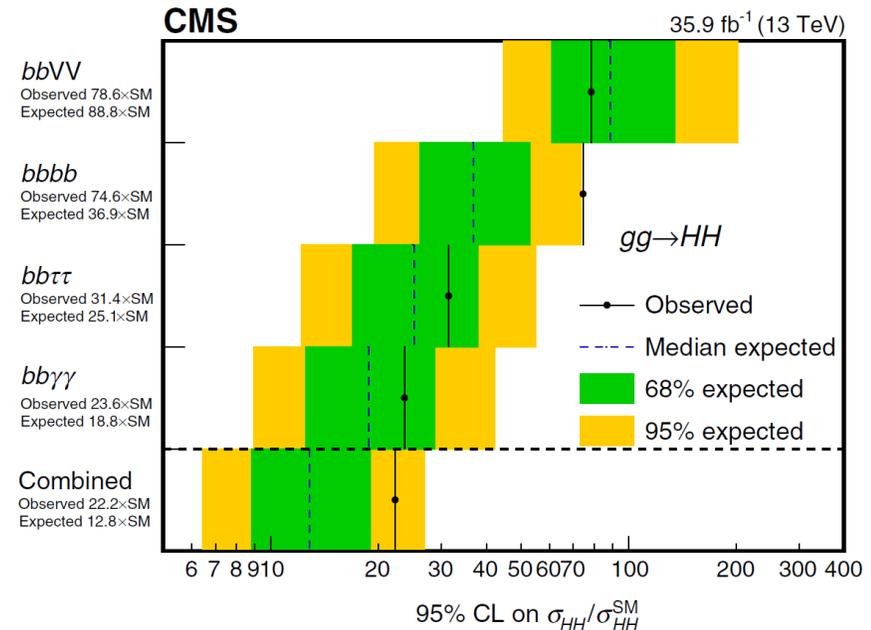
BACKUP

Latest $HH \rightarrow bb\tau\tau$ results (2016 dataset)



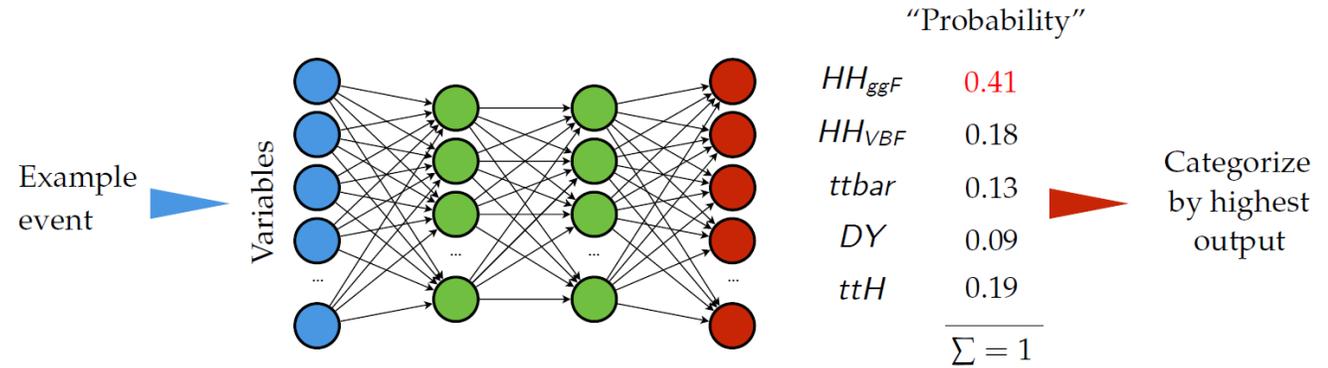
- Most recent $bb\tau\tau$ result: [PLB 778 \(2018\) 101](#)
- Using 2016 data: 35.0 fb^{-1}
- 95% CL upper limit on $\sigma(gg \rightarrow HH)$ obs (exp) ~ 30 (25) $\times \sigma_{SM}$

- Combination of 2016 HH analysis: [Phys. Rev. Lett. 122, 121803](#)
- The $bb\tau\tau$ channel is the second most sensitive HH channel.

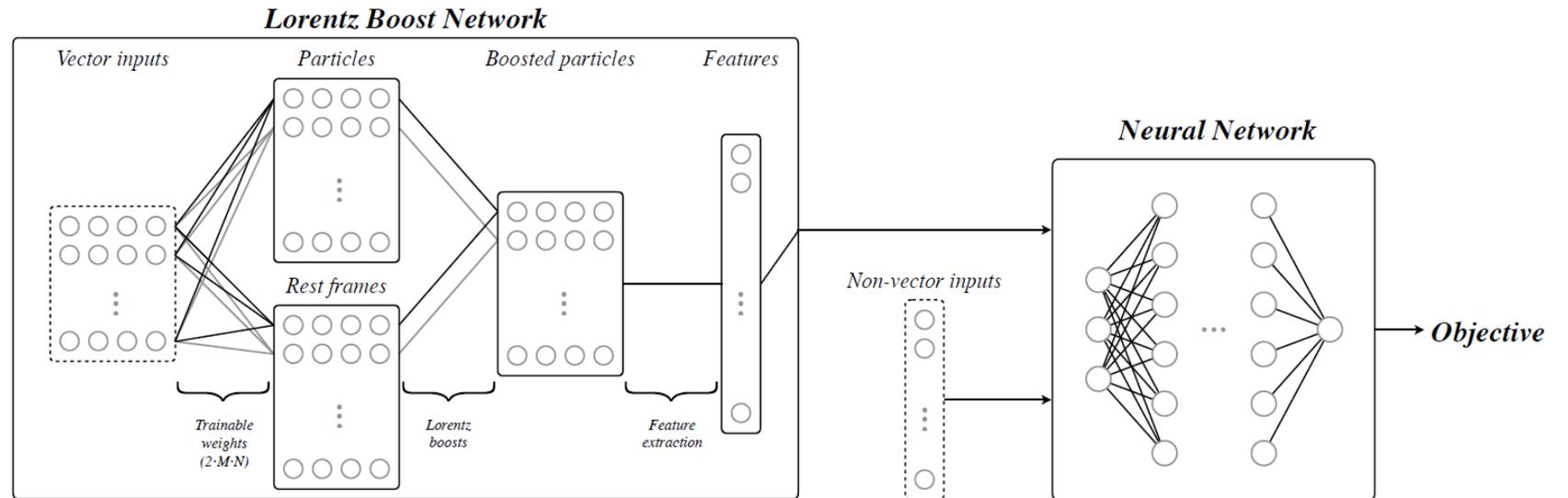


Multi-class classification for VBF categorization

- In order to separate the VBF signal from the ggF and other background contamination, we classify the events in this category into 5 subcategories using a multi-class classification.
- In this strategy, machine learning techniques are used to assign probability estimates for an event to belong to categories associated to any of the relevant physics processes under consideration.

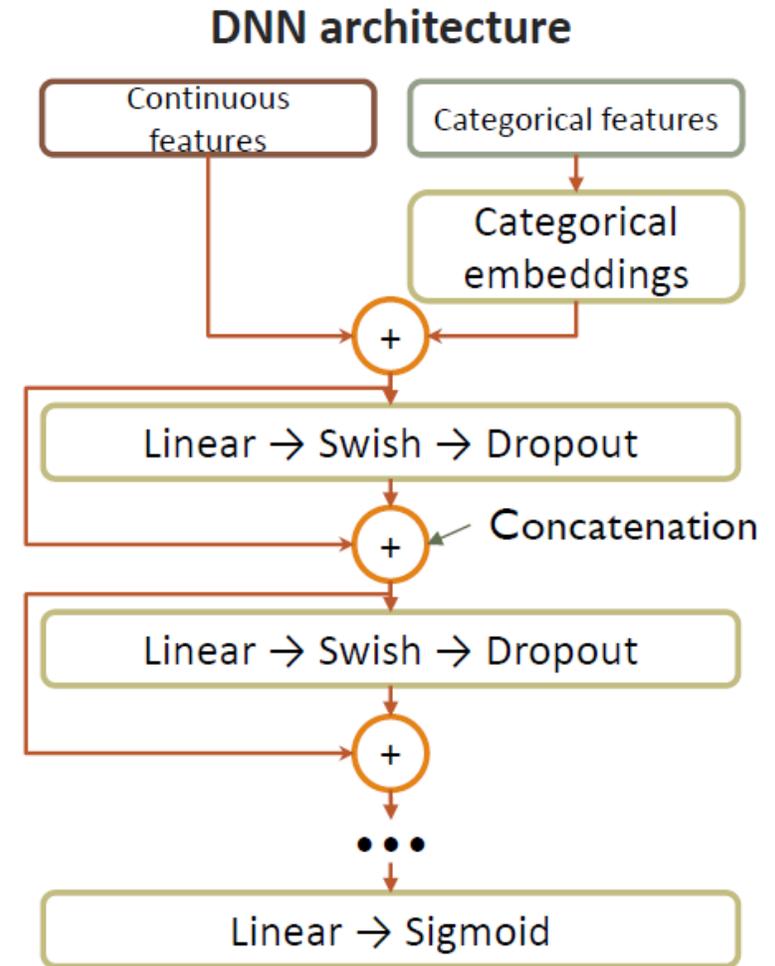


LBN takes N input four-vectors, creates M particles, boosts them to their rest frames and extracts generic features from them and their combinations.



DNN for signal extraction

- A dedicated ML discriminator that provides a binary signal vs background classification is trained.
- Trained with events belonging to the resolved and boosted categories and considering as signal only the ggF LO sample.
- Each discriminator is an ensemble of ten neural networks trained via ten-fold cross validation.
 - The final score is a weighted sum of the DNN scores from the ensemble
- Binned shape of the output score is used for the final signal extraction in all analysis categories.



DNN for signal extraction

- An extensive set of high- and low-level features is considered as possible inputs.
- The most influential inputs are selected using a series of steps:
 - Cleaning of linearly correlated features
 - Iterative cleaning based on feature importance using Random Forest classifiers.
 - Pruning of features with a high mutual dependency among the remaining ones.
- The final NN ensemble is trained on 20 continuous and 6 categorical features.

Some of the most important continuous features
$M(HH)$ (KinFit)
χ^2 (KinFit)
$M(\tau\tau)$ (SVFit)
$\Delta R(\tau, \tau) \cdot p_t(H(\tau\tau)(SVFit))$
$M_T(\tau_1, MET)$
Deepflavour CvsB of the first b jet
...

Categorical features
Channel
Year
Boosted flag
VBF jets presence
Highest deepflavour WP from the first b
Highest deepflavour WP from the second b

Background composition ($\tau_\mu\tau_h$, 2018)

Sample	Boosted	res1b	res2b	classVBF	classGGF	classTT	classDY	classTTH
tt	46.71	18670.46	15577.77	369.8	151.13	4786.5	198.89	1613.41
dy+jets	15.12	6150.25	638.72	86.7	22.77	66.59	87.72	11.21
qcd	0	1978.46	398.11	52.4	0.36	164.78	27.91	0
tth	0.83	19.75	23.43	0.37	0.59	3.92	0.64	10.81
others	13.89	5876.12	783.2	61.57	15.83	312.37	44.56	94.73
BKG SUM	76.55	32695.04	17421.23	570.84	190.68	5334.16	359.72	1730.16
ggHH SM	0.15	0.72	0.8	0.066	0.09	0.012	0.0071	0.022
VBFHH SM	0.0014	0.013	0.011	0.042	0.0037	0.0032	0.0014	0.0018

Background composition ($\tau_e\tau_h$, 2018)

Sample	Boosted	res1b	res2b	classVBF	classGGF	classTT	classDY	classTTH
tt	32.57	11466.95	9418.7	224.51	91.95	3031.34	132.82	1111.3
dy+jets	14.16	5080.67	474.66	65.45	15.85	69.61	83.49	16.27
qcd	0	743.47	0	0	0	50.05	10.05	0
tth	0.53	12.34	14.72	0.20	0.36	2.50	0.41	7.05
others	12.82	3387.06	680.1	26.24	12.14	205.79	30.84	65.1
BKG SUM	60.08	20690.49	10588.18	316.4	120.3	3359.29	257.61	1199.72
ggHH SM	0.15	0.72	0.8	0.066	0.09	0.012	0.0071	0.022
VBFHH SM	0.0014	0.013	0.011	0.042	0.0037	0.0032	0.0014	0.0018

Background composition ($\tau_h\tau_h$, 2018)

Sample	Boosted	res1b	res2b	classVBF	classGGF	classTT	classDY	classTTH
tt	0.42	260.90	220.41	17.52	2.91	119.56	4.54	26.63
dy+jets	8.27	829.12	90.16	14.66	8.51	23.25	36.64	1.75
qcd	1.07	769.12	58.13	18.1	1.56	80.61	9.89	0.26
tth	0.25	3.68	2.93	0.15	0.27	0.62	0.11	0.86
others	4.97	133.53	27.55	7.06	3.3	12.6	5.83	3.34
BKG SUM	14.98	1996.35	399.18	57.49	16.55	236.64	57.01	32.84
ggHH SM	0.14	0.7	0.82	0.051	0.066	0.0088	0.0039	0.008
VBFHH SM	0.0016	0.013	0.011	0.024	0.0025	0.0015	0.00076	0.00041

Results: Upper limits @SM

ATLAS RESULT r (ggF+VBF) SM only!

Table 5: Observed and expected upper limits at 95% CL on the cross-section of non-resonant HH production according to SM-like kinematics, and on the cross-section of non-resonant HH production divided by the SM prediction. The $\pm 1 \sigma$ and $\pm 2 \sigma$ variations around the expected limit are also shown.

		Observed	-2σ	-1σ	Expected	$+1 \sigma$	$+2 \sigma$
$\tau_{\text{had}} \tau_{\text{had}}$	$\sigma_{\text{ggF+VBF}}$ [fb]	145	70.5	94.6	131	183	245
	$\sigma_{\text{ggF+VBF}}/\sigma_{\text{ggF+VBF}}^{\text{SM}}$	4.95	2.38	3.19	4.43	6.17	8.27
$\tau_{\text{lep}} \tau_{\text{had}}$	$\sigma_{\text{ggF+VBF}}$ [fb]	265	124	167	231	322	432
	$\sigma_{\text{ggF+VBF}}/\sigma_{\text{ggF+VBF}}^{\text{SM}}$	9.16	4.22	5.66	7.86	10.9	14.7
Combined	$\sigma_{\text{ggF+VBF}}$ [fb]	135	61.3	82.3	114	159	213
	$\sigma_{\text{ggF+VBF}}/\sigma_{\text{ggF+VBF}}^{\text{SM}}$	4.65	2.08	2.79	3.87	5.39	7.22

- CMS $HH \rightarrow bb\tau\tau$ upper limits in $\sigma/\sigma_{\text{SM}}$:

ggF + VBF	3.3 (5.2)
VBF	124 (154)