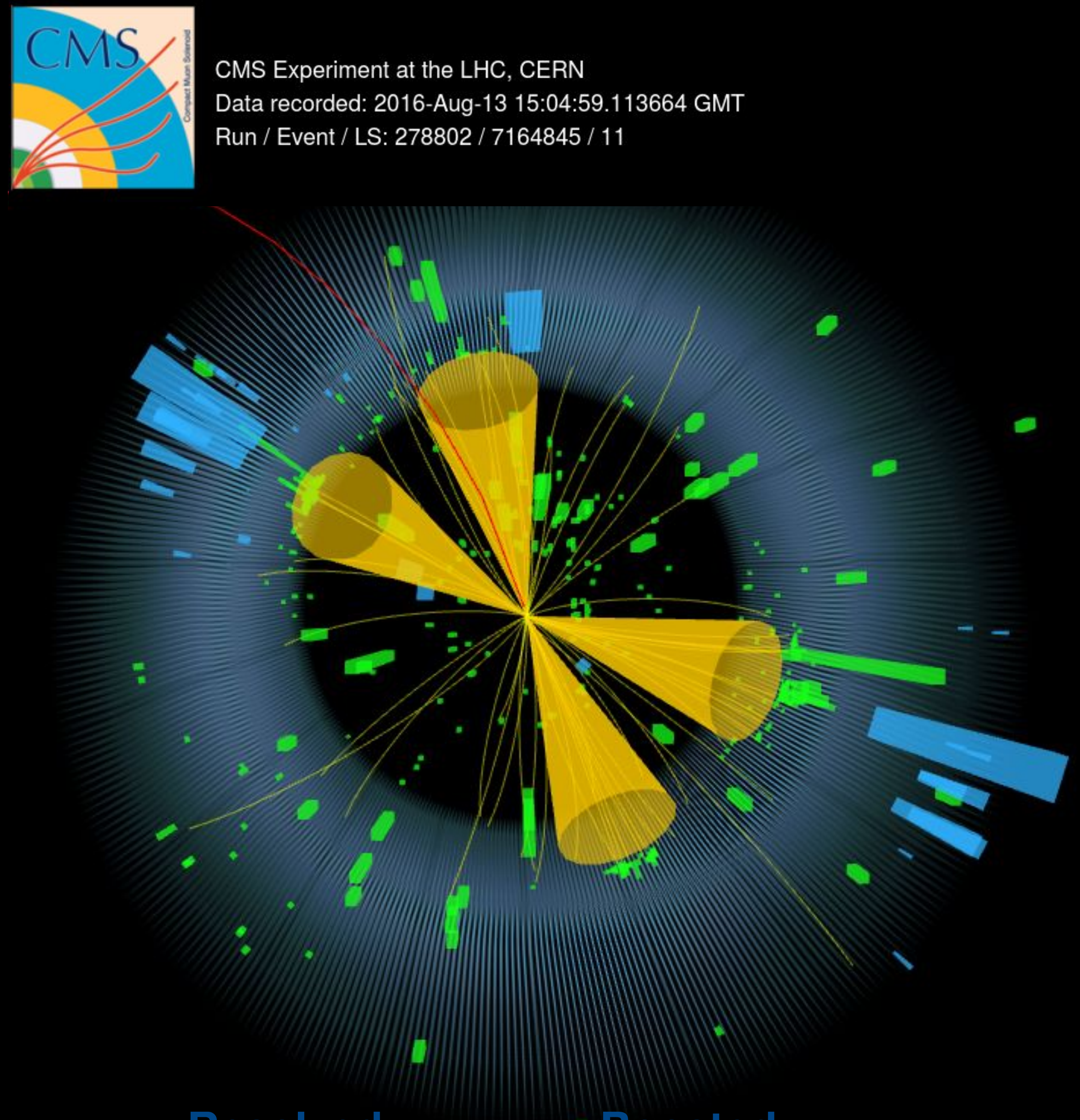
ATLAS + CMS Non-resonant HH→bbbb

Daniel Guerrero
and
Rafael Teixeira de Lima
on behalf of the ATLAS and CMS Collaborations

Outline



- The HH→ bbbb channel
- ATLAS and CMS Run-2 Overview
  - Jet tagging & Higgs reconstruction
  - Trigger
  - Event categorization
  - Background modeling
  - Discriminant observables
  - Systematics
  - Results: SM, anomalous kl and k2v
- Conclusions

More details in references: ATLAS (CONF-2022-035) [Resolved] and CMS (ArXiv:2202.09617 [Resolved], ArXiv:2205.06667 [Boosted])

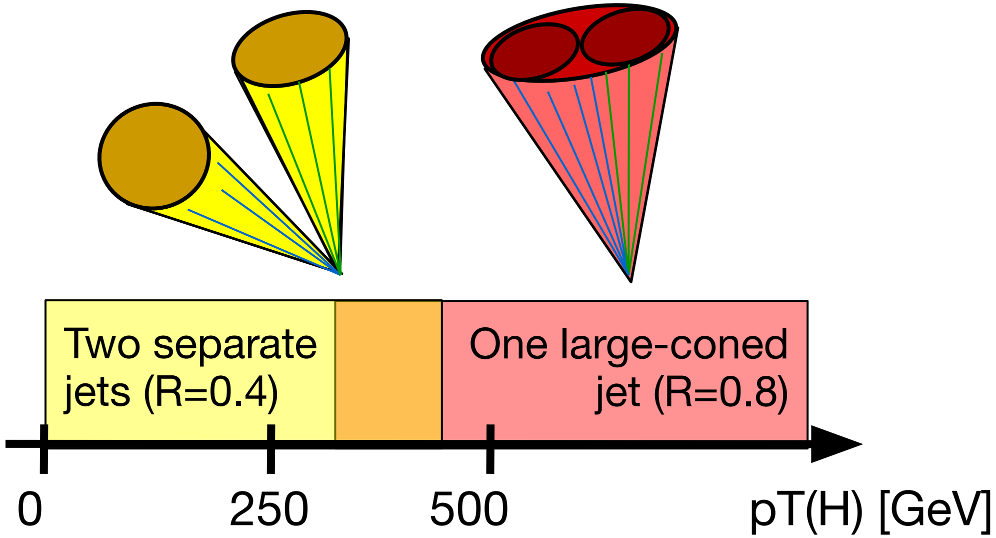
NEW RESULT! (ATLAS)

Both submitted to PRL! (CMS)

The bbbb decay channel

It has the HH largest rate ~ 1500 SM events produced in the LHC Run-2 period (L~140 fb^{-1})!
But searching for signal events is challenged by
the large production of multi-jet bkg. events (QCD multijet~90-95%, top quarks~5-10%)

H→ bb experimental fingerprint depends on Higgs pT



Jet clustering with Anti-kT (AK) algorithm

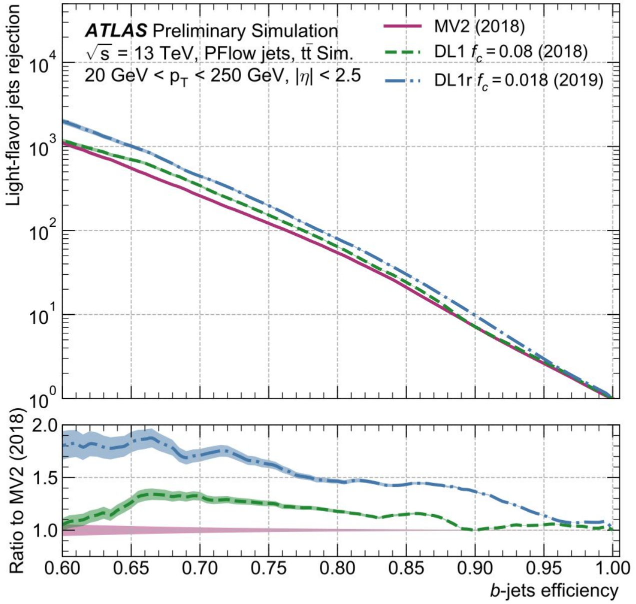
Experimental challenges:

- b quark jet or H jet identification w.r.t large usdg/c jet bkg
- Online trigger algorithms are complex
  - Depends on L1 seed, HLT tracking, jet reco/cal, b-tagging, etc
  - Constrained by L1 rate, HLT CPU limit & output rate
  - Consistency with offline algorithms (e.g. b-tagging)
- Higgs boson reconstruction affected by
  - Large jet combinatorics
  - Missing energy from neutrinos in semi-leptonic B decays
  - Jet constituents from ISR, FSR & Pile-up
- Precise model and rejection of multijet bkg are crucial

Small-R jet b-tagging (resolved)

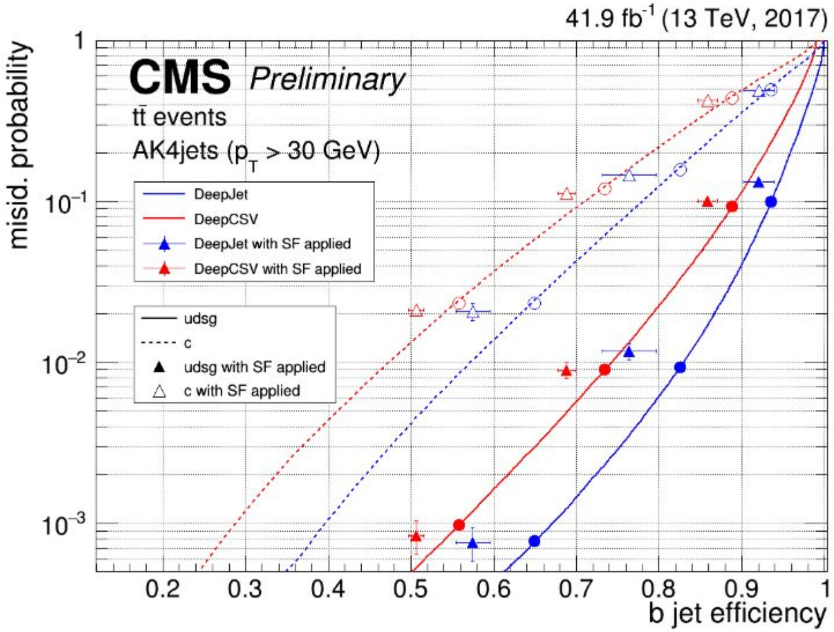
The b-jet candidates are identified using machine learning algorithms

ATLAS



DL1r performance plots

CMS



DeepJet performance plots

Best performing AK4 b-jet tagger: DL1r

- DNN + RNN (tracks' impact parameters)
- 10% improvement in efficiency for same bkg rejection

b-jet candidates:

- Pass DL1r at 77% b-jet efficiency (measured in ttbar)
- Muon-in-jet information used to correct jet pT

Best performing AK4 b-jet tagger: DeepJet

- DNN multioutput classifier
- 10% eff. Improvement w.r.t previously used (DeepCSV)

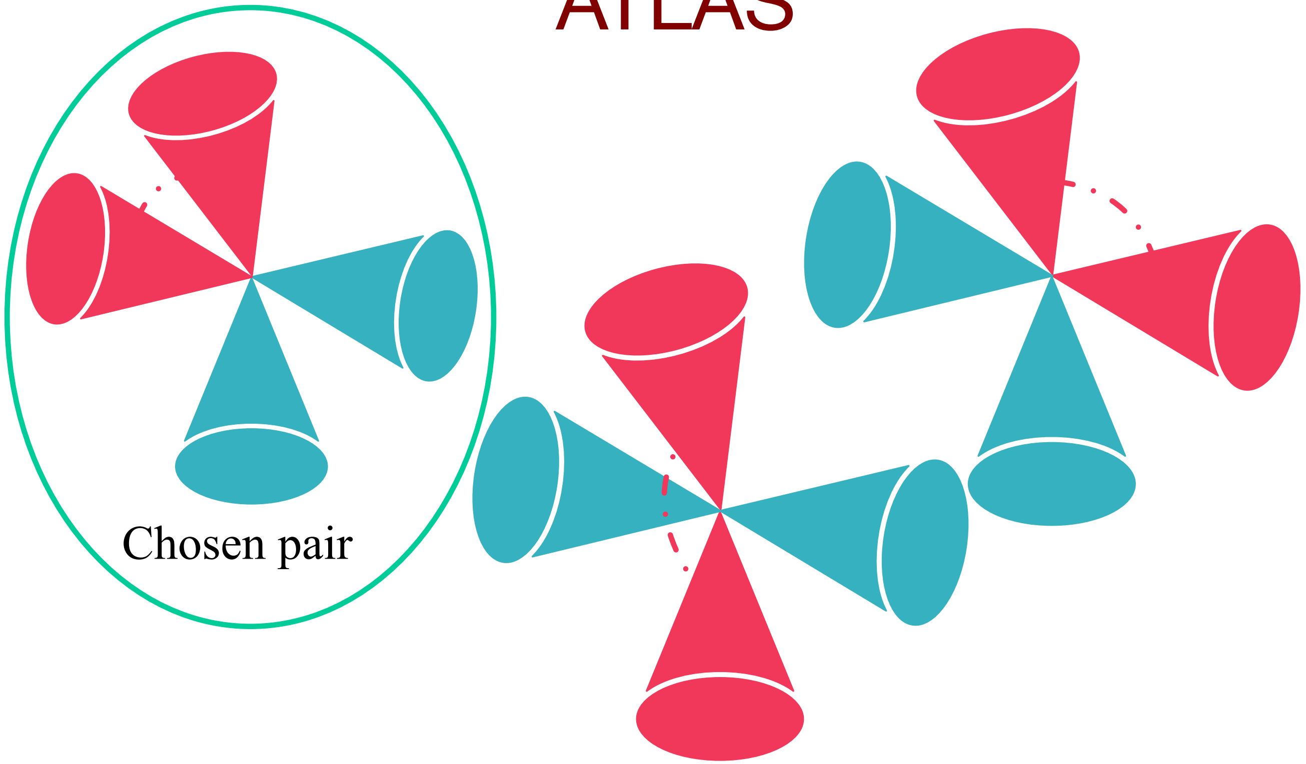
b-jet candidates:

- Pass DeepJet WP with ~75% efficiency (measured in ttbar)
- DNN b-jet energy regression is used to correct jet pT

Higgs boson reconstruction (resolved)

The challenge: Four b-jet candidates→ 3 pairing possibilities

ATLAS



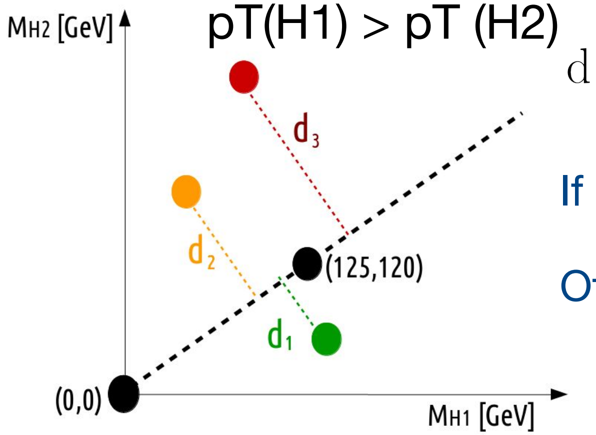
Choose pairing that minimizes ΔR distance between jets in the leading Higgs candidate (H1)

90% accurate for SM ggF events

where *4 b-jets = 4 b-quarks from Higgs decays*

CMS

Jet pairing based on m_{H1}-m_{H2} plane information



$$d = \frac{|M_{H1} - k\,M_{H2}|}{\sqrt{1+k^2}}$$

d1 < d2 < d3

k=125/120=1.04

If |d1-d2| ≥ 30 GeV:

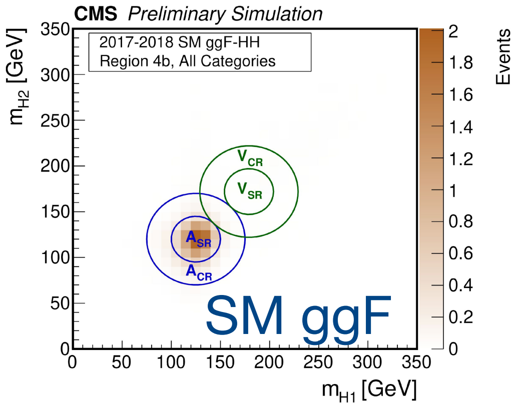
d1 (closest to diagonal)

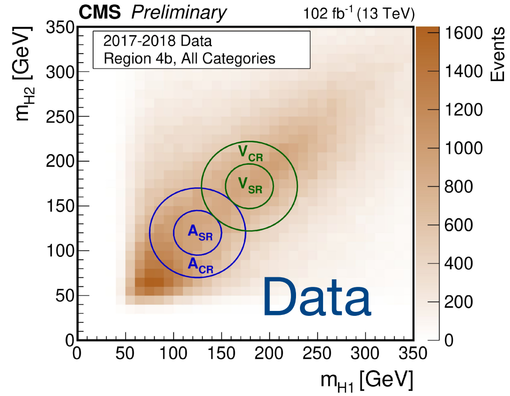
Otherwise:

d1 or d2 maximizing pT(H) in 4-jet frame

Very good performance (e.g. 96% accurate for SM ggF)

Maximizing signal collection w/o bkg. sculpting near the mH





H-tagging & reconstruction (CMS boosted)

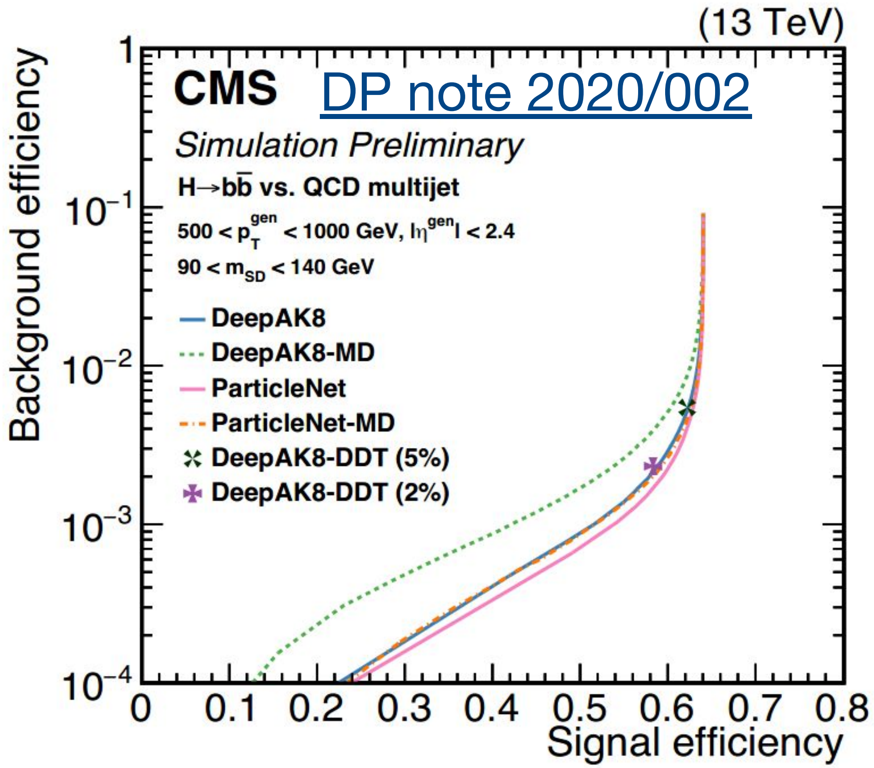
Best performing AK8 jet tagger: ParticleNet

- Architecture based on Graph Neutral Networks
- Mass-decorrelated version identifies 2-prong hadronic decays of highly boosted particles (e.g. X→bb)

D_{bb} discriminant = P[X→bb] / P[X→bb] + P[QCD]

- Input features:
  - Jet constitutents PF candidates & secondary vertices
- Performance: 0.1% bkg eff. at 50% signal eff:
  - Bkg rejection improved by a ~factor of 2 per jet than previously
- Same architecture is used for regression of the jet mass (m_{reg})

More details on
'ATLAS+CMS jet and flavor tagging' talk
by Loukas tomorrow!



Calibration (DP note 2022/005)

Trigger Strategy

ATLAS

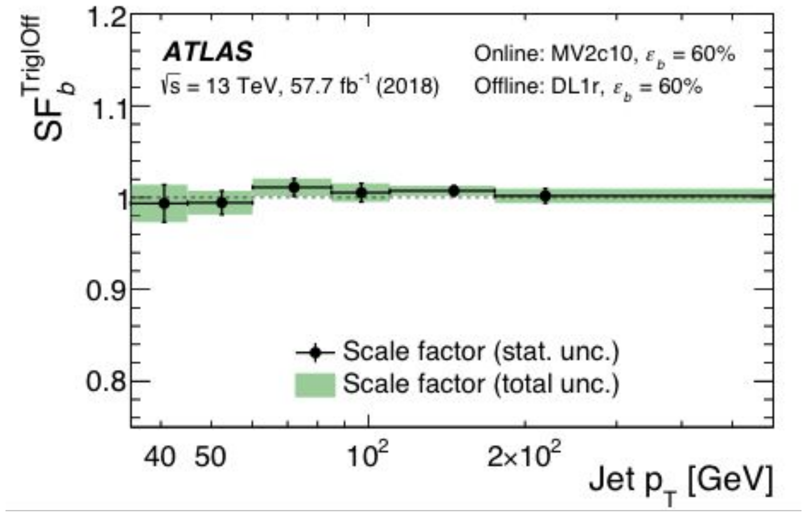
2b2j: 2 b-tagged jets + 2 extra jets

- $p_T > 35$ GeV for all jets
- b-jet eff. between 40% and 70%
- Important for low m_{HH} events

2b1j: 2 b-tagged jets + 1 extra jet

- $p_T > 35$ GeV for b-jets, $p_T > 100$~150 GeV for extra jet
- b-jet eff. between 40% and 70%
- Important for high m_{HH} events

ATLAS b-jet trigger



CMS

Resolved:

Triggers require 4 jets with 3 b-tagged jets

- 2016:
  - 4j with pT> 45 GeV
  - 2j with pT > 30 GeV, 2j with pT> 90 GeV
- 2017 (2018):
  - 4j with pT > 40,45,60,75 GeV, HT> 300 (330) GeV

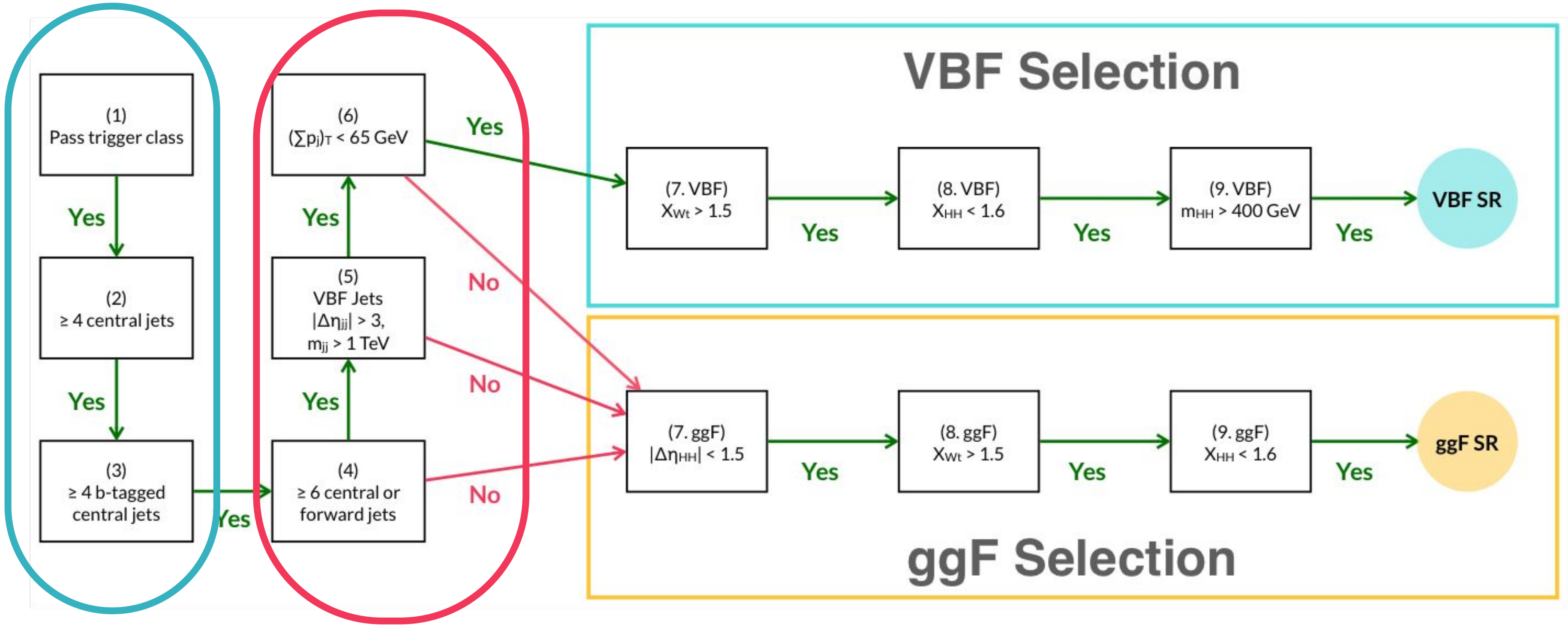
Boosted:

Combination of several single-jet and HT triggers:

- Requirements on:
  - HT, jet pT, Trimmed mass & double b-tagging
- Fully efficient for Jet pT>500 GeV

Event Categorization (ATLAS resolved)

Same analysis selection is performed in 2016, 2017 & 2018 datasets, L_{total}=126 fb^{-1}



Common selection

ggF vs VBF separation

Categories

$\|\Delta\eta_{HH}\| < 1.5$	$\|\Delta\eta_{HH}\| > 1.5$
VBF Cat 1	VBF Cat 2

ggF Cats.	$\|\Delta\eta_{HH}\| < 0.5$	$0.5 > \|\Delta\eta_{HH}\|$ $\|\Delta\eta_{HH}\| < 1.0$	$1.0 > \|\Delta\eta_{HH}\|$ $\|\Delta\eta_{HH}\| < 1.5$
$X_{HH} < 0.95$	1	2	3
$0.95 > X_{HH}$ $X_{HH} < 1.6$	4	5	6

Top veto cut

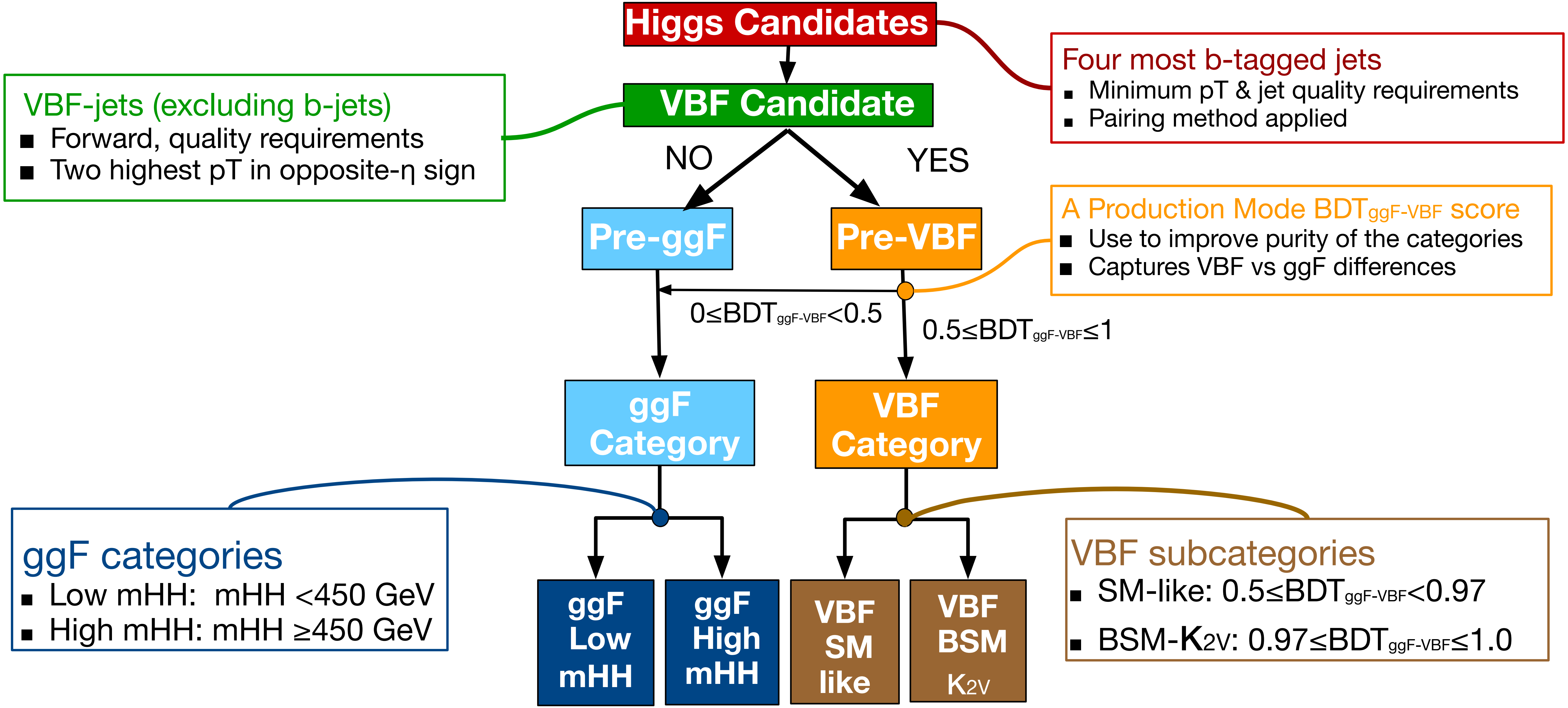
$$X_{Wt} = \sqrt{\left(\frac{m_W - 80.4\text{ GeV}}{0.1\, m_W}\right)^2 + \left(\frac{m_t - 172.5\text{ GeV}}{0.1\, m_t}\right)^2}$$

Signal region cut

$$X_{HH} = \sqrt{\left(\frac{m_{H1} - 124\text{ GeV}}{0.1\, m_{H1}}\right)^2 + \left(\frac{m_{H2} - 117\text{ GeV}}{0.1\, m_{H2}}\right)^2}$$

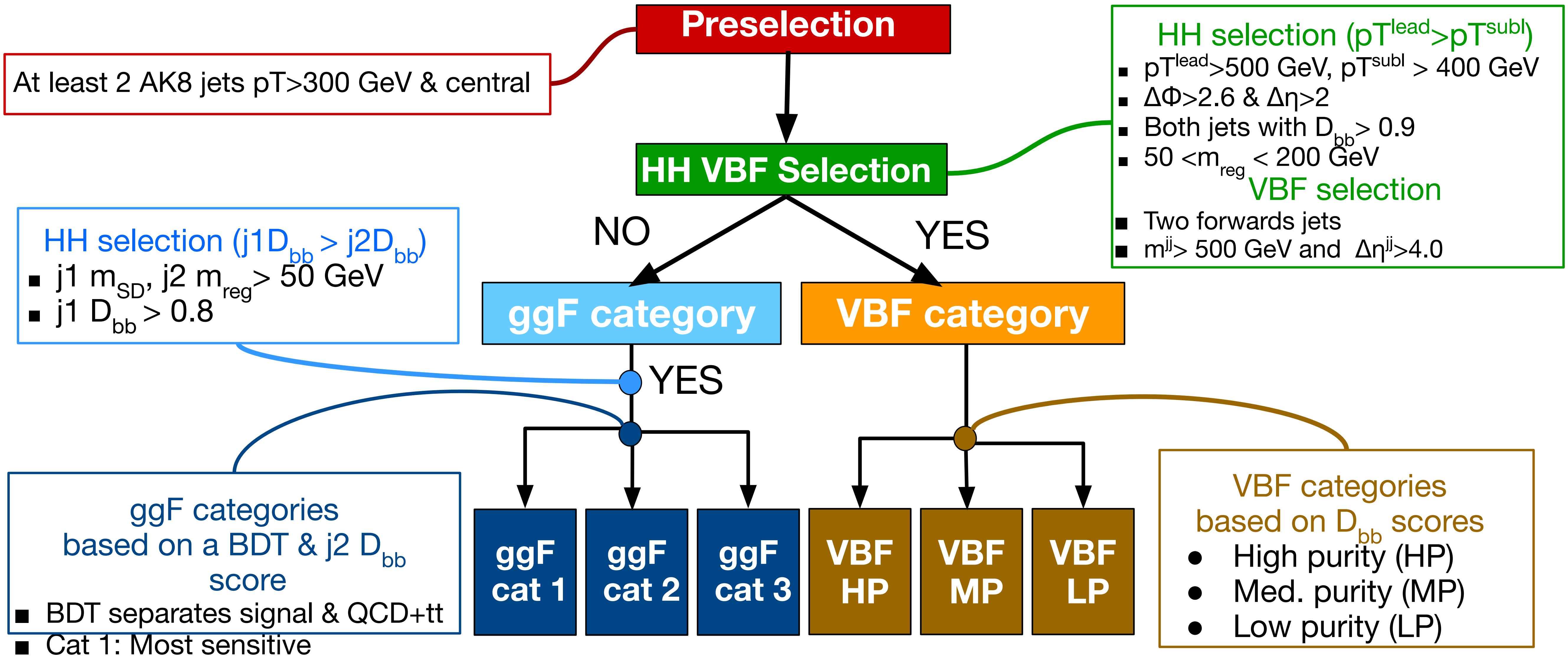
Event Categorization (CMS resolved)

Same analysis selection is performed in 2016 & 2017-2018 datasets, L_{total}=138 fb^{-1}



Event Categorization (CMS boosted)

Same analysis selection is performed in 2016, 2017 & 2018 datasets, L_{total}=138 fb^{-1}

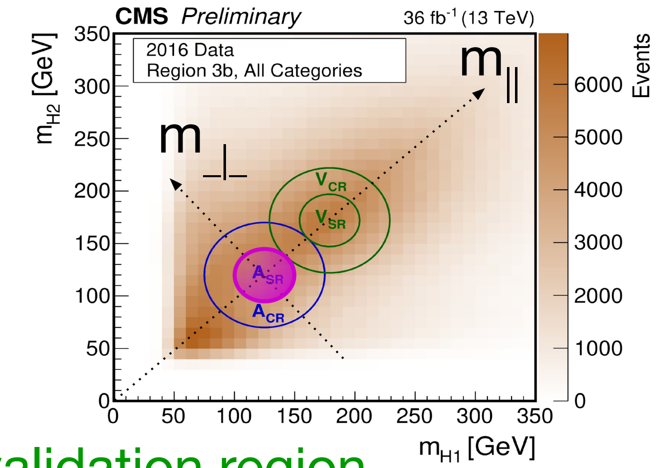


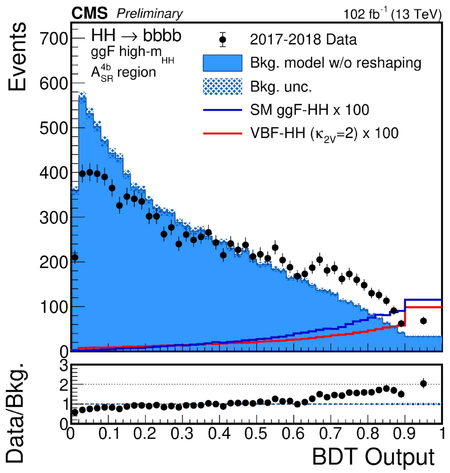
Background estimation (resolved)

Data-driven multijet bkg model using ‘low-btag’ data to derive ‘4-btag’ background distributions

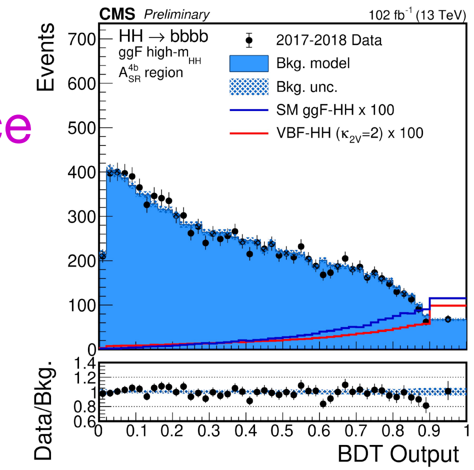
CMS

- 3b data → 4b full bkg
- Derived in CR data (ring)
  - Transfer factor (4b/3b)
  - Residual differences corrected via weights from BDT re-weighting
- Applied to SR (circle)
- Method is fully verified in validation region





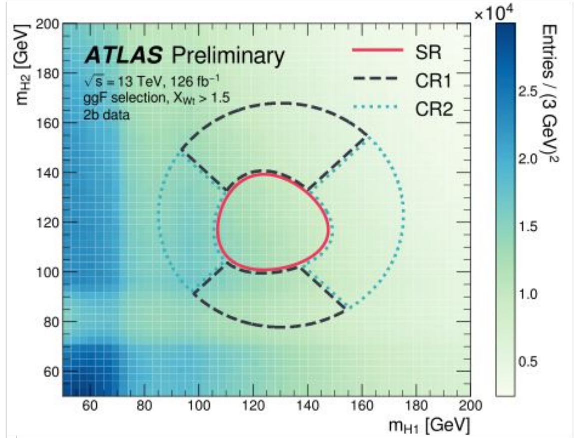
Performance in SR(4b) region



ATLAS

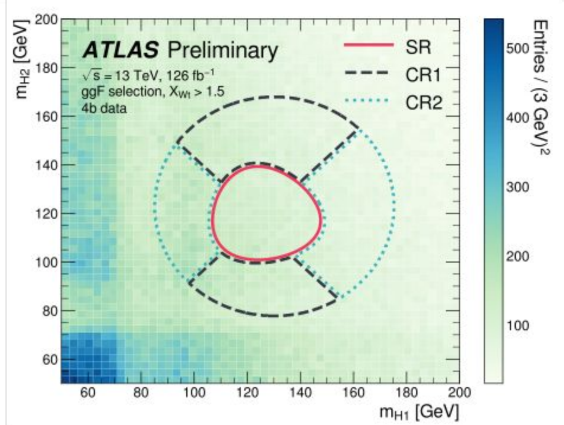
Similar to CMS method, but three main differences:

- Use 2b data instead of 3b (use 3b as validation)
- Split ring in two: nominal and alternative models
- Use DNN instead of BDT



Many validation regions:

- Shifted region (like CMS)
- 3b+1f region
- Reversed $|\eta_{HH}|$ selection



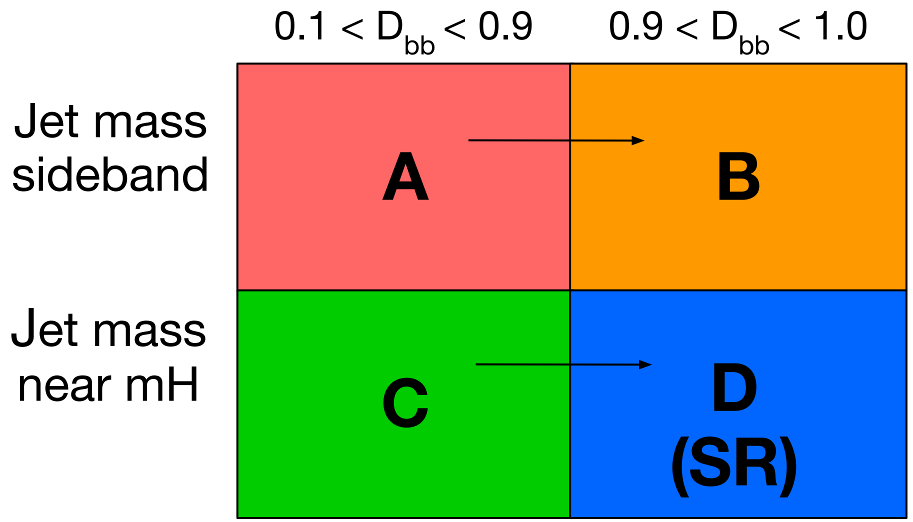
Background estimation (CMS boosted)

ggF Categories

- QCD bkg. is derived using parametric alphabet method:
  - Fail region: j2 D_{bb} score < 0.95 & BDT>0.03
  - Model in pass region = fail region shape x t.f. (j2 m_{reg})
- ttbar bkg. is modeled from MC:
  - Correction from ttbar(had) & ttbar(semilep.) in CRs
    - ttbar recoil & D_{bb} shape corrections
- Other minor bkgs (Single H, VV, etc) from MC

VBF Categories

- QCD bkg derived using ABCD method:
  - Model D = Shape from C region x tf



- ttbar bkg is modeled from MC
  - Corrections from semileptonic tt events
    - Normalization & bb-mistagging SFs

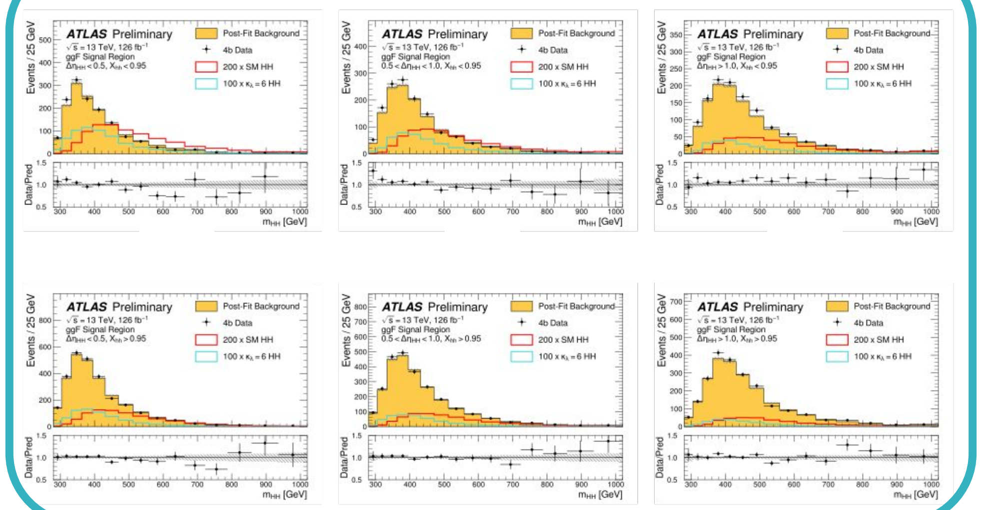
More details on
<u>Data-driven QCD background methods talk</u>
by Matej tomorrow!

Discriminant observables (resolved)

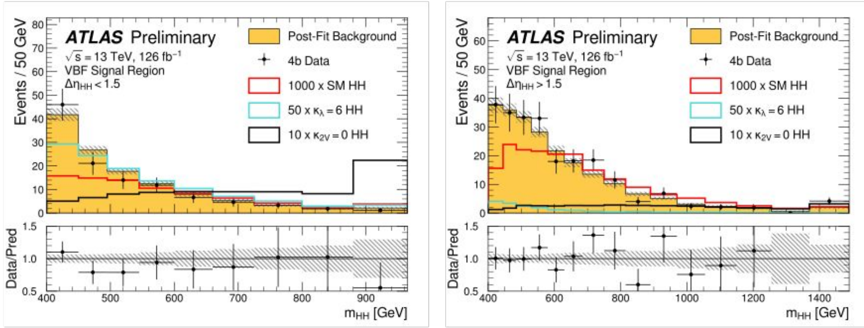
The selected observables maximize the analysis sensitivity

ATLAS

Gluon-gluon fusion



VBF

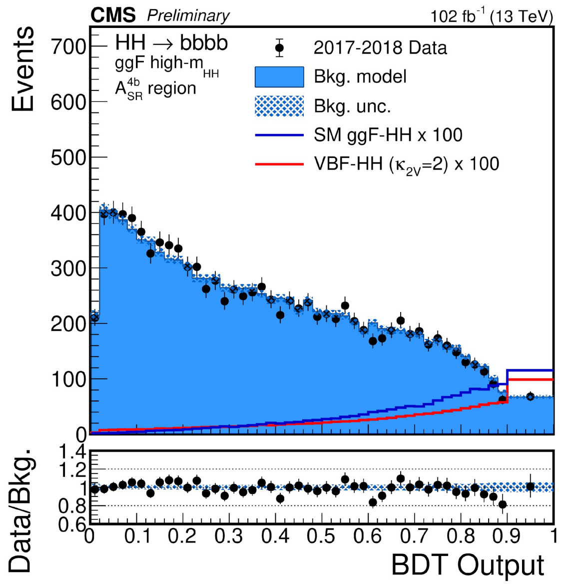


CMS

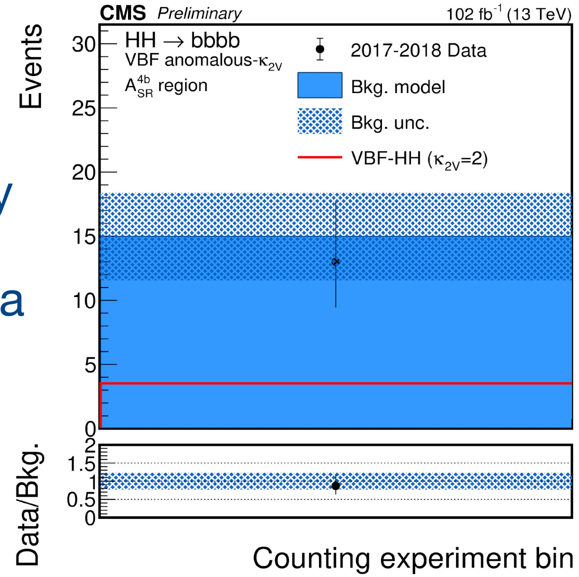
ggF categories: BDT distribution

- Bkg model stats. enables the model of a BDT discriminant
- Discriminant is trained by category (16 variables)

VBF SM category: mHH distribution
VBF BSM category: Counting experiment



B-only Fit to data



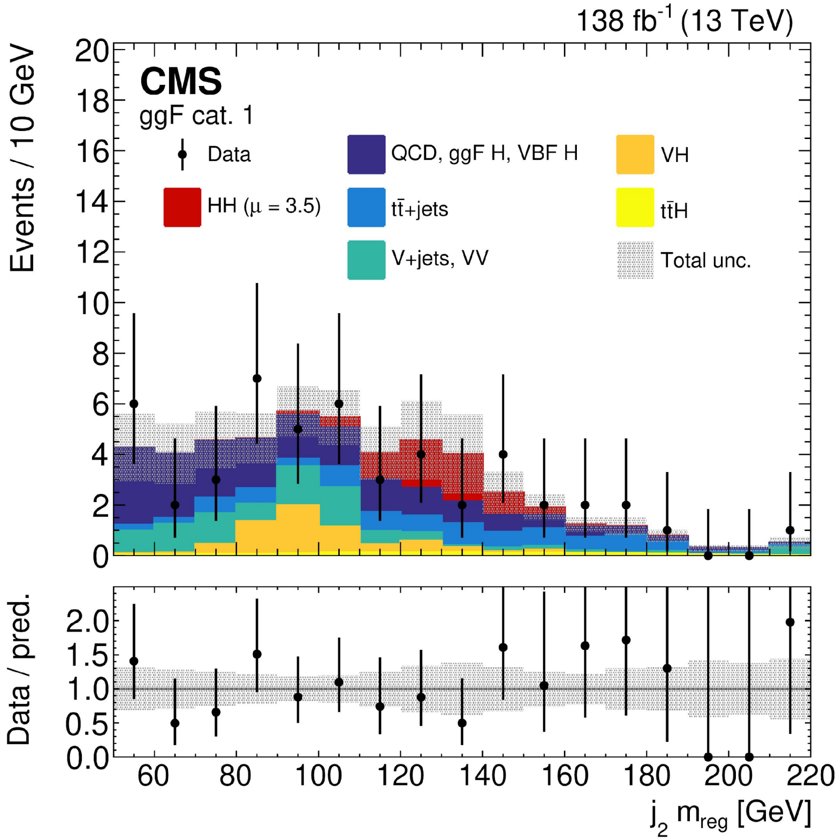
Signal extraction performed by fitting signal region distributions of all categories/datasets

Discriminant observables (CMS boosted)

The selected observables maximize the analysis sensitivity

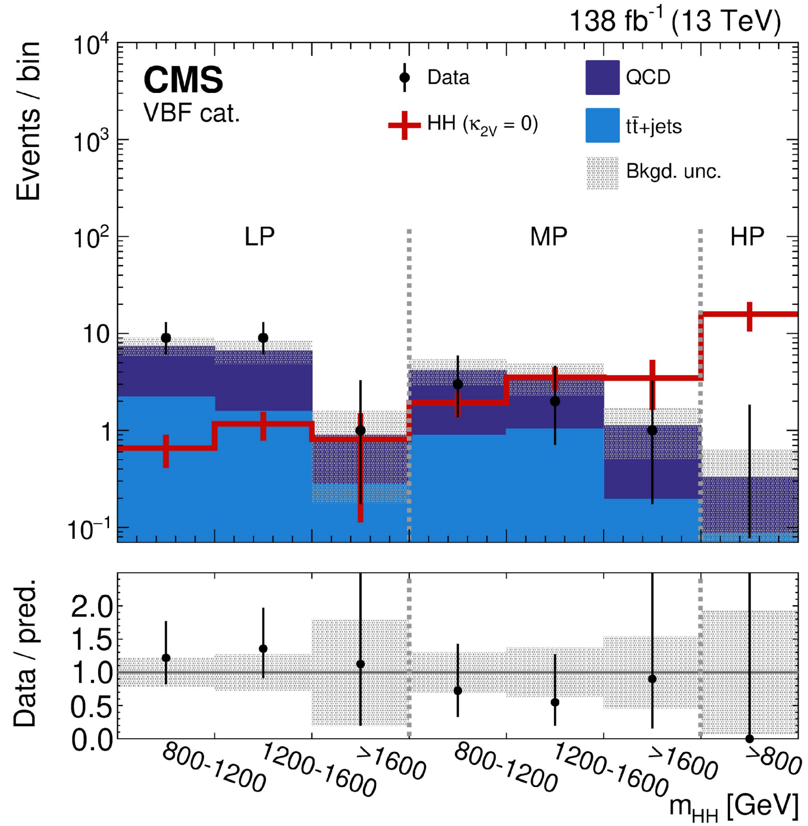
ggF categories: j2 regressed mass

S+B fit to data on most sensitive ggF category



VBF categories: mHH distribution (3 or 1 bin)

B-only fit to data



Signal extraction performed fitting all categories in signal and control regions

Systematics (resolved)

The bbbb resolved analyses are dominated by background estimate uncertainties!

ATLAS

Background uncertainties:

- Statistical: 2b statistics + DNN variation under bootstrapped deep ensembles (100 trainings)
- Alternative vs nominal estimate (CR1 vs CR2)
- 3b1f region non-closure
- Normalization uncertainty from 2b/4b CR

CMS

Background uncertainties:

- Statistical uncertainty in SR(3b) region (dominant)
- Transfer factor from CR 4b/3b regions
- Shape variation (Alternative CR definition vs nominal)
- Validation yields non-closure
- Validation limited statistical precision

ATLAS & CMS signal uncertainties

- Luminosity , PU, trigger
- b-tagging, jet scale and resolution,
- PS, PDF, factorization scales
- PS dipole recoil in VBF signals (CMS-only)
- Cross section & branching ratio

Systematics (CMS boosted)

Breakdown of the impact of systematics on measured SM signal strength (μ=3.5) grouped by their statistical, systematic and theoretical nature

Uncertainty source	$\Delta\mu$	
Statistical	$+2.55$	-2.30
Signal extraction	$+2.32$	-2.06
QCD multijet modeling	$+1.12$	-1.01
$t\bar{t}$ modeling	$+0.28$	-0.19
Systematic	$+2.09$	-0.89
Simulated sample size	$+0.55$	-0.55
$D_{b\bar{b}}$ selection	$+0.72$	-0.32
Jet energy and mass scale and resolution	$+0.54$	-0.39
Trigger selection	$+0.26$	-0.03
Luminosity measurement	$+0.13$	-0.04
Pileup modeling	$+0.05$	-0.06
Other experimental uncertainties	$+0.05$	-0.03
Theoretical	$+0.63$	-0.63
Total	$+3.30$	-2.47

Results: SM

No excess of events is observed relative to the background-only hypothesis
95% CL upper limits are set on the SM HH (ggF+VBF) production

ATLAS

Observed (expected) limit (resolved):

5.4 (8.1) x SM prediction (~3 x more sensitive than before)

CMS

Observed (expected) limit (resolved):

3.9 (7.8) x SM prediction (~5 x more sensitive than before)

Best observed constraint on bbbb channel!

Observed (expected) limit (boosted):

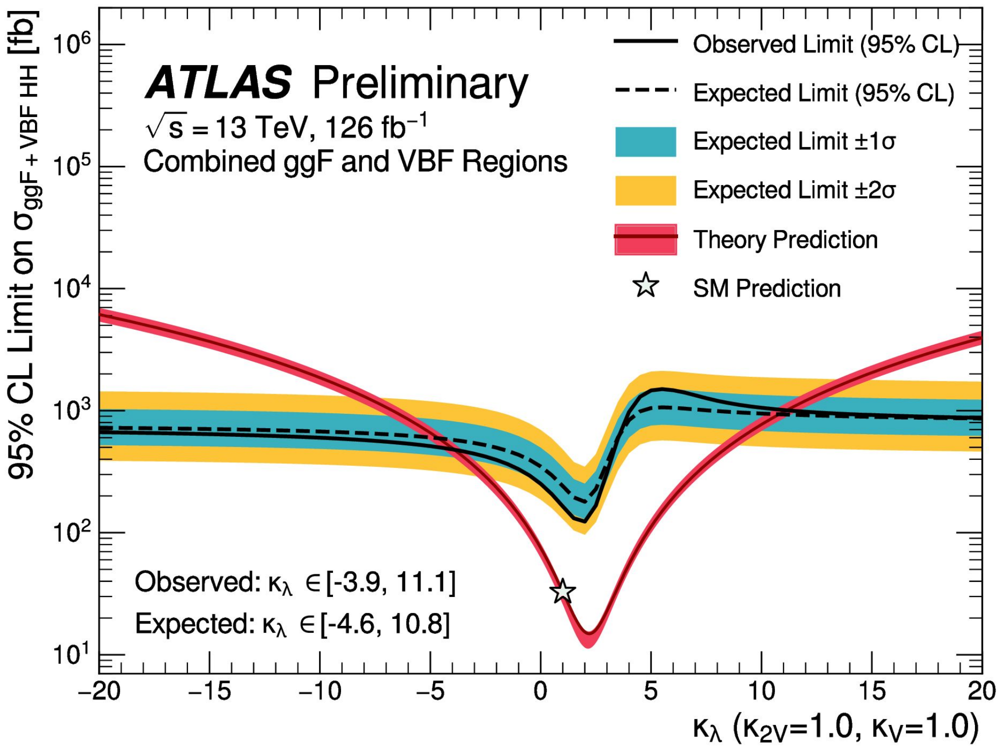
9.9 (5.1) x SM prediction (~30 x more sensitive than before)

Significant improvements using full Run-2 data!

Results: Anomalous κ_λ

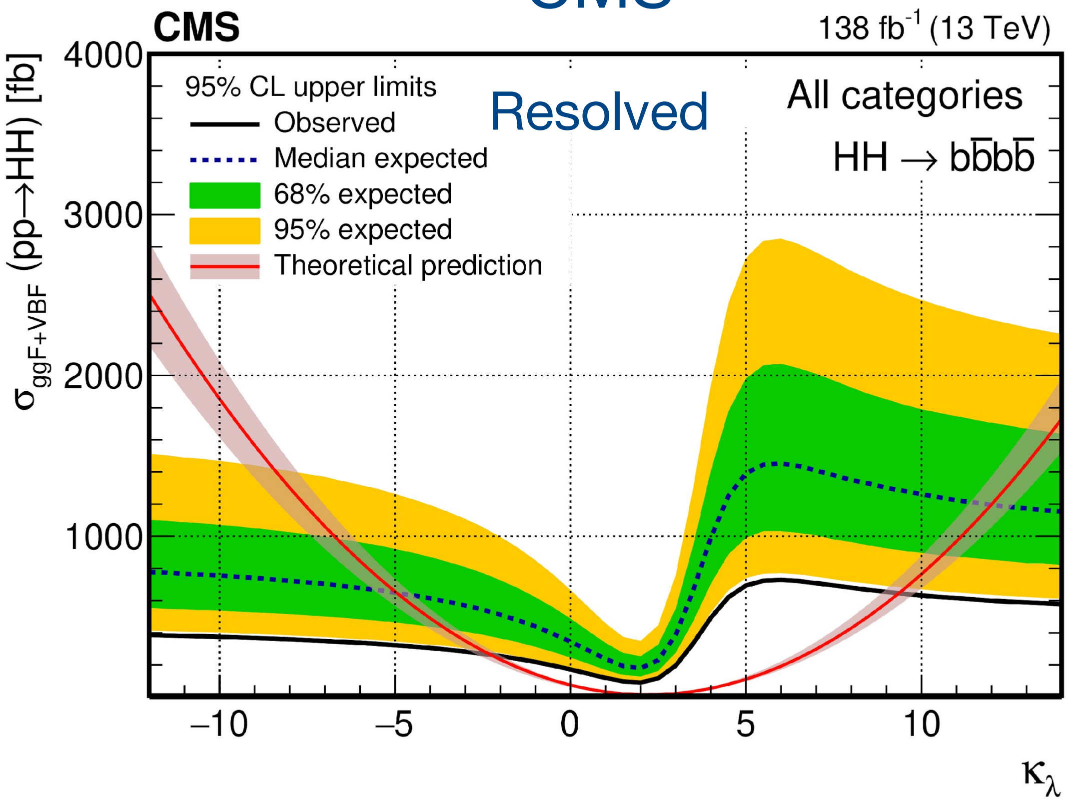
95% CL upper limits are set on the HH production cross section vs k_λ values

ATLAS



Obs. (exp.) constraint (resolved)
[-3.5,11.3] ([-5.4,11.4])

CMS

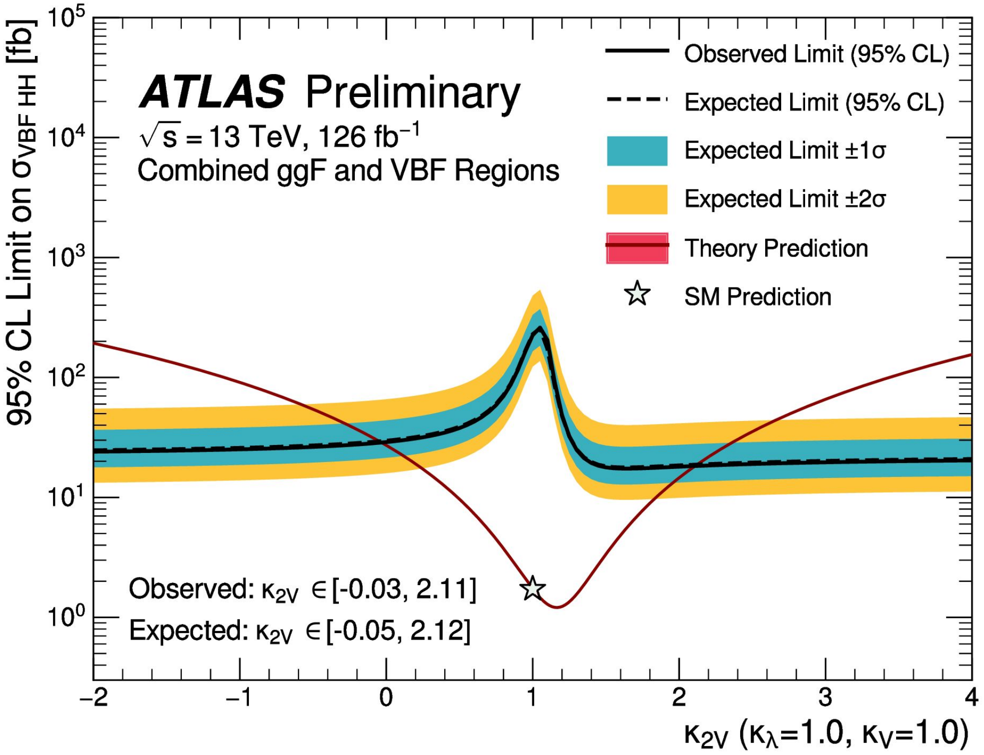


Obs. (exp.) constraint (resolved)
[-2.3, 9.4] ([-5.0,12.0])
Obs. (exp.) constraint (boosted)
[-9.9,16.9] ([-5.1,12.2])

Results: Anomalous κ_{2V}

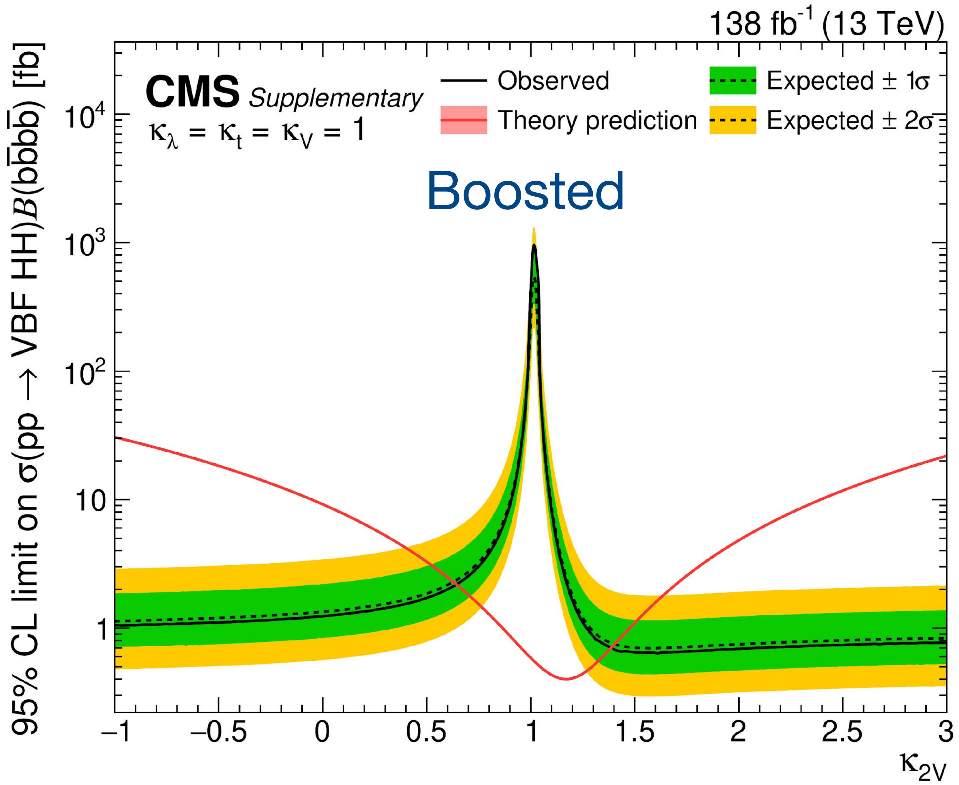
95% CL upper limits are set on the HH production cross section vs κ_{2V} values

ATLAS



Obs. (exp.) constraint (resolved)
[0, 2.1] ([-0.1,2.1])

CMS



Obs. (exp.) constraint (resolved)
[-0.1, 2.2] ([-0.4,2.5])
Obs. (exp.) constraint (boosted)
[0.6,1.4] ([0.7,1.4])

The κ_{2V}=0 hypothesis is excluded for the first time, with a 6.3 s.d. significance ($\kappa_\lambda=\kappa_V=\kappa_t=1$)

Conclusions

The bbbb channel is one of the main players in the HH game

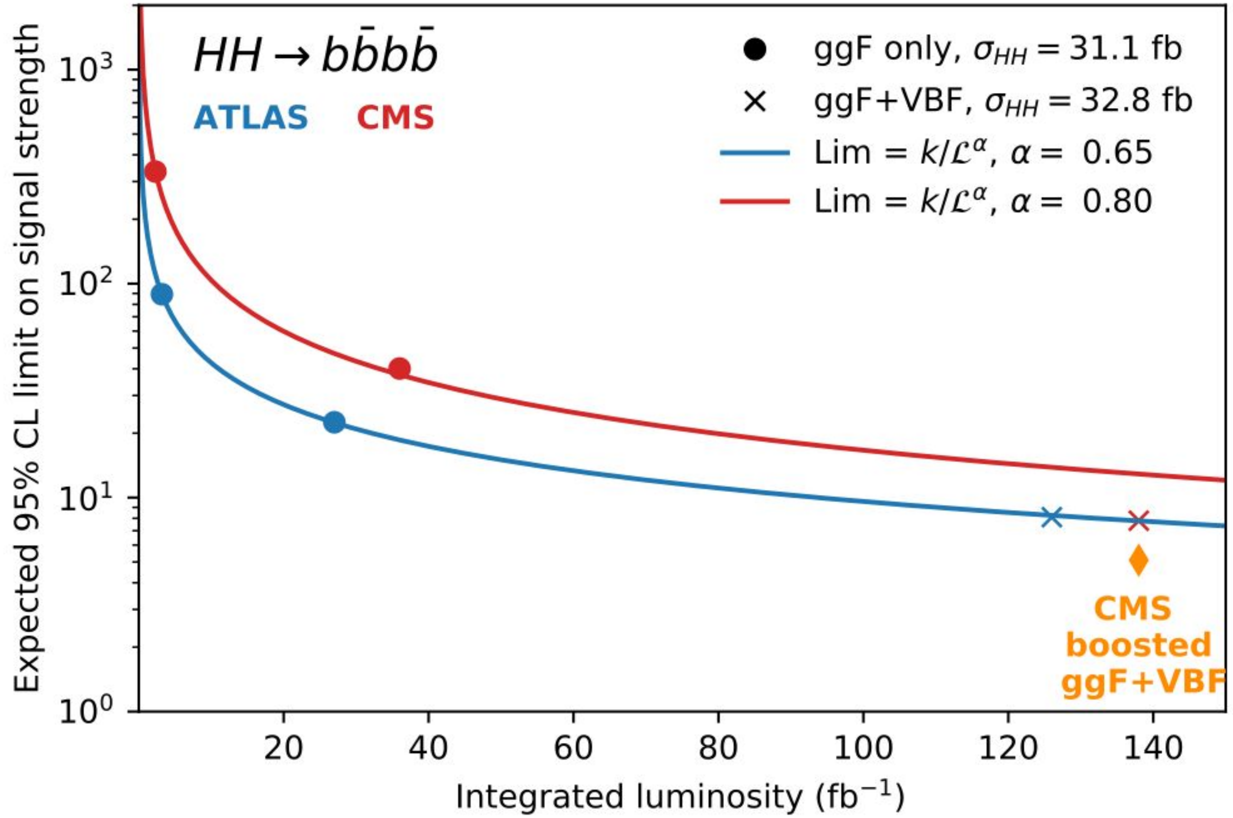
Significant improvements in both ATLAS and CMS Run-2 analyses!

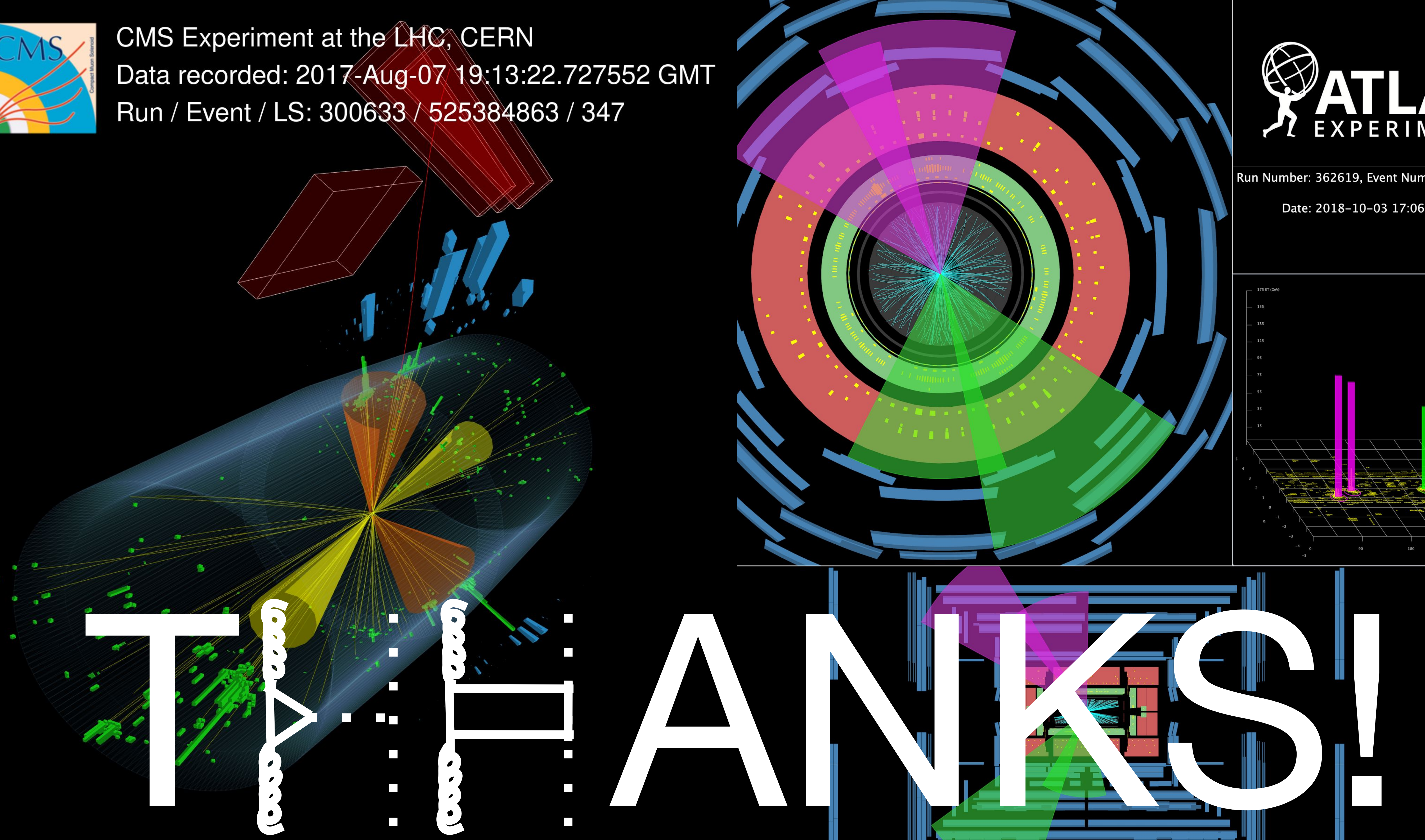
- Sophisticated b-tagging & Higgs-tagging using AI/ML
- New jet pairing methods
- Dedicated ggF and VBF mode categorizations
- Novel background estimation using ML-based reweighting

Looking forward to the future bbbb analyses!

- How can we improve our background estimate?
- How can we make it more sensitive in the low mass region?
- How can we cope with the increasing trigger thresholds in the (HL) LHC's future?

Stay tuned!




CMS Experiment at the LHC, CERN
Data recorded: 2017-Aug-07 19:13:22.727552 GMT
Run / Event / LS: 300633 / 525384863 / 347
ATLA
EXPERIM
Run Number: 362619, Event Num
Date: 2018-10-03 17:06
THANKS!

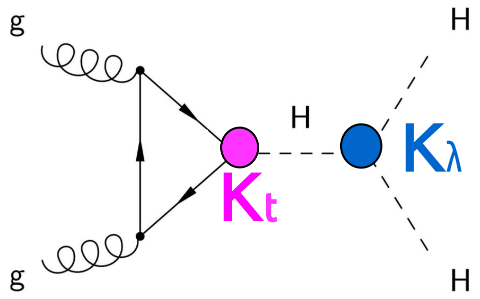
Non-resonant HH production at the LHC

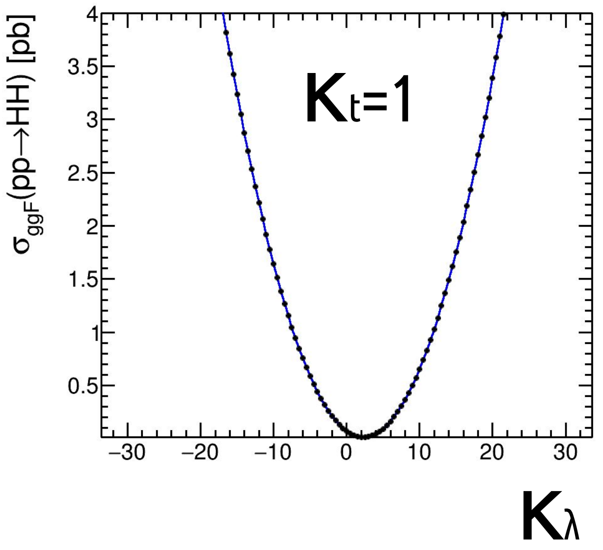
SM production gives access direct access to self-coupling (λ) → Reconstruct the Higgs potential
It's hard to measure it at current LHC datasets (let's see how far we can go!)

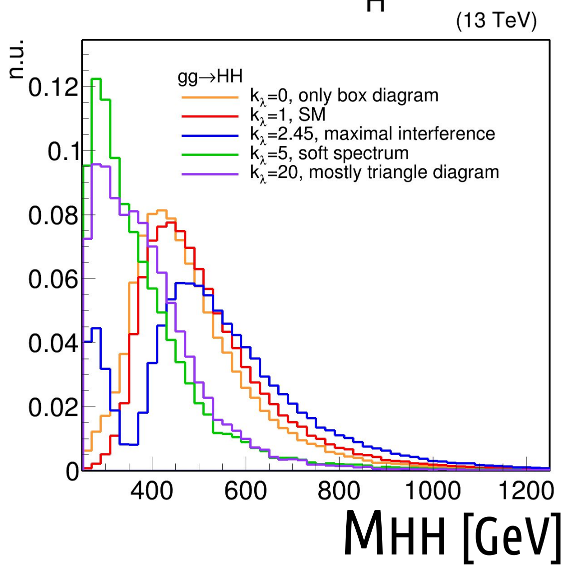
BSM physics may modify Higgs couplings or activate vertices → Potential for discovery at the LHC!

- Anomalous Higgs couplings are studied w.r.t. the SM using a **K**-framework, e.g. $\mathbf{K}_\lambda = \lambda / \lambda^{SM}$
- Large changes in cross sections & kinematics predicted

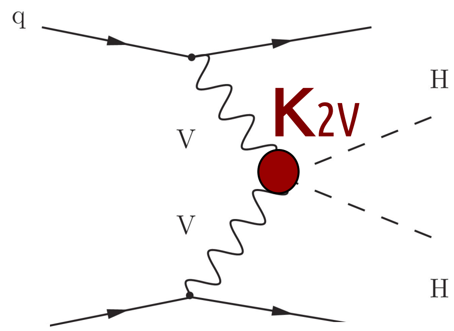
BSM self-couplings in ggF mode

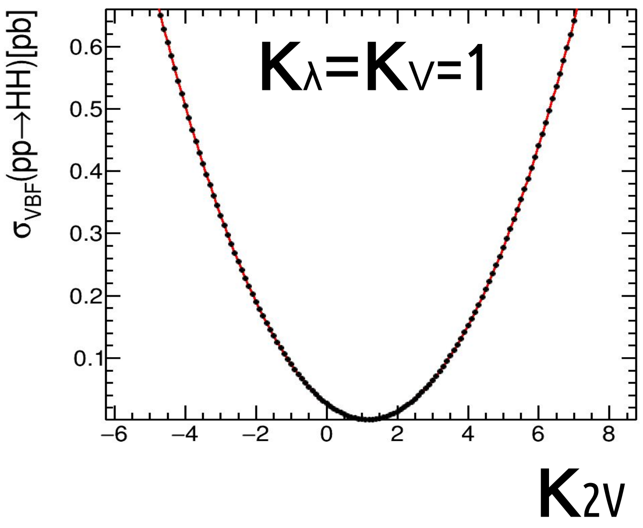


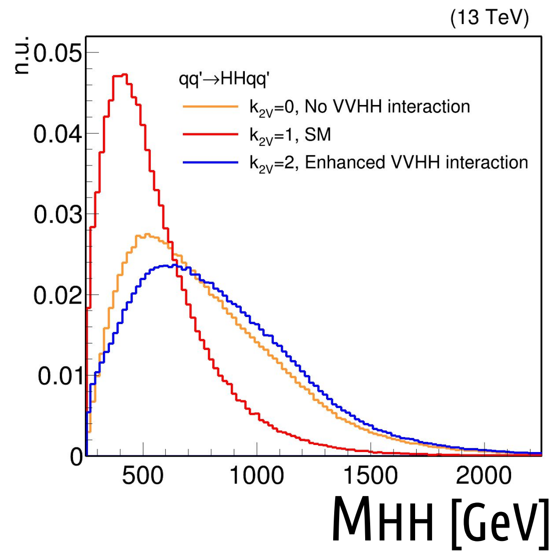




BSM VVHH couplings in VBF mode

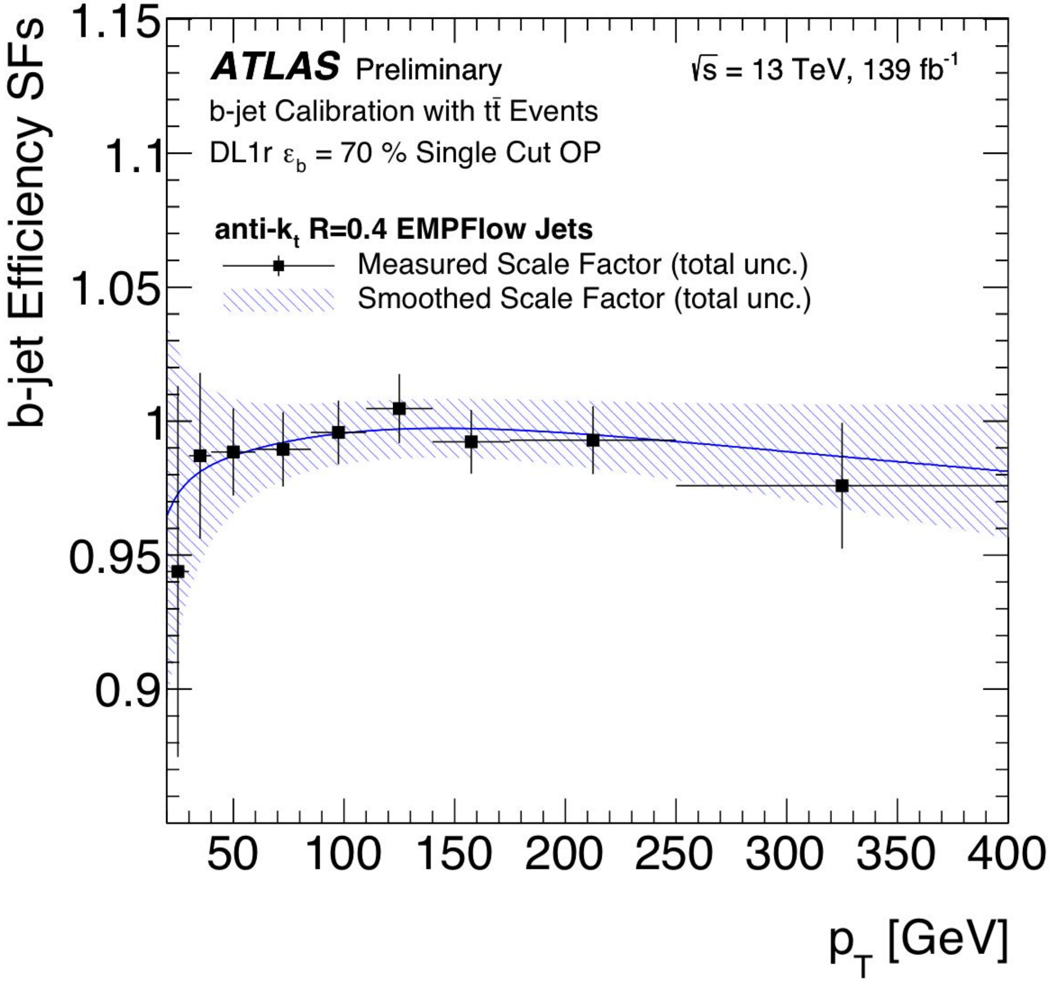






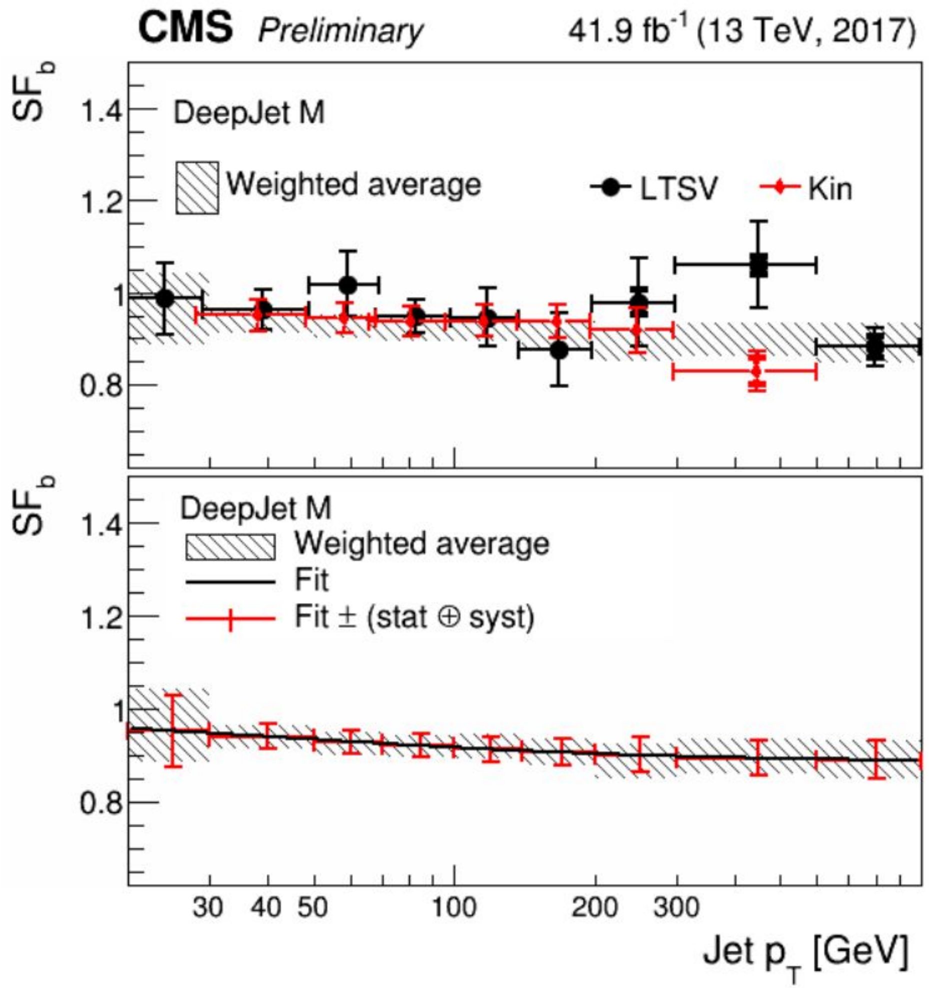
Small-R b-tagging calibration

ATLAS



DL1r calibration plots

CMS



DeepJet calibration plots

Production Mode BDT (CMS resolved)

Target: Pre-VBF events

Signal (S): VBF-HH (k2v=2)

- Signature with strongest contribution from longitudinal scattering amplitude V(L)V(L)→HH
- VBF-HH (k2v=0) has similar response

Background (B): NLO SM ggF-HH

Variable	Meaning
$p_T(H_1)$ ($p_T(H_2)$)	Tranverse momentum of the H_1 (H_2) candidate
$p_T(j_1)$ ($p_T(j_2)$)	Tranverse momentum of the j_1 (j_2) candidate
$\|\eta(jj)\|$	VBF-jet pair pseudorapity
$M(jj)$	VBF-jet pair invariant mass
$\Delta R(H_1, H_2)$	ΔR distance between two Higgs bosons
$\Delta R(H_1, j_1)$	ΔR distance between H_1 and j_1
$\Delta R(H_1, j_2)$	ΔR distance between H_1 and j_2
$\Delta R(H_2, j_1)$	ΔR distance between H_2 and j_1
$\Delta R(H_2, j_2)$	ΔR distance between H_2 and j_2
$\|\cos(\theta)^*(j1)\|$	$\|\cos(\theta)\|$ of j_1 in the six-jet center of mass frame
$\|\cos(\theta)^*(j2)\|$	$\|\cos(\theta)\|$ of j_2 in the six-jet center of mass frame
H1-centrality · H2-centrality	Product of the Higgs boson centralities

where:

$$\text{H1-centrality} \cdot \text{H2-centrality: } \exp\left[-\left(\frac{\eta(H_1)-\eta_{avg}}{\Delta\eta}\right)^2 - \left(\frac{\eta(H_2)-\eta_{avg}}{\Delta\eta}\right)^2\right],$$

$$\Delta\eta = \eta(j_1) - \eta(j_2) \qquad \eta_{avg} = \frac{\eta(j_1)+\eta(j_2)}{2}$$

BDT-reweighting variables (CMS resolved)

ggF categories

BDT Reweighter Input variables
Regressed p_T of the leading-p_T b jet of the H_1 candidate
Regressed p_T of the trailing-p_T b jet of the H_1 candidate
Regressed p_T of the leading-p_T b jet of the H_2 candidate
Regressed p_T of the trailing-p_T b jet of the H_2 candidate
Mass of the H_1 candidate, $M(H_1)$
Mass of the H_2 candidate, $M(H_2)$
Mass of the Higgs pair system, m_{HH}
Transverse momentum of the H_1 candidate, $P_T(H_1)$
Transverse momentum of the H_2 candidate, $P_T(H_2)$
Pseudorapidity separation between the two Higgs candidates, $\Delta\eta(H_1, H_2)$
ΔR distance between two b jets of the H_1 candidate, $\Delta R(H_1(bb))$
ΔR distance between two b jets of the H_2 candidate, $\Delta R(H_2(bb))$
$\lvert\cos(\theta)^*\ (H)\rvert$ in HH frame
$\lvert\cos(\theta)^*\ (b)\rvert$ in H_1 frame
Sum of four b jets' regressed p_T
Transverse momentum of the HH system, $p_T(HH)$
Number of tight b-tags in 3 hightest b-tags
Sum of 3b's resolution scores
Minimal ΔR distance between two b jets, $\text{Min}\lvert\Delta R(bb)\rvert$
Maximum pseudorapidity separation between two b jets, $\text{Max}\lvert\Delta\eta(bb)\rvert$

VBF categories

BDT Reweighter Input variables
Regressed p_T of the leading-p_T b jet of the H_1 candidate
Regressed p_T of the trailing-p_T b jet of the H_1 candidate
Regressed p_T of the leading-p_T b jet of the H_2 candidate
Regressed p_T of the trailing-p_T b jet of the H_2 candidate
Mass of the H_1 candidate, $M(H_1)$
Mass of the H_2 candidate, $M(H_2)$
Mass of the Higgs pair system, m_{HH}
Transverse momentum of the H_1 candidate, $P_T(H_1)$
Transverse momentum of the H_2 candidate, $P_T(H_2)$
Pseudorapidity separation between the two Higgs candidates, $\Delta\eta(H_1, H_2)$
Azimuthal angle separation between the two Higgs candidates, $\Delta\phi(H_1, H_2)$
Mass of the VBF-jet pair system, $M(jj)$
Pseudorapidity separation between the two VBF jets, $\Delta\eta(j1, j2)$
PMMVA score

NN-reweighting variables (ATLAS resolved)

ggF	VBF
1. $\log(p_{\mathrm{T}})$ of the 2^{nd} leading Higgs boson candidate jet 2. $\log(p_{\mathrm{T}})$ of the 4^{th} leading Higgs boson candidate jet 3. $\log(\Delta R)$ between the closest two Higgs boson candidate jets 4. $\log(\Delta R)$ between the other two Higgs boson candidate jets 5. Average absolute η value of the Higgs boson candidate jets 6. $\log(p_{\mathrm{T}})$ of the di-Higgs system 7. ΔR between the two Higgs boson candidates 8. $\Delta\phi$ between jets in the leading Higgs boson candidate 9. $\Delta\phi$ between jets in the subleading Higgs boson candidate 10. $\log(X_{Wt})$ 11. Number of jets in the event 12. Trigger class index as one-hot encoder	1. Maximum di-jet mass out of the possible pairings of the four Higgs boson candidate jets 2. Minimum di-jet mass out of the possible pairings of the four Higgs boson candidate jets 3. Energy of the leading Higgs boson candidate 4. Energy of the subleading Higgs boson candidate 5. Second smallest ΔR between the jets in the leading Higgs boson candidate (out of the three possible pairings for the leading Higgs candidate) 6. Average absolute η value of Higgs boson candidate jets 7. $\log(X_{Wt})$ 8. Trigger class index as one-hot encoder 9. Year index as one-hot encoder (for years inclusive training)

Discriminant observables (CMS resolved)

Chosen observables maximize the analysis sensitivity
Their definition is constrained by available statistics of the bkg model

ggF categories: BDT distribution

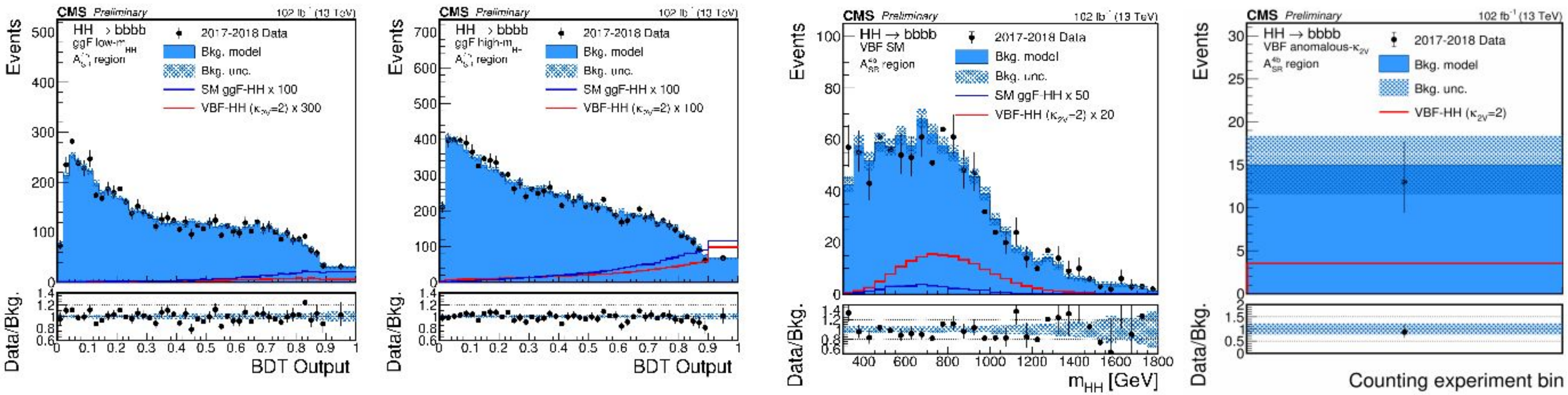
- Bkg model enables to model a BDT discriminant
- Discriminant is trained by category
  - Signal (SM ggF) vs background (bkg model)
  - 16 variables

VBF category 1,2:

- mHH distribution, counting experiment

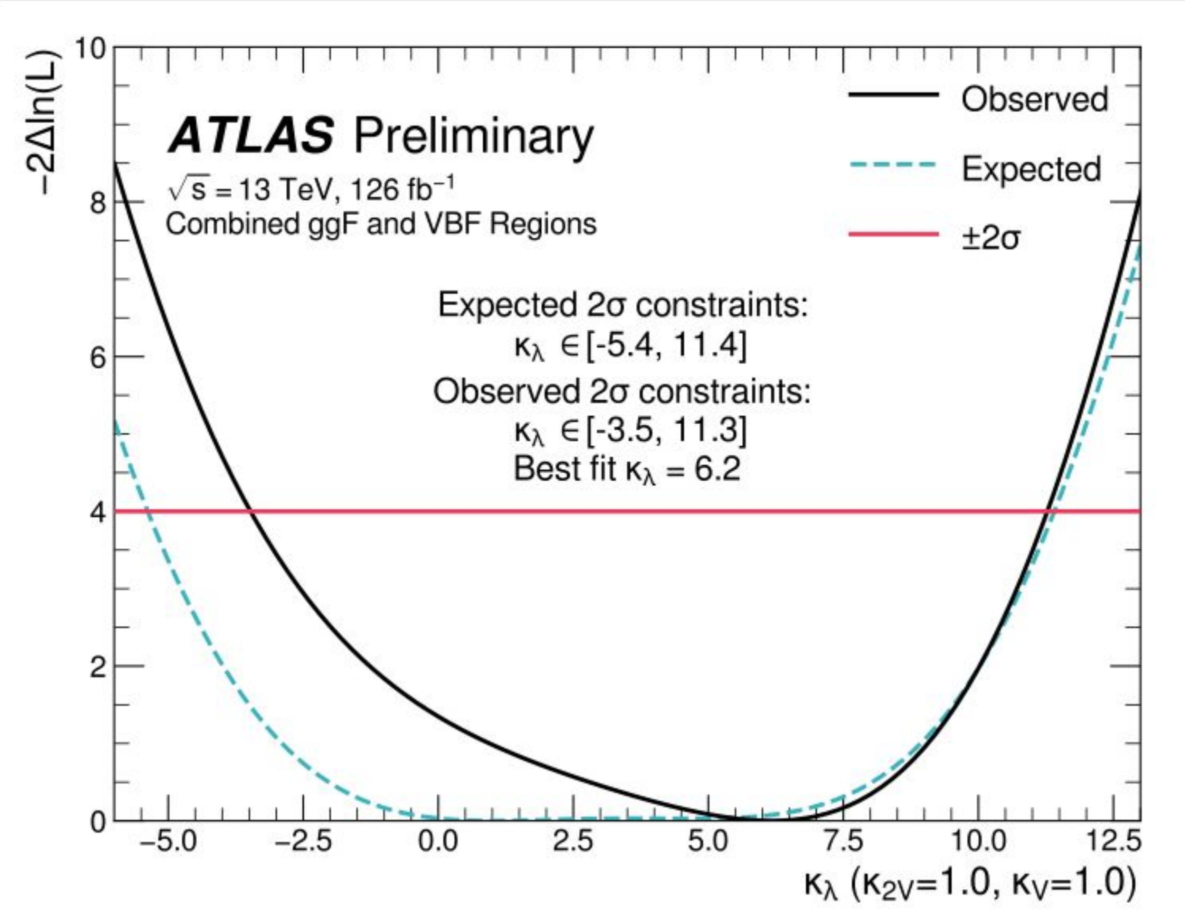
BDT input variables	
M(H1), M(H2), M(HH)	|cos*θ(b)| in H1-frame
pT(H1), pT(H2), pT(HH)	Max|Δηbb|
Scalar sum of 4b's pT	Min|ΔRbb|
Δη(H1,H2), ΔRbb(H1), ΔRbb(H2)	Sum of 3b's resolution scores
|cos*θ(H)| in HH-frame	N. tight b-tags in 3 highest b-tags

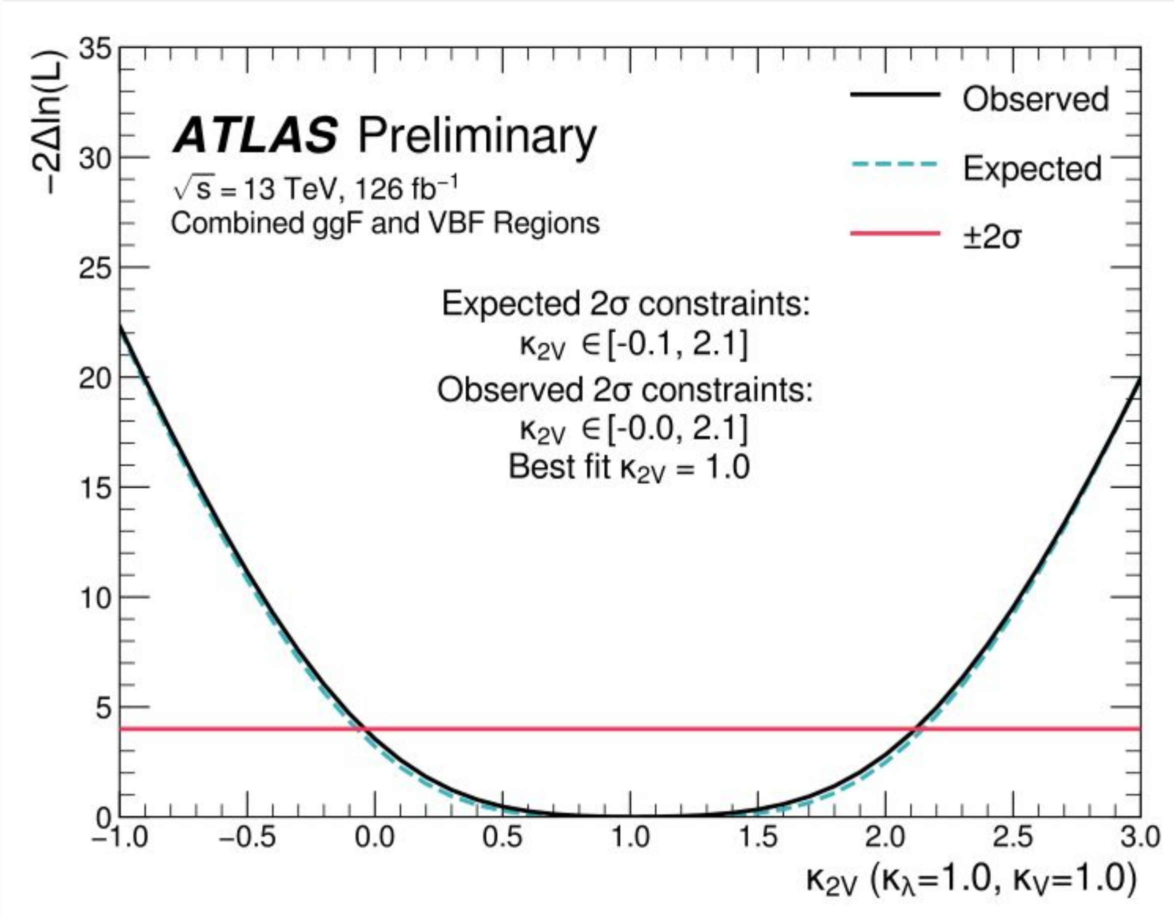
2017-2018 Observables (post-fit)



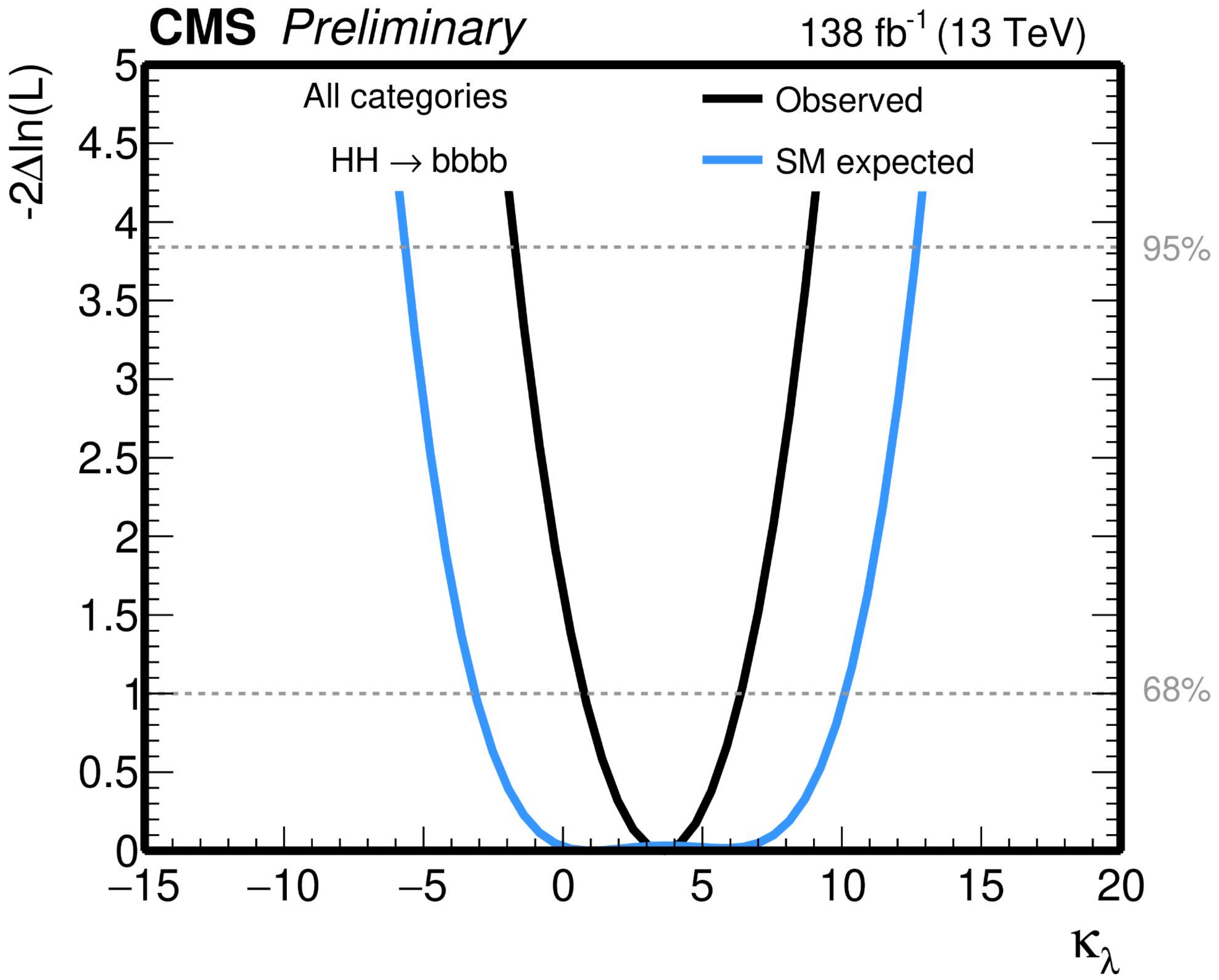
Goodness of fit tests showed that data & bkg model are compatible in all eight observables (p>23%)

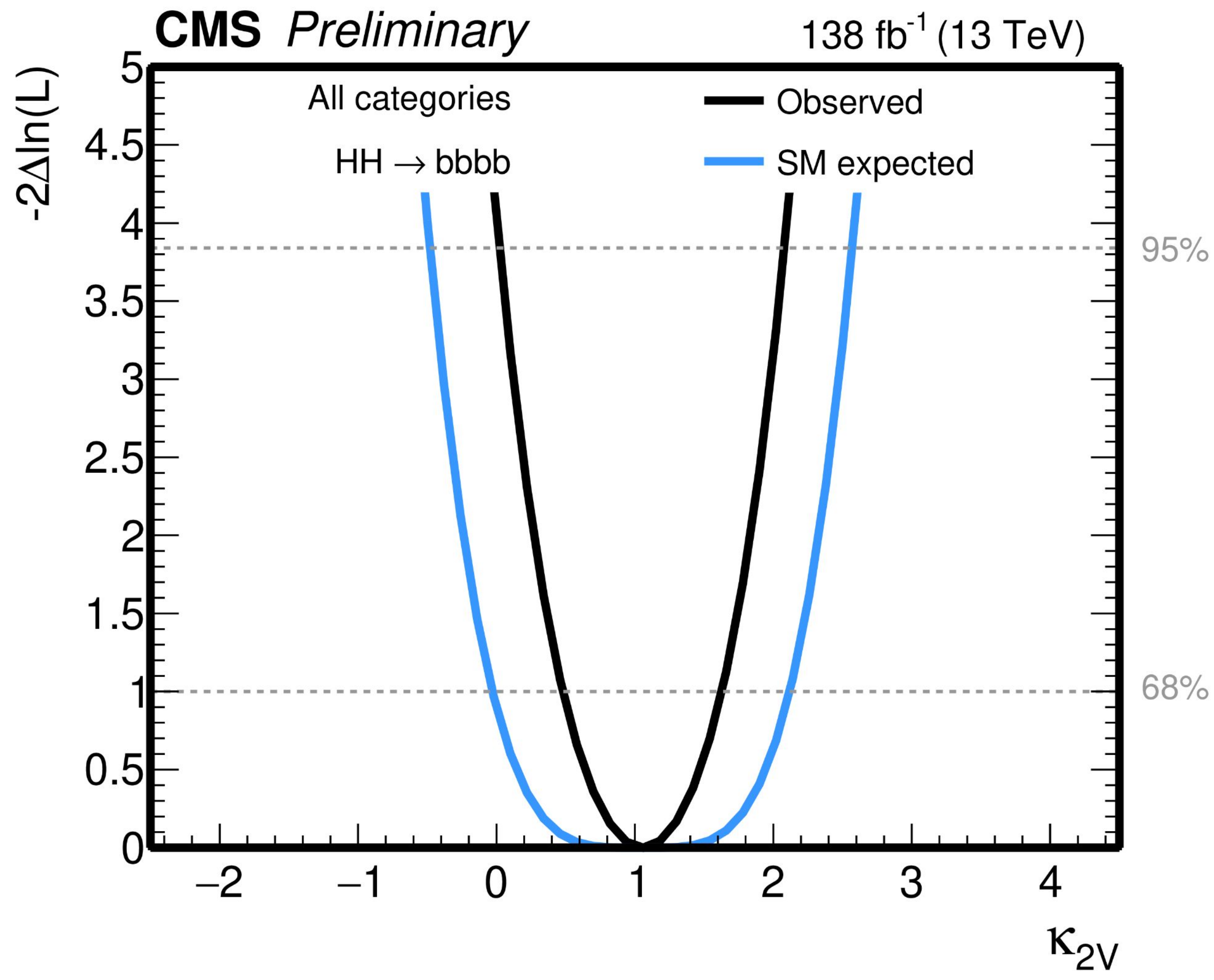
Likelihood scans (ATLAS resolved)





Likelihood scans (CMS resolved)





Likelihood scans (CMS boosted)