Search for Higgs boson pairs in CMS

Daniel Guerrero on behalf of the CMS Collaboration

SM@LHC Conference: Young Scientist Forum

July 11, 2023 - Fermilab, US

Outline

- HH theoretical and experimental overview
- Highlights of the CMS bbbb (resolved) analysis (PRL 129, 081802)
- CMS HH Run-2 combination (Nature 607, 60–68(2022))
- Conclusions and prospects

Related talks tomorrow:

Pier Paolo Giardino (theory)

Maximiliam Swiatlowski (experiment)

Non-resonant Higgs boson pairs (HH) at the LHC

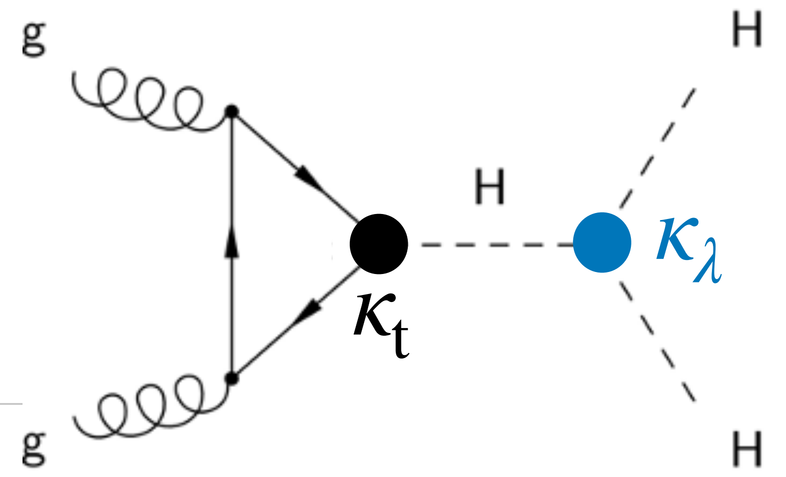
SM production gives direct access to the self-coupling (λ), thus the reconstruction of the Higgs potential

It's an elusive process at the current LHC datasets, but let's see how far we can go!

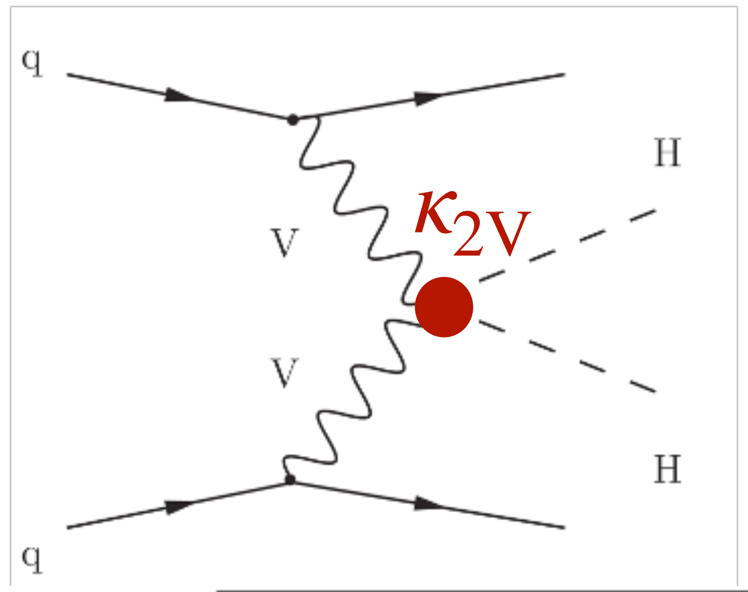
New physics may modify Higgs couplings (or activate new vertices)

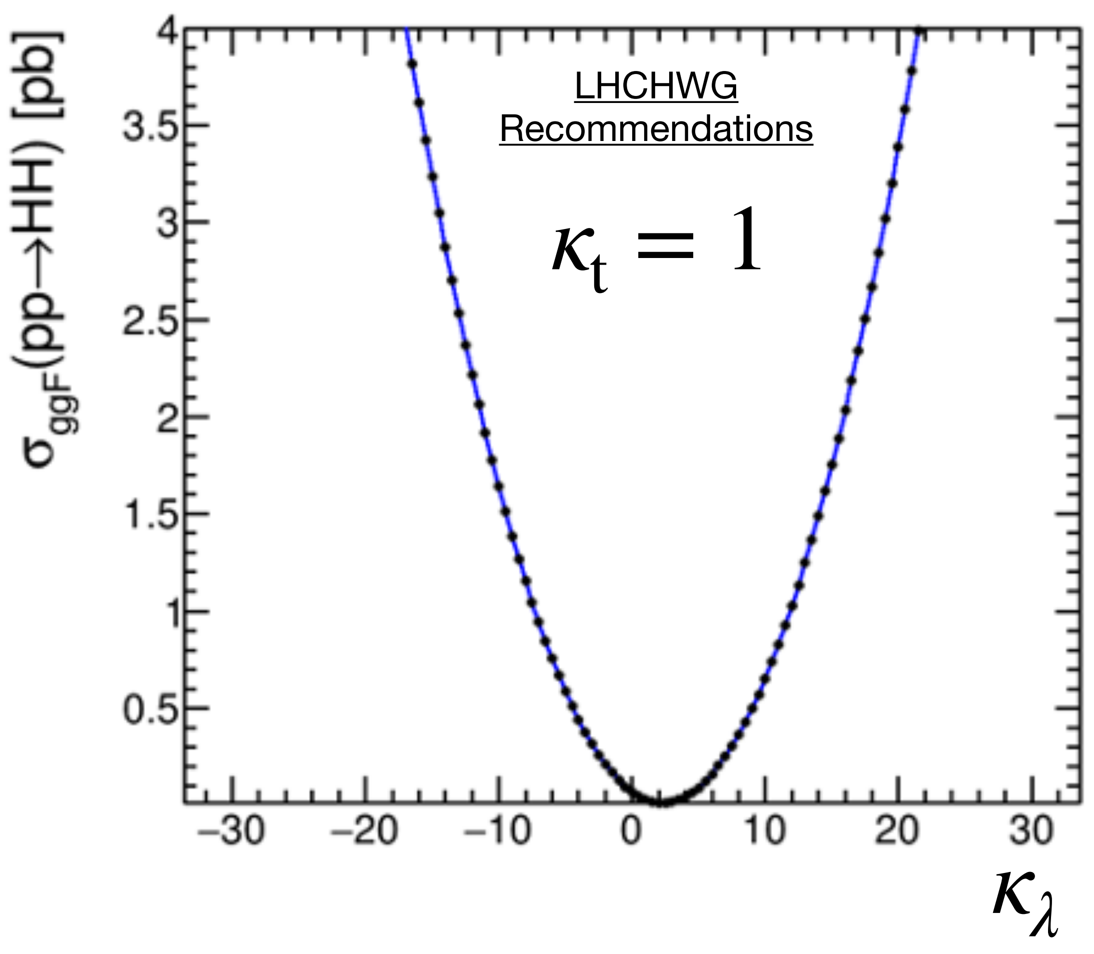
- Anomalous couplings are studied w.r.t. the SM using a κ-framework, e.g. $\kappa_\lambda = \lambda/\lambda_{SM}$
- Enhancement in cross sections and kinematics are predicted

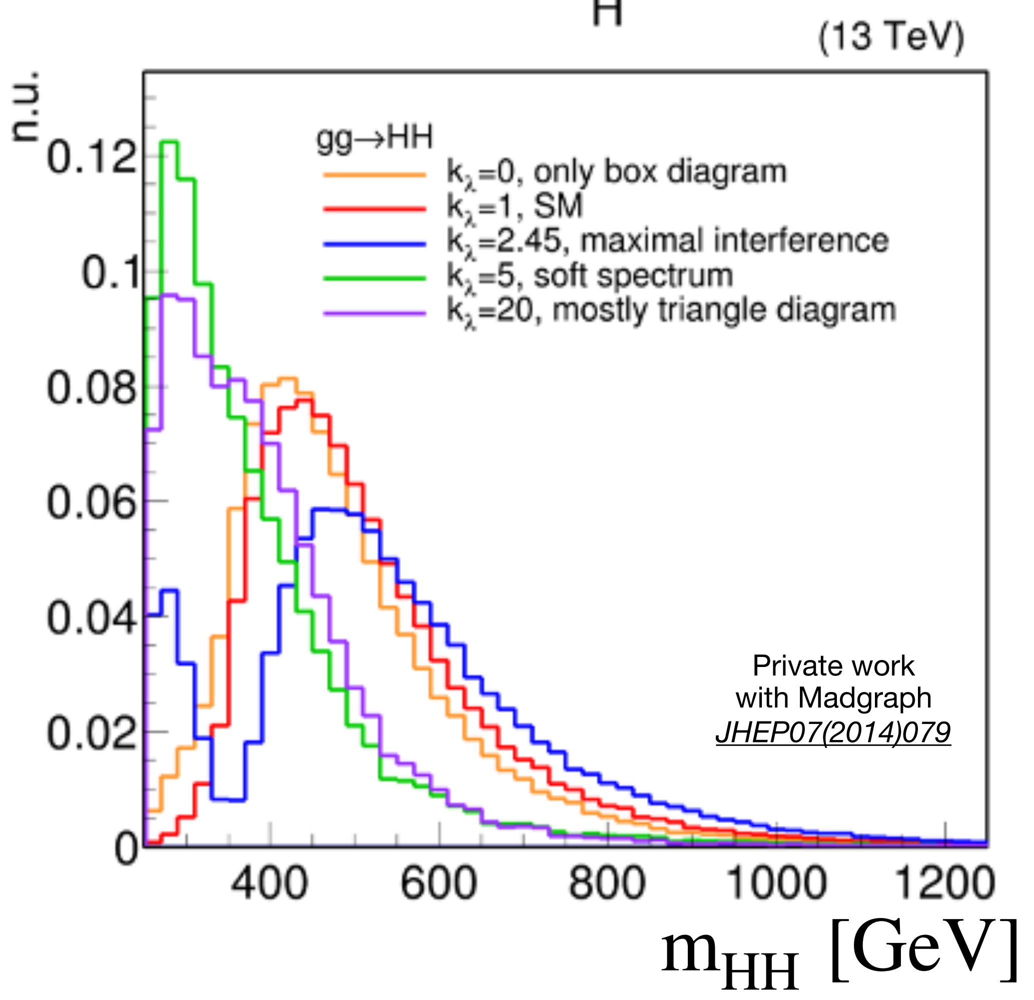
BSM self-couplings in ggF HH mode

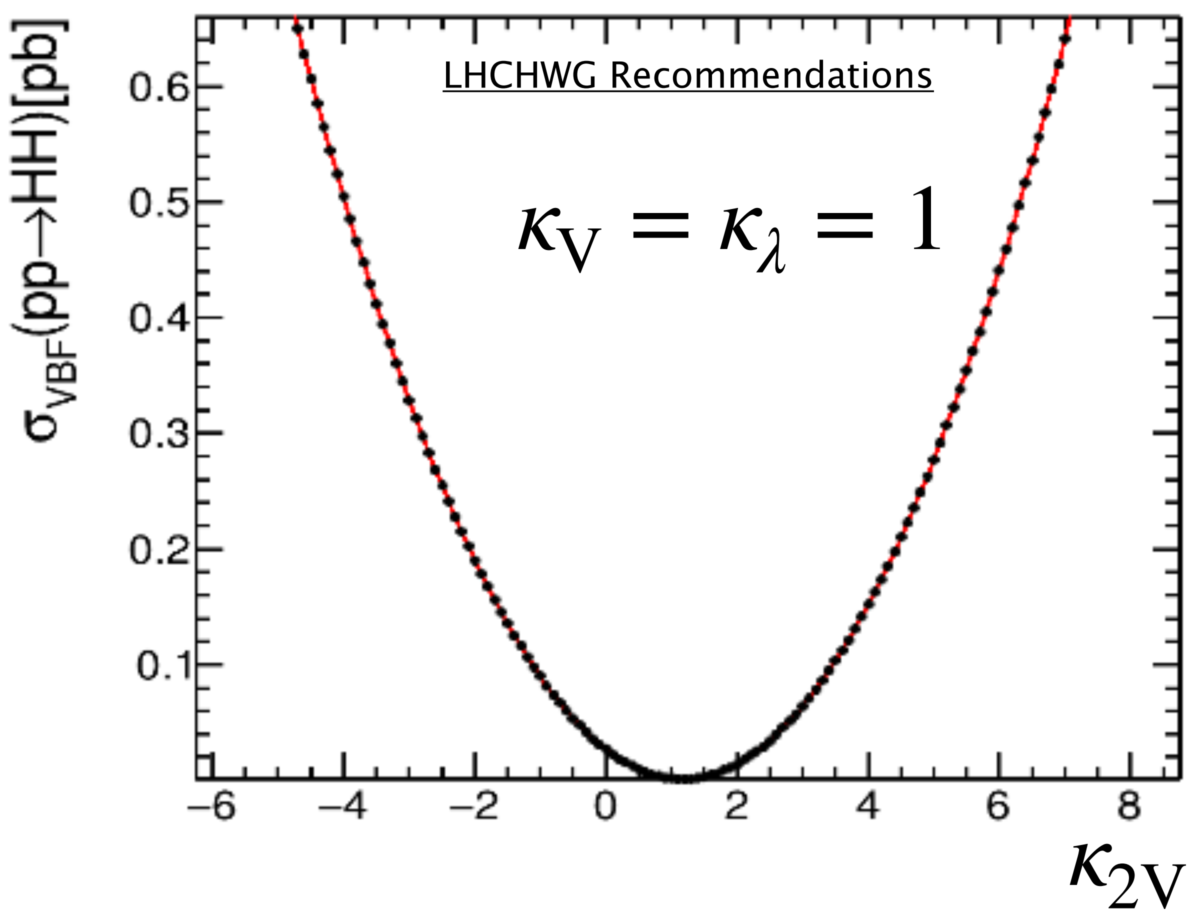


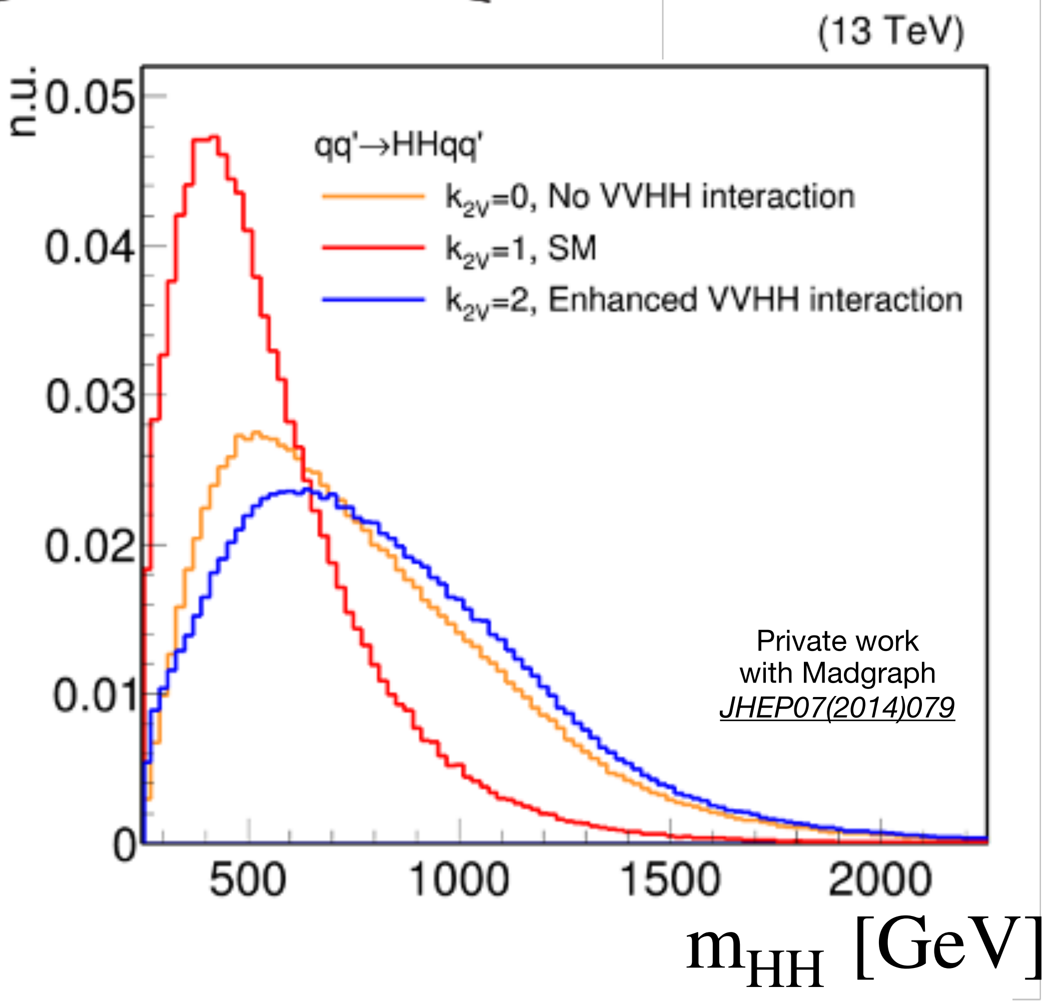
BSM VVHH couplings in VBF HH mode









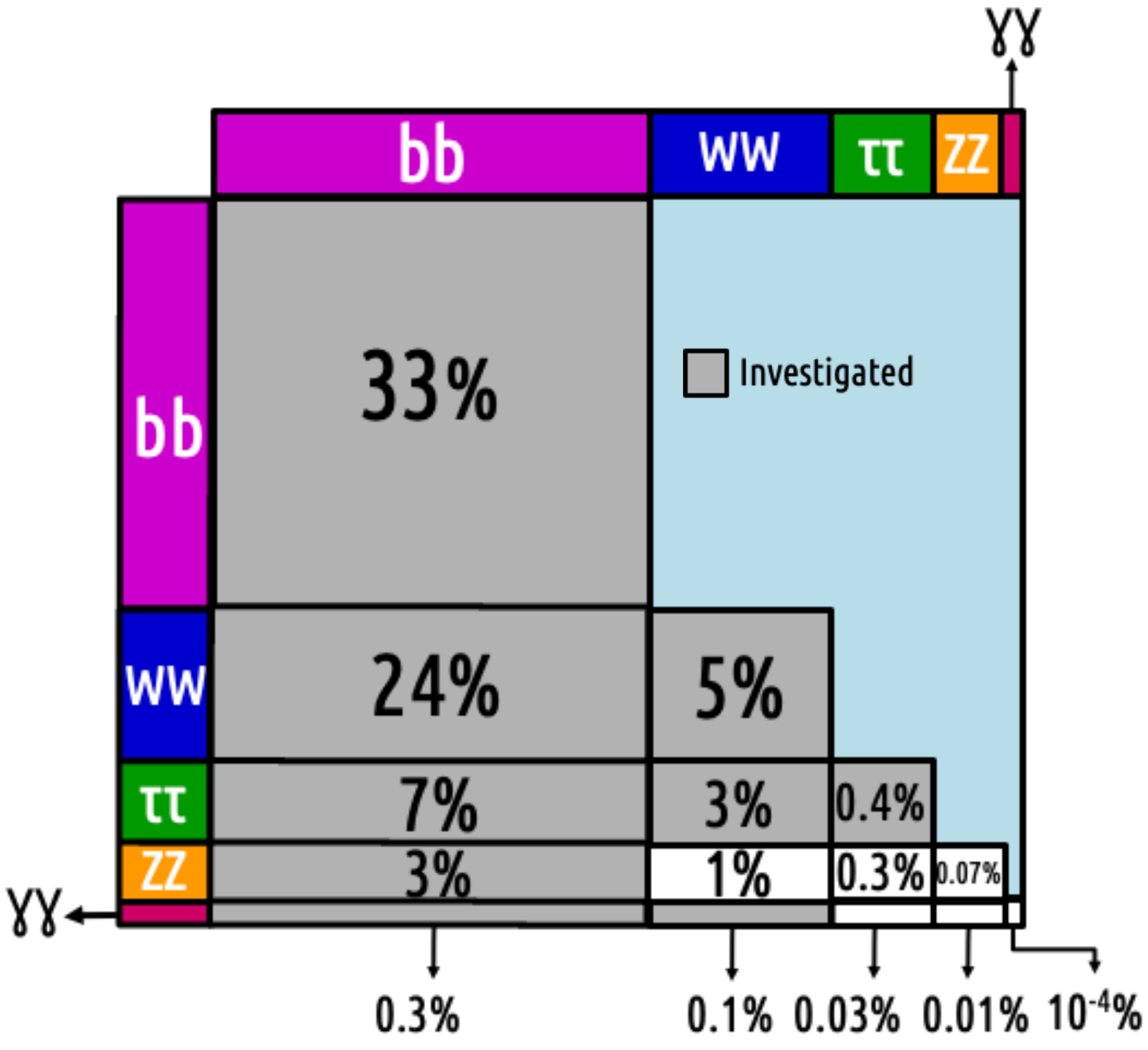


Sensitivity to discovery BSM physics already at the LHC!

How do we search for it at the LHC?

Rich variety of decay channels to explore!

Run-2 searches combine five Higgs boson decays: bb, WW, $\tau\tau$, ZZ and $\gamma\gamma$



Branching fractions (m_H = 125 GeV)

No 'golden' HH channel

Three important and sensitive channels:

- bbbb: Largest rate but large and challenging QCD background
- bb$\tau\tau$: Medium rate and background ($t\bar{t}$, DY and QCD)
- bb$\gamma\gamma$: Small rate but small background

Covered in this talk

Other complementary channels:

bbWW, bbZZ

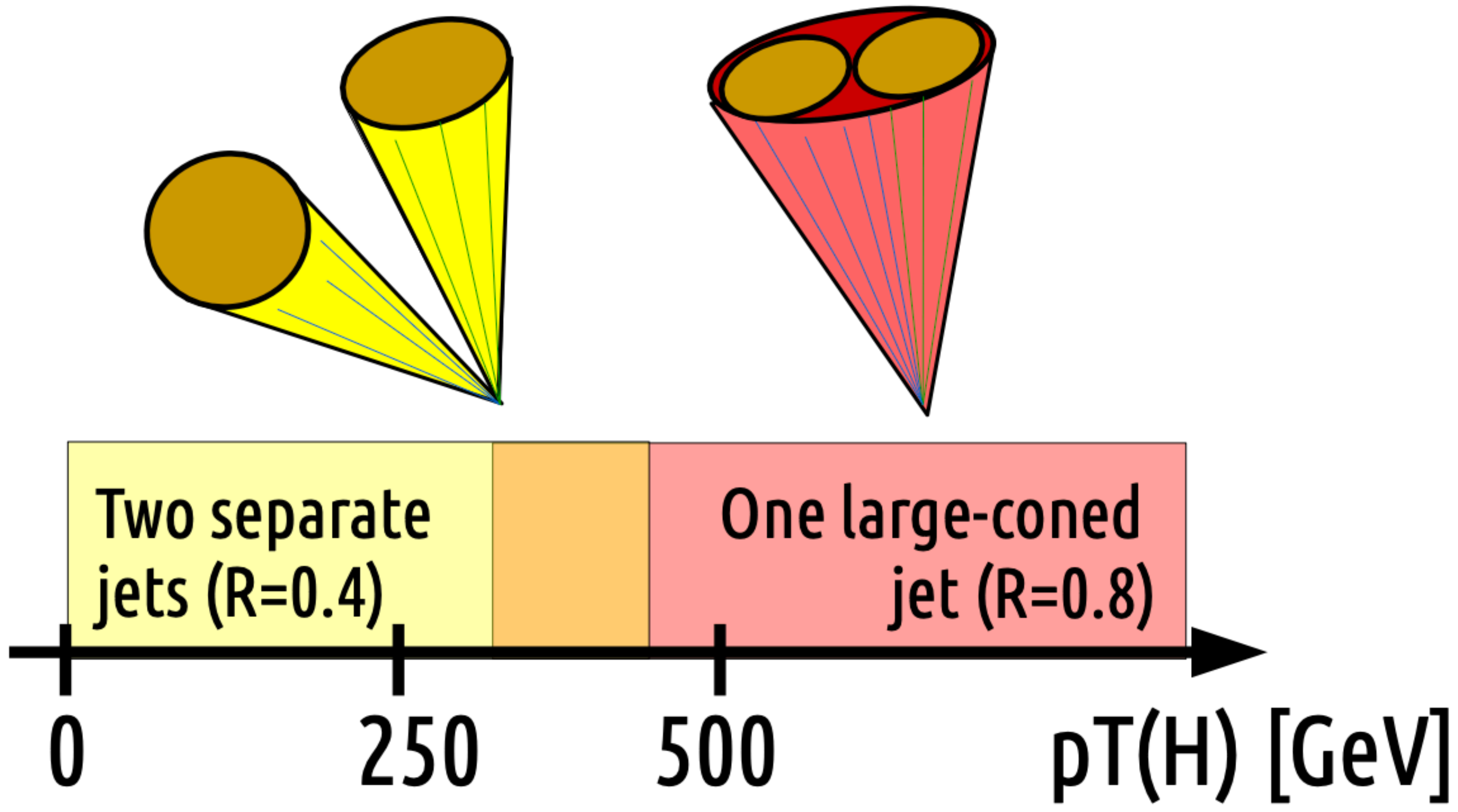
Multilepton (WWWW, WW$\tau\tau$, $\tau\tau\tau\tau$) and WW$\gamma\gamma$

More in additional material

HH→bbbb channel

Benefits from the largest HH branching ratio, but . . .
searching for a signal is ferociously challenged by the overwhelming amount of multi-jet background

The $H \rightarrow bb$ experimental fingerprint depends on Higgs pT



Experimental Challenges

- b jet or H jet identification from large g/light/c jet bkg
- Complex trigger algorithms
  - Depends on L1 seed, PV finding, HLT tracking, jet cal., b-tagging
  - Constrained by L1 rate, rate and CPU limit at HLT and rate
- Higgs boson reconstruction
  - Large jet combinatorics
  - Missing energy from semi-leptonic B hadron decays
- Precise bkg modeling and good bkg rejection are needed

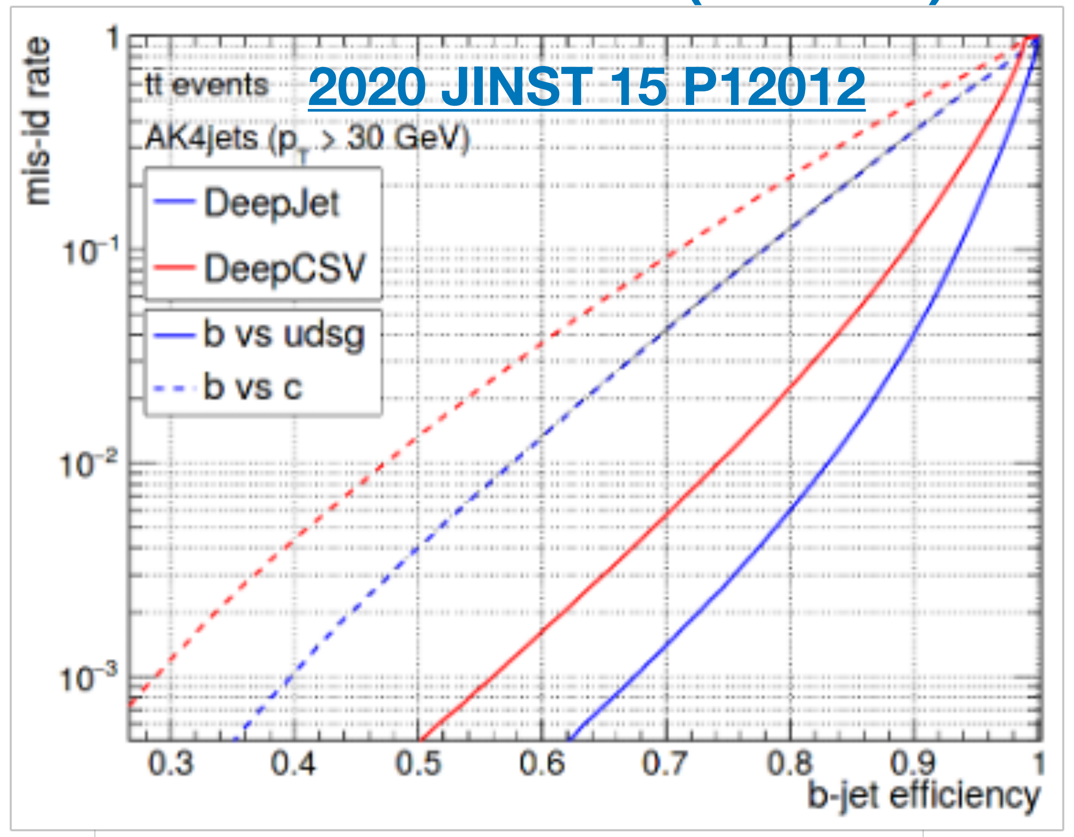
HH→bbbb (resolved): Strategy

PRL 129, 081802

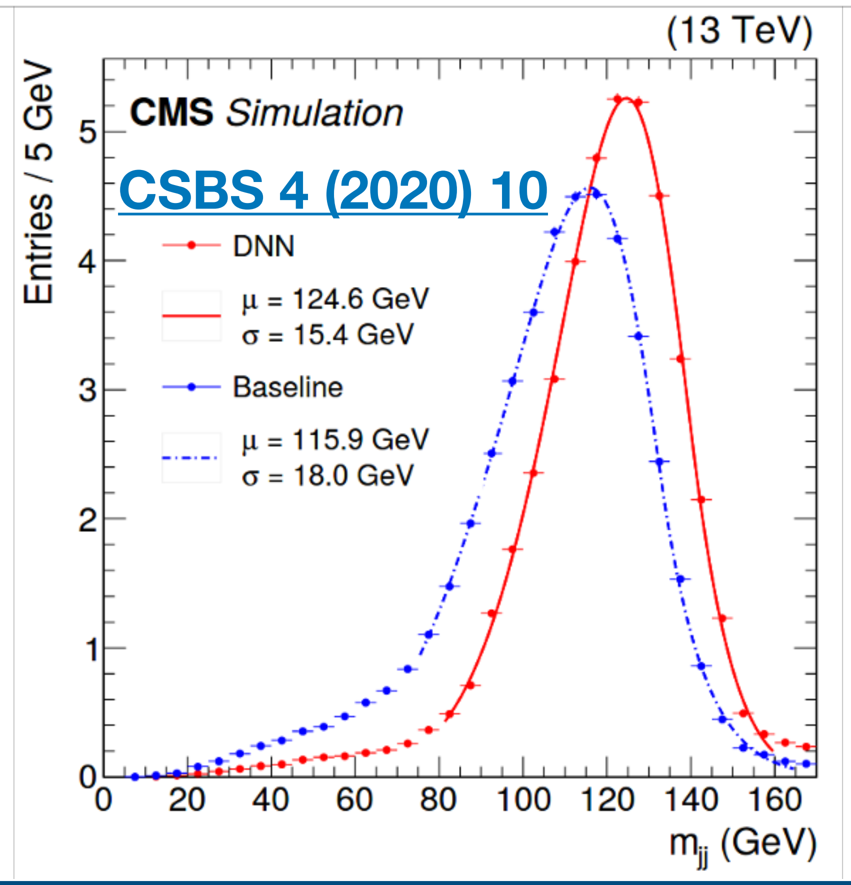
The analysis is carried out using the CMS full Run-2 dataset (L=138 fb^{-1})
It combines several methods to maximize the analysis sensitivity

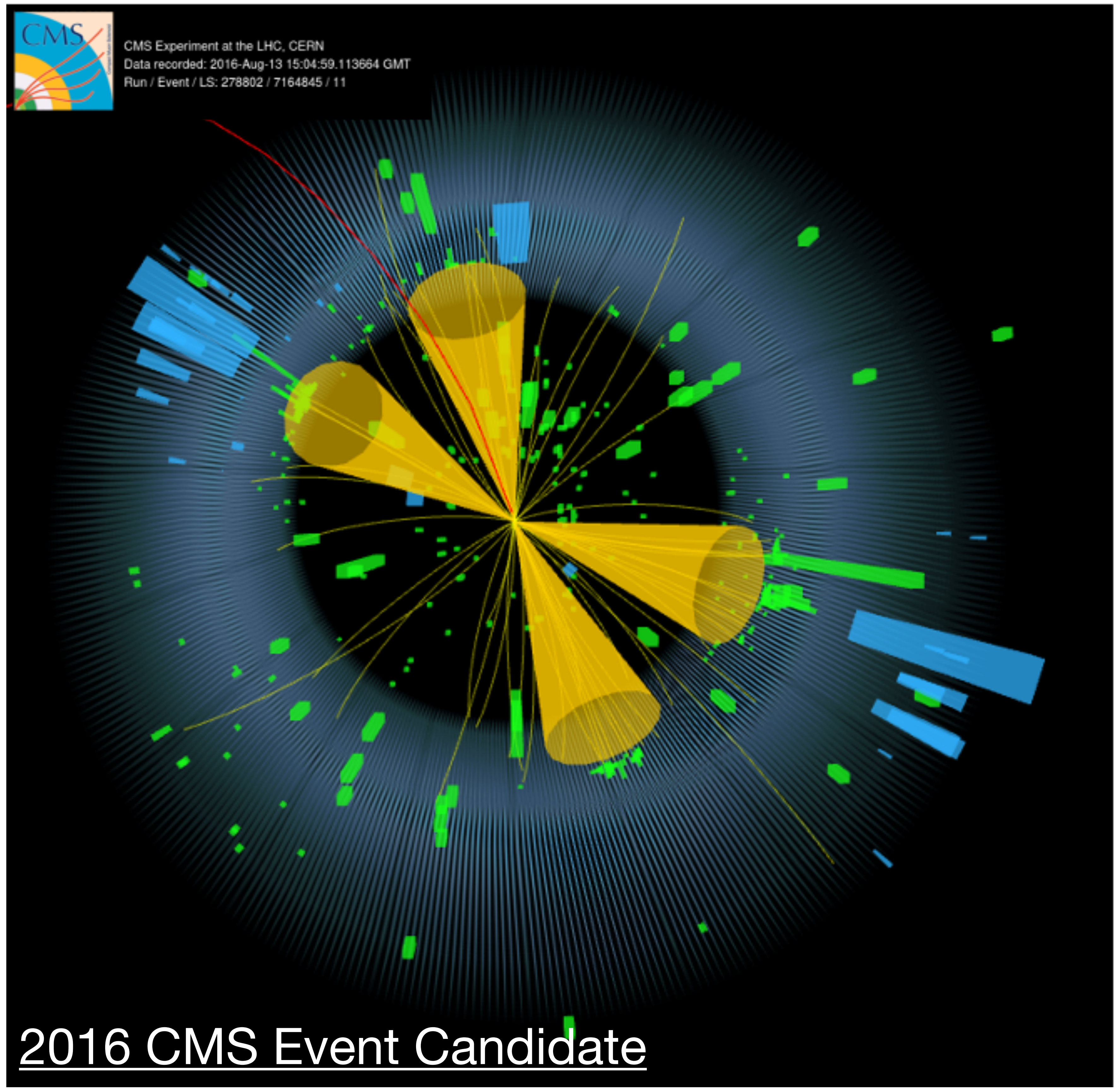
Advanced jet algorithms based on
A deep neural network (DNN)

DeepJet b-tagging algorithm



b-jet energy regression





2016 CMS Event Candidate

Developed using
novel analysis methods

- Novel jet pairing for Higgs candidate identification
- Advanced ggF and VBF categorization
- Powerful bkg modeling using a ML method

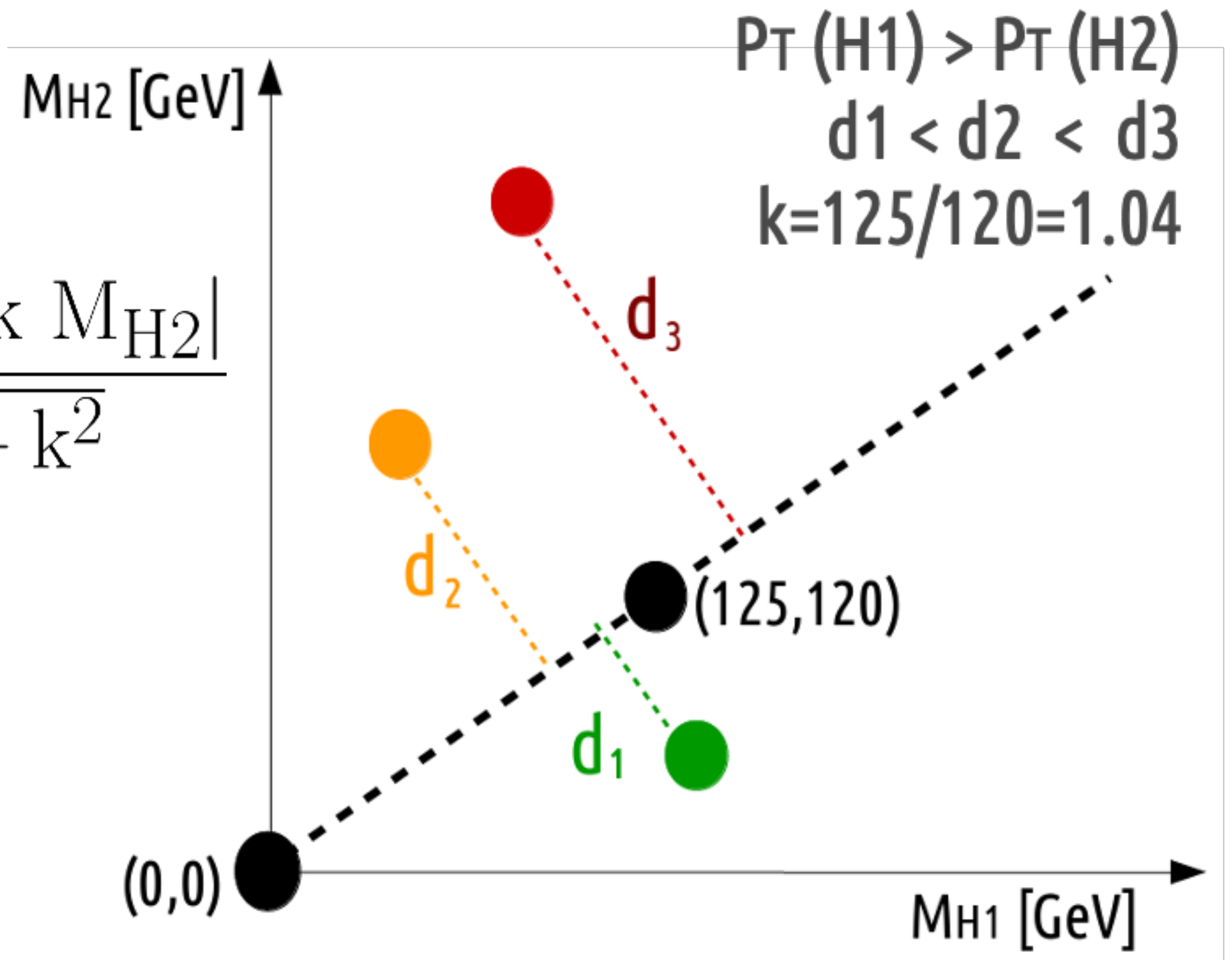
More on boosted results in back-up

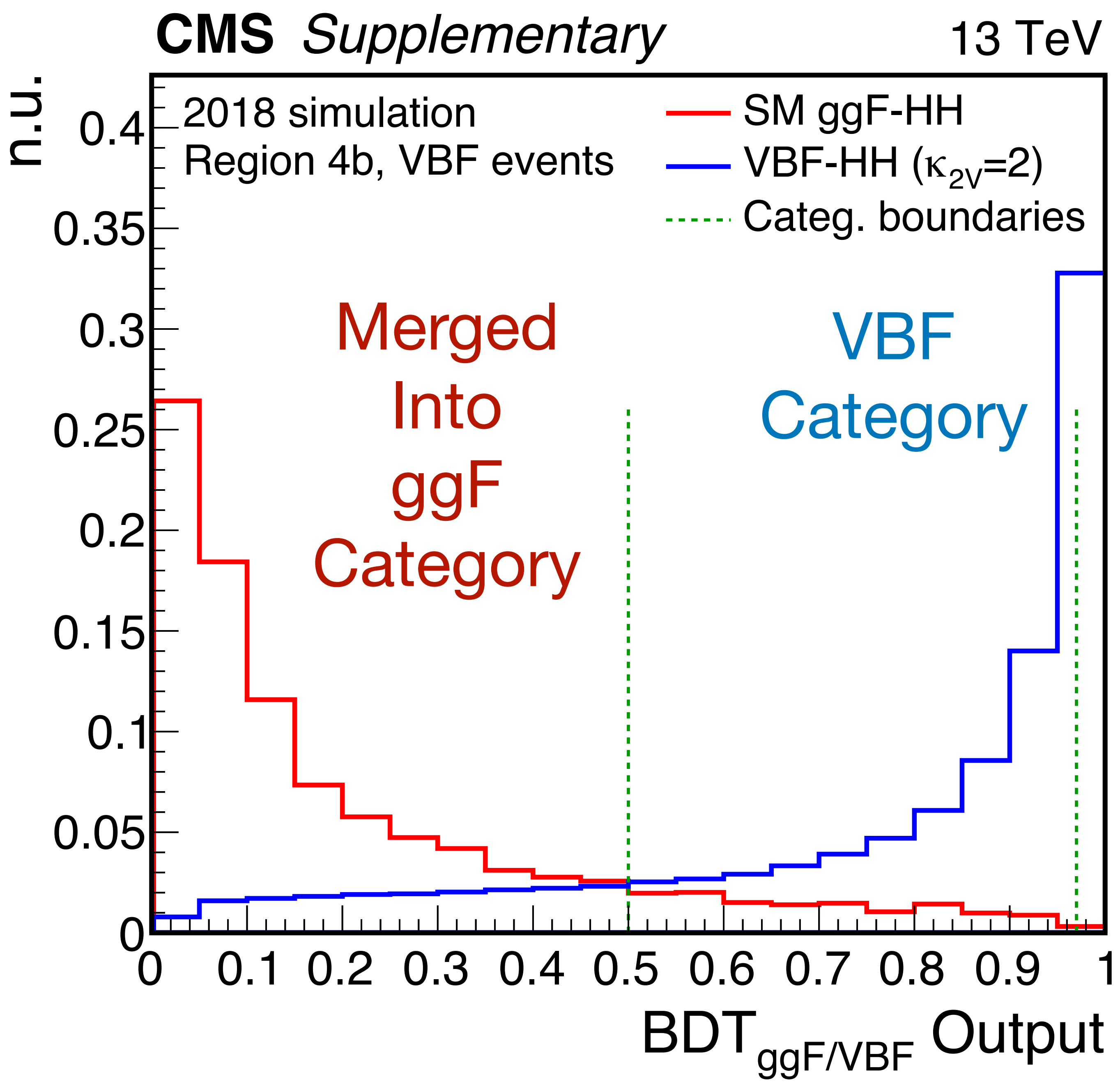
HH→bbbb (resolved): Event Selection

PRL 129, 081802

- Multijet triggers with 4 central jets, 3 b-tagged jets
- Four b-jets: highest b-tagged jets (DeepJet Medium WP)
- Higgs candidates identified with d-pairing algorithm
  - If $|d_1 - d_2| \geq 30$ GeV: d1 (closest to diagonal line)
  - Else: d_1 or d_2 based largest Higgs p_T in 4-jet CM frame
- Production mode separation:
  - Presence of VBF jets
  - $\mathrm{BDT}_{\mathrm{ggF/VBF}}$ discriminant
- Two ggF subcategories:
  - Low-m_{HH} , m_{HH} < 450 GeV
  - High-m_{HH} , m_{HH} > 450 GeV
- Two VBF subcategories:
  - SM-like VBF: 0.5 < $\mathrm{BDT}_{\mathrm{ggF/VBF}}$ < 0.97
  - Anomalous-κ_{2V} VBF: 0.97 < $\mathrm{BDT}_{\mathrm{ggF/VBF}}$ < 1.0

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





2016 and 2017-2018 datasets are analyzed separately and combined for final result

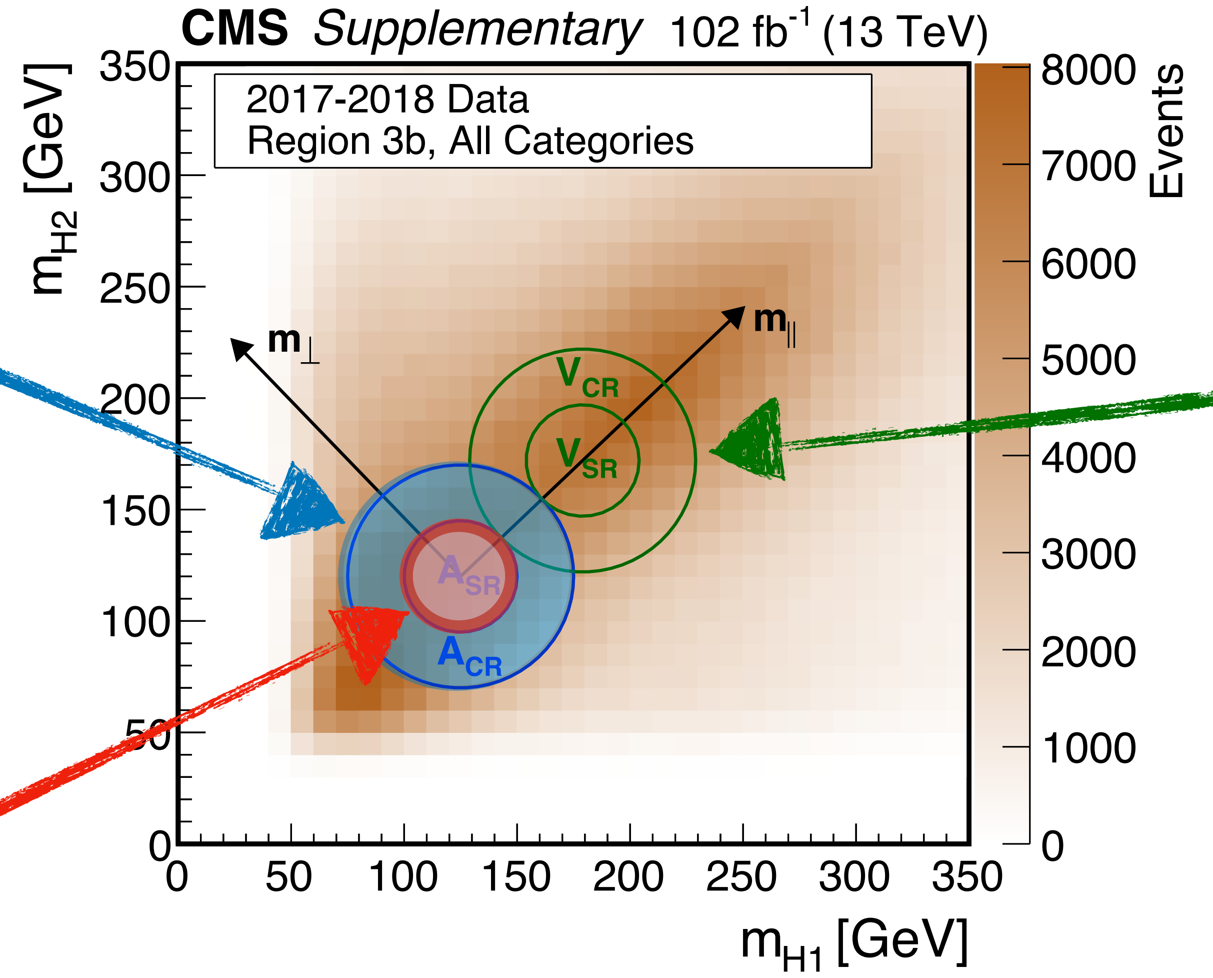
HH→bbbb (resolved): Background model

PRL 129, 081802

Data-driven method to derive ‘4b’ multijet background estimate using ‘3b’ region data
3b-to-4b shape differences are corrected with BDT re-weighting

The bkg model is built using CR(3b) and CR(4b) data
Normalization: Transfer factor
- ggF, it considers ‘parallel’ mass ($m_{||}$) dependency
- VBF, constant

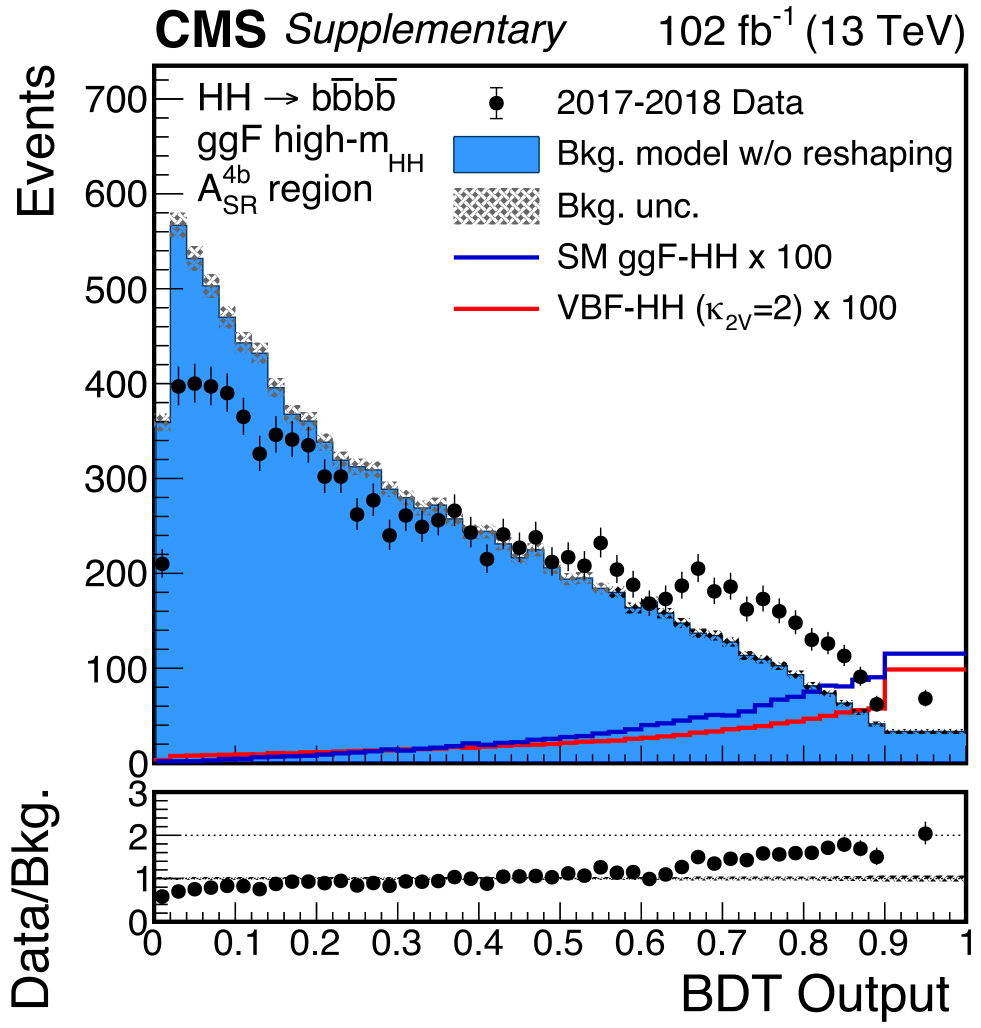
Shape variables mis-modeling: weights from BDTReweighter

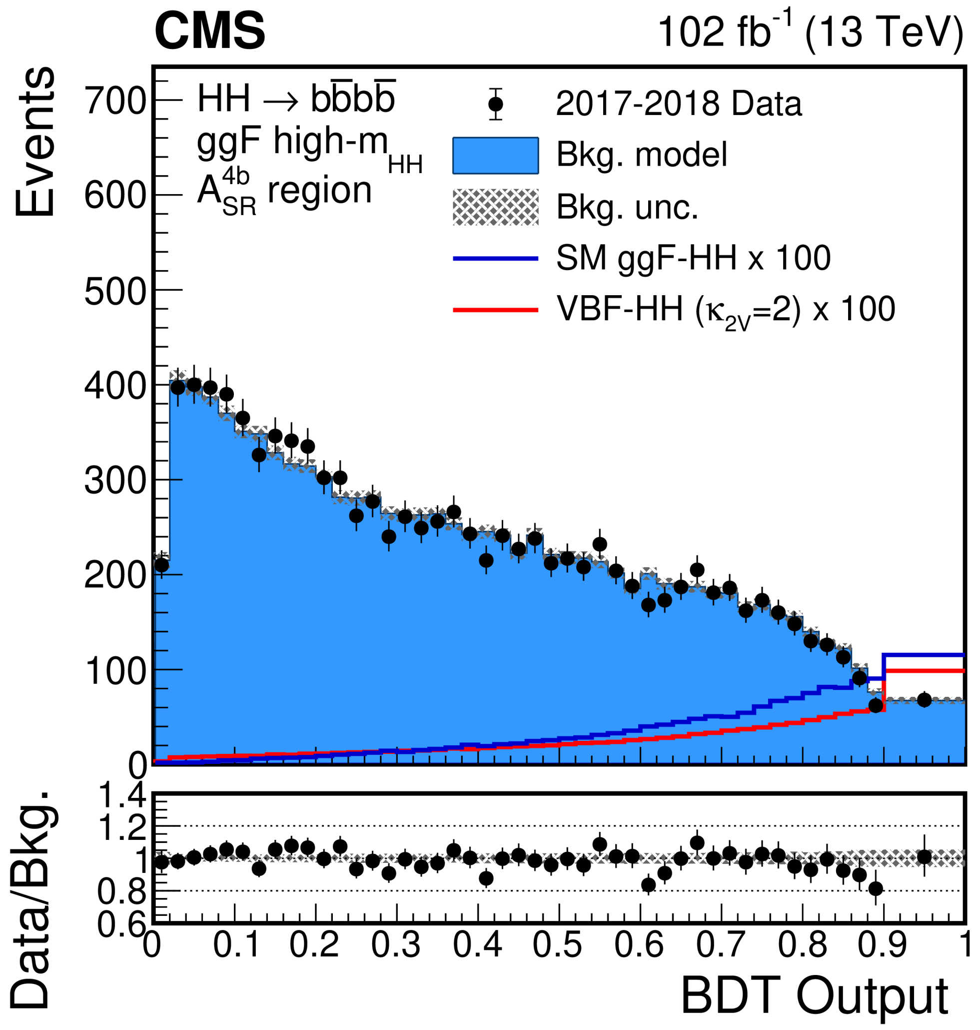


Data/model closure is fully verified in validation region

SR(4b) bkg model = Model applied to SR(3b) data
Normalization: Transfer factors from CR
Shape: SR(3b) distribution reshaped by CR BDTReweighter

Performance in SR(4b) region





The bkg model uncertainties are the dominant systematics in the analysis

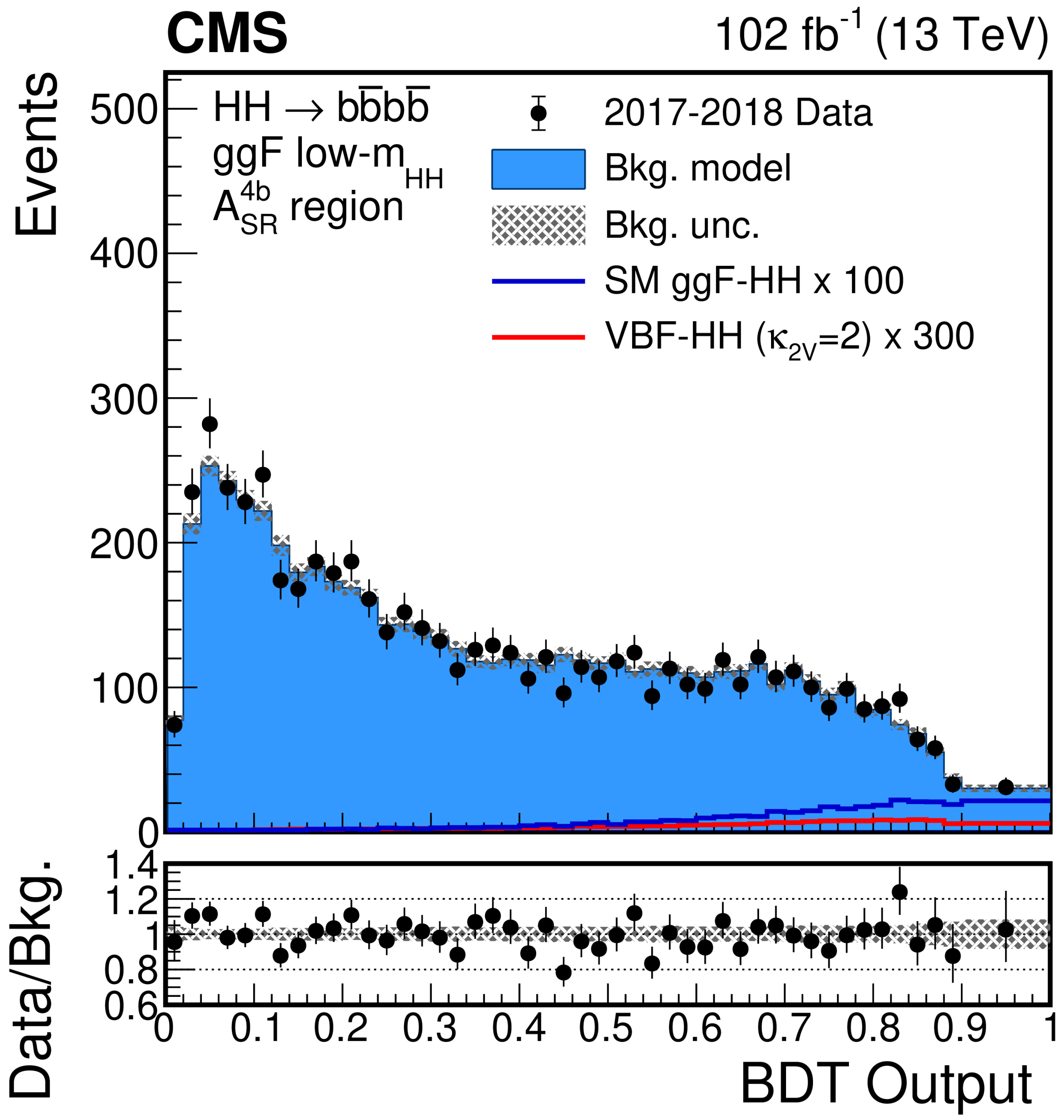
HH→bbbb (resolved): Signal extraction

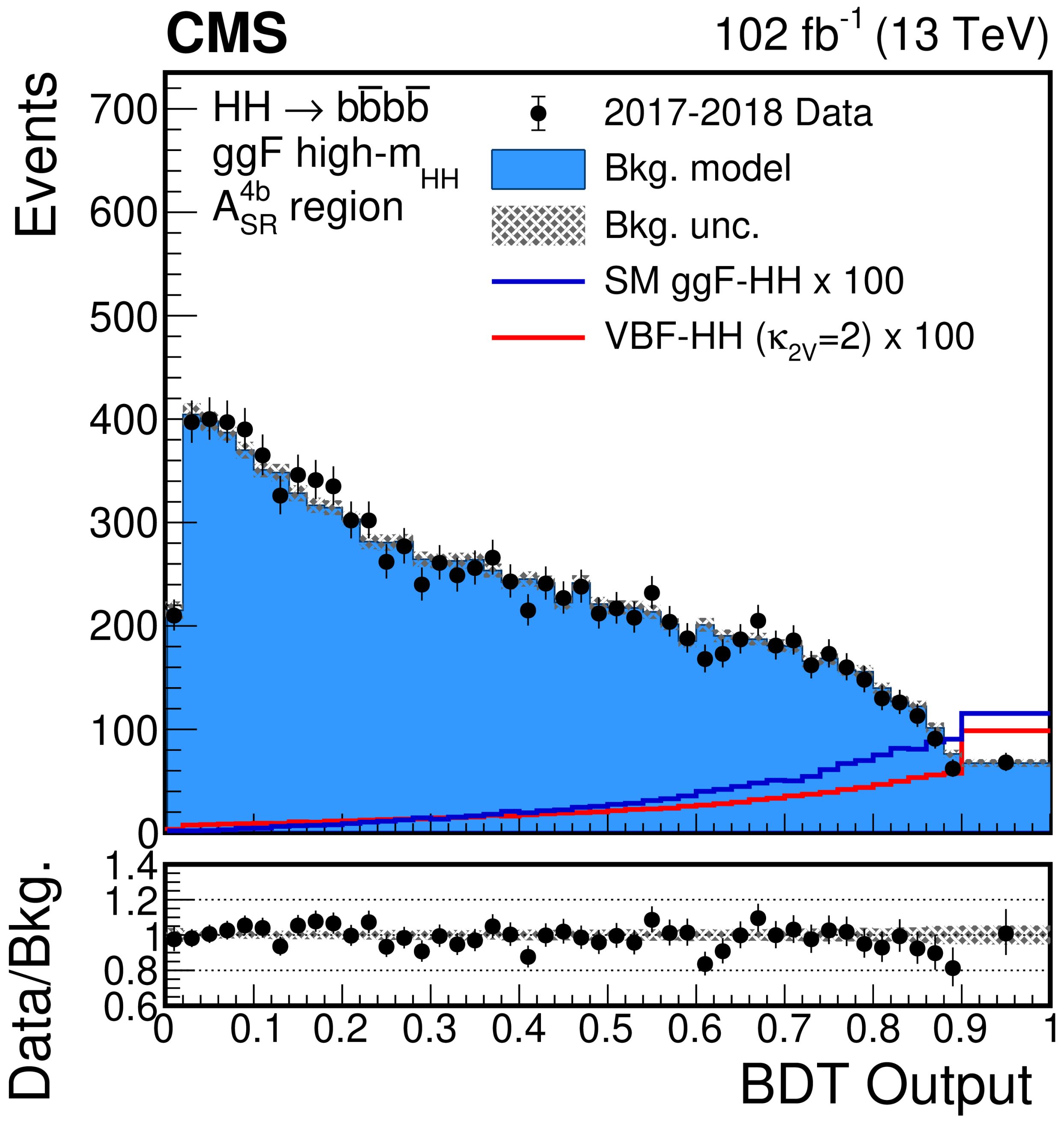
PRL 129, 081802

Chosen observables maximize the analysis sensitivity

ggF Categories: BDT output

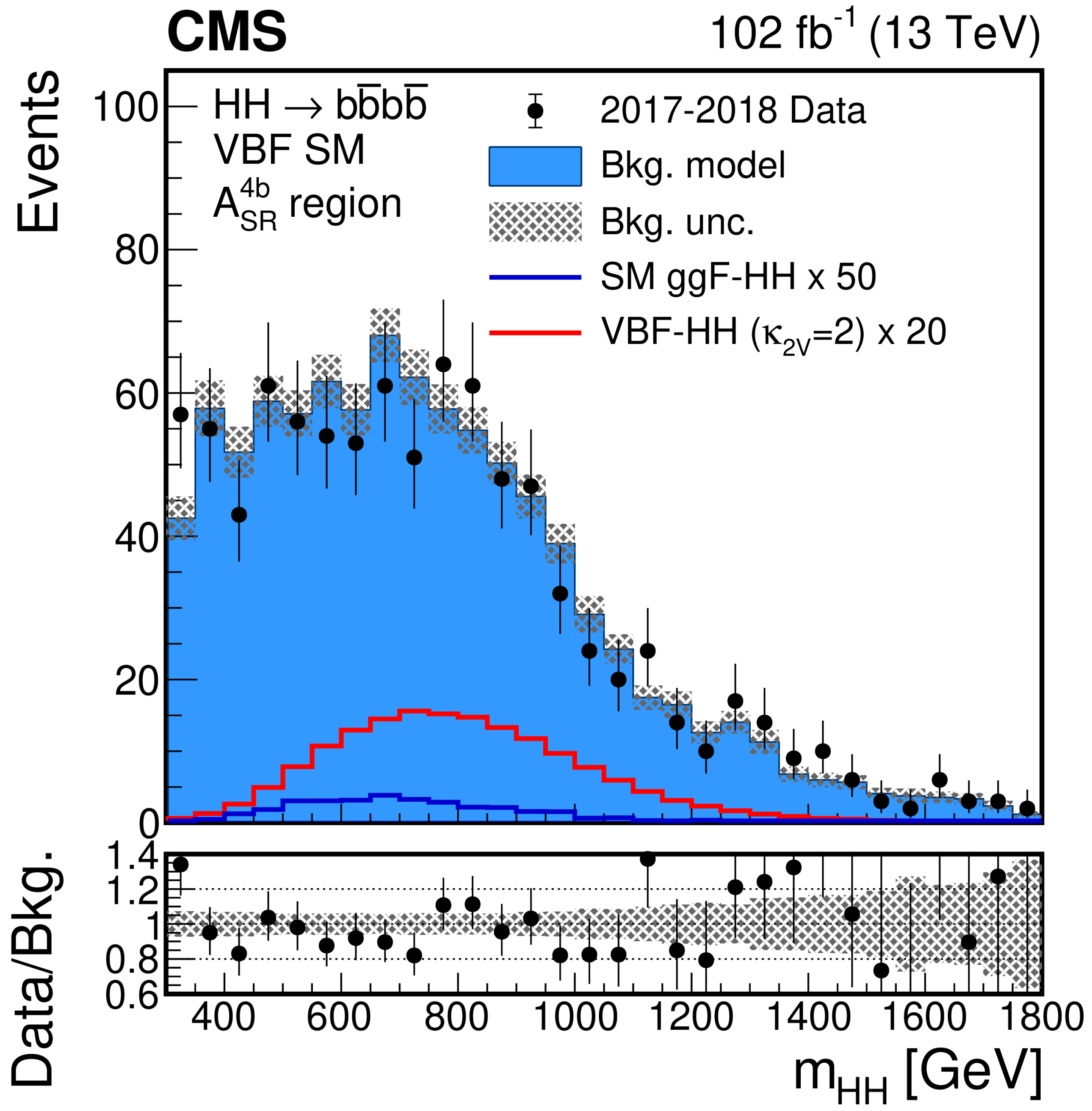
- Bkg model enables to model ML discriminant
- BDT is trained by category using 16 variables





VBF Categories

- SM-like: m_{HH} distribution
- Anomalous-κ_{2V}: Counting experiment



Data and model are compatible in all analysis observables

More in additional material

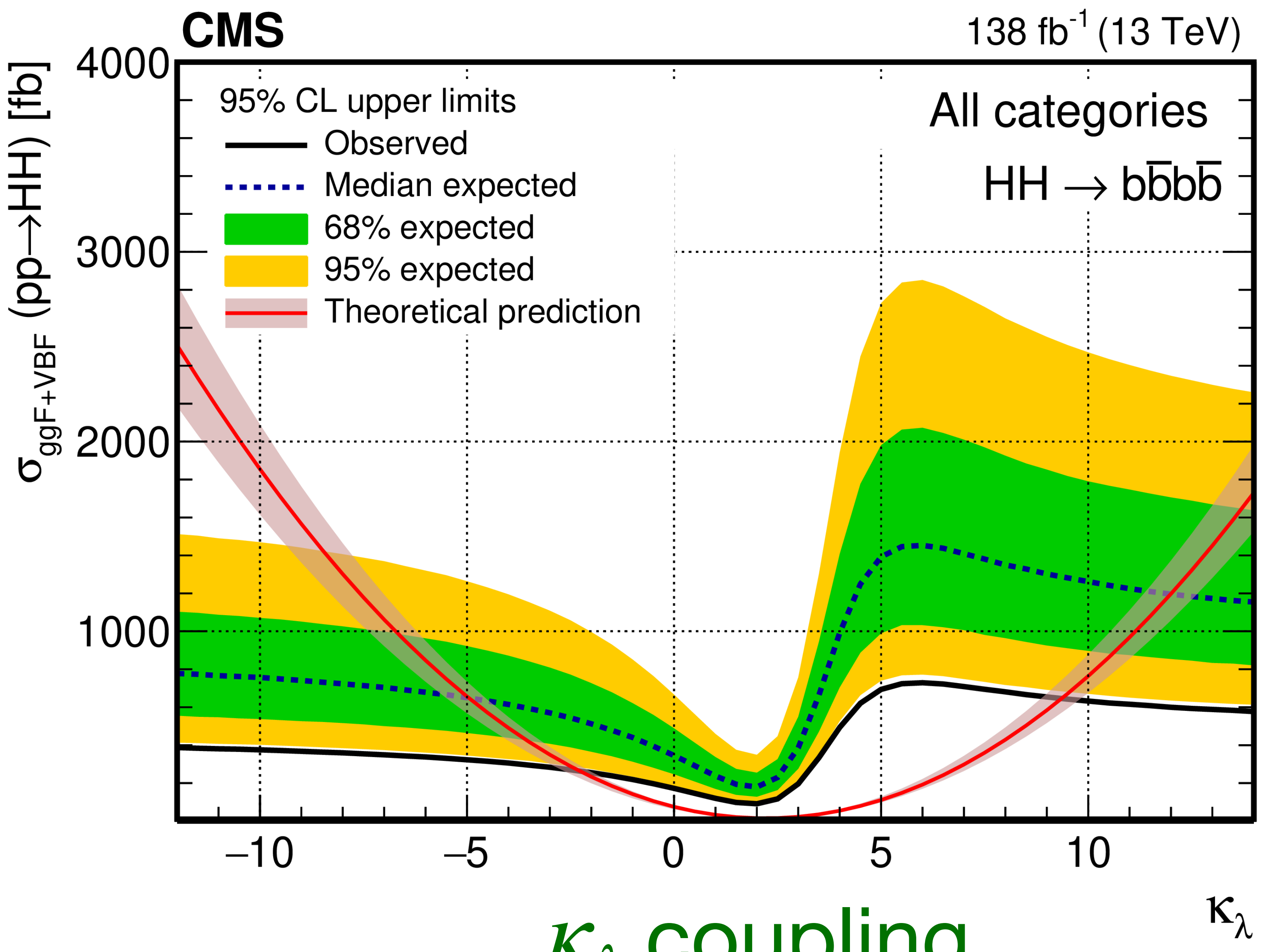
HH→bbbb (resolved): Results

PRL 129, 081802

No excess of data is observed relative to the background expectation

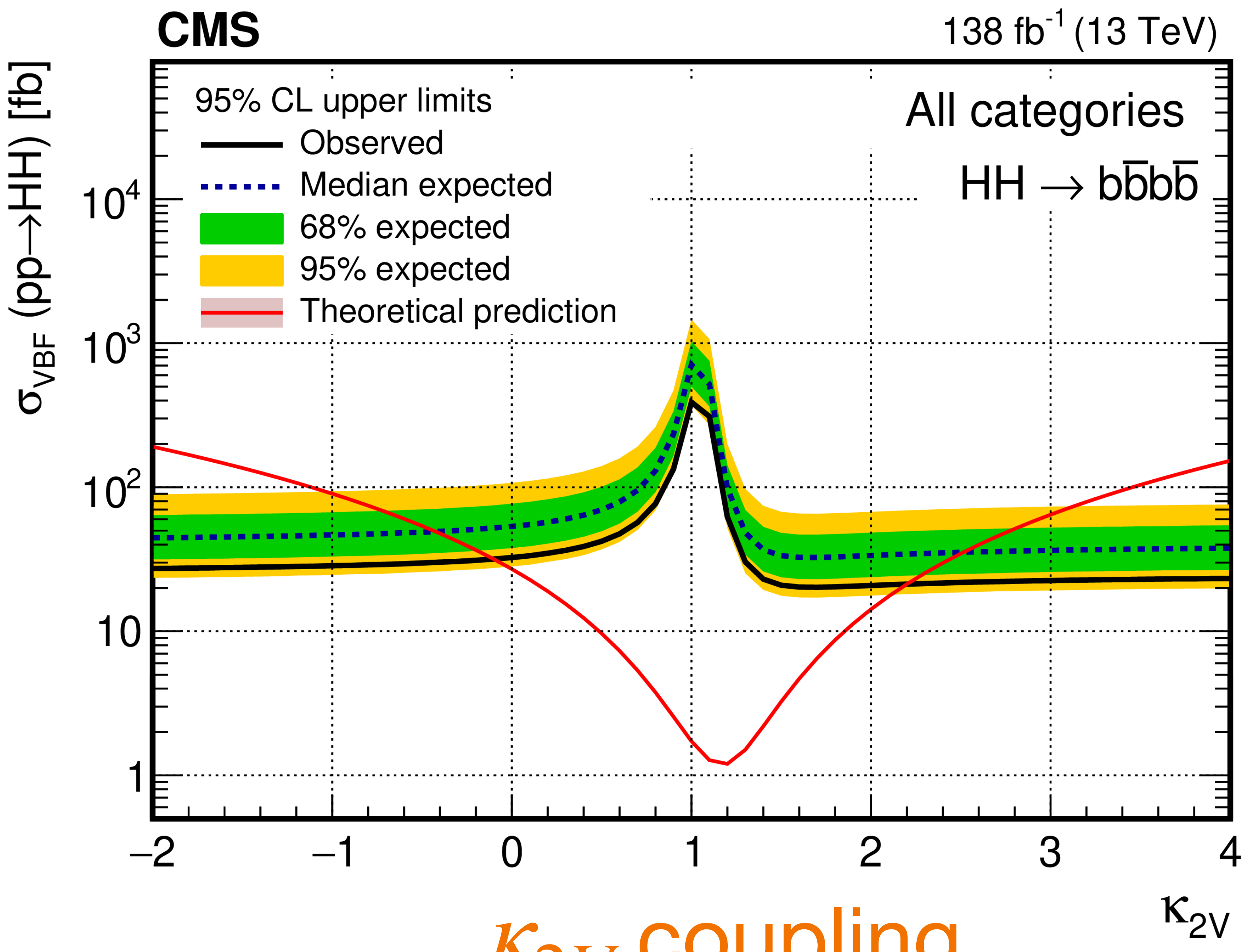
95% CL upper limits are set on the SM and BSM production cross sections

- Constraint on SM: Obs. (exp.) limit on σ/σ_{theory} is 3.9 (7.8)
- Constraints on anomalous Higgs couplings



κ_λ coupling

[-2.3, 9.4] ([-5.0,12.0])



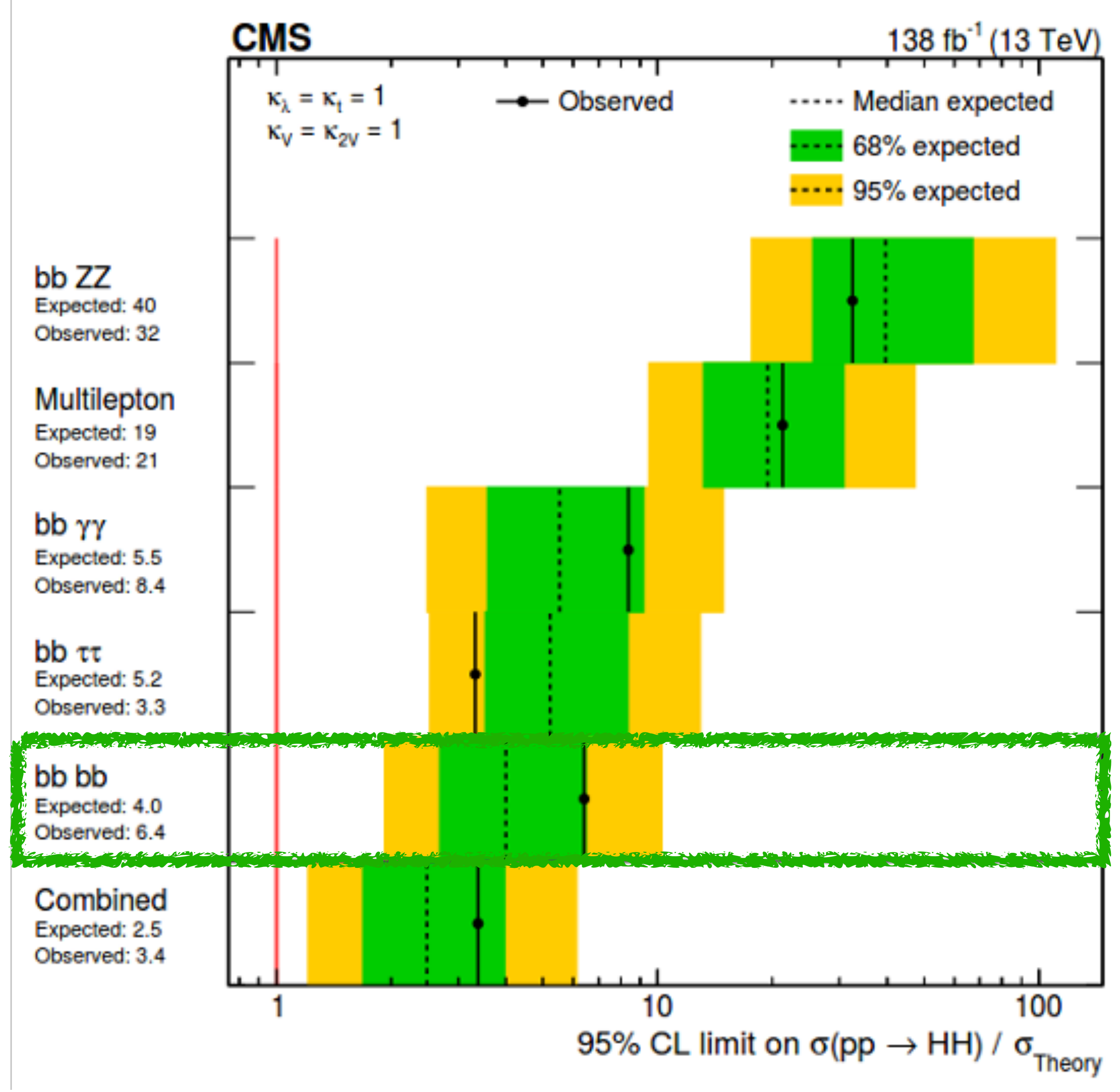
κ_{2V} coupling

[-0.1, 2.2] ([-0.4,2.5])

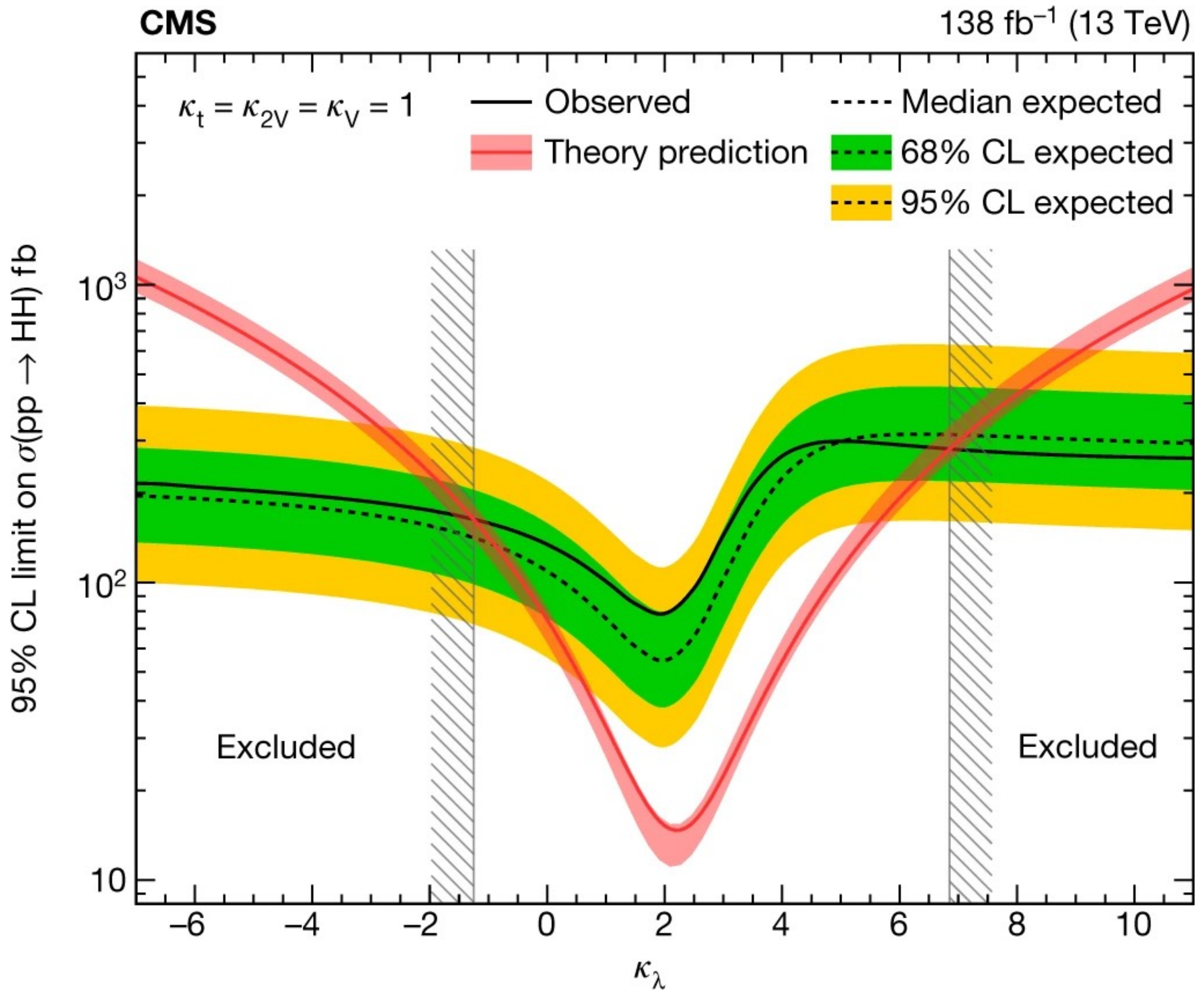
More in additional material

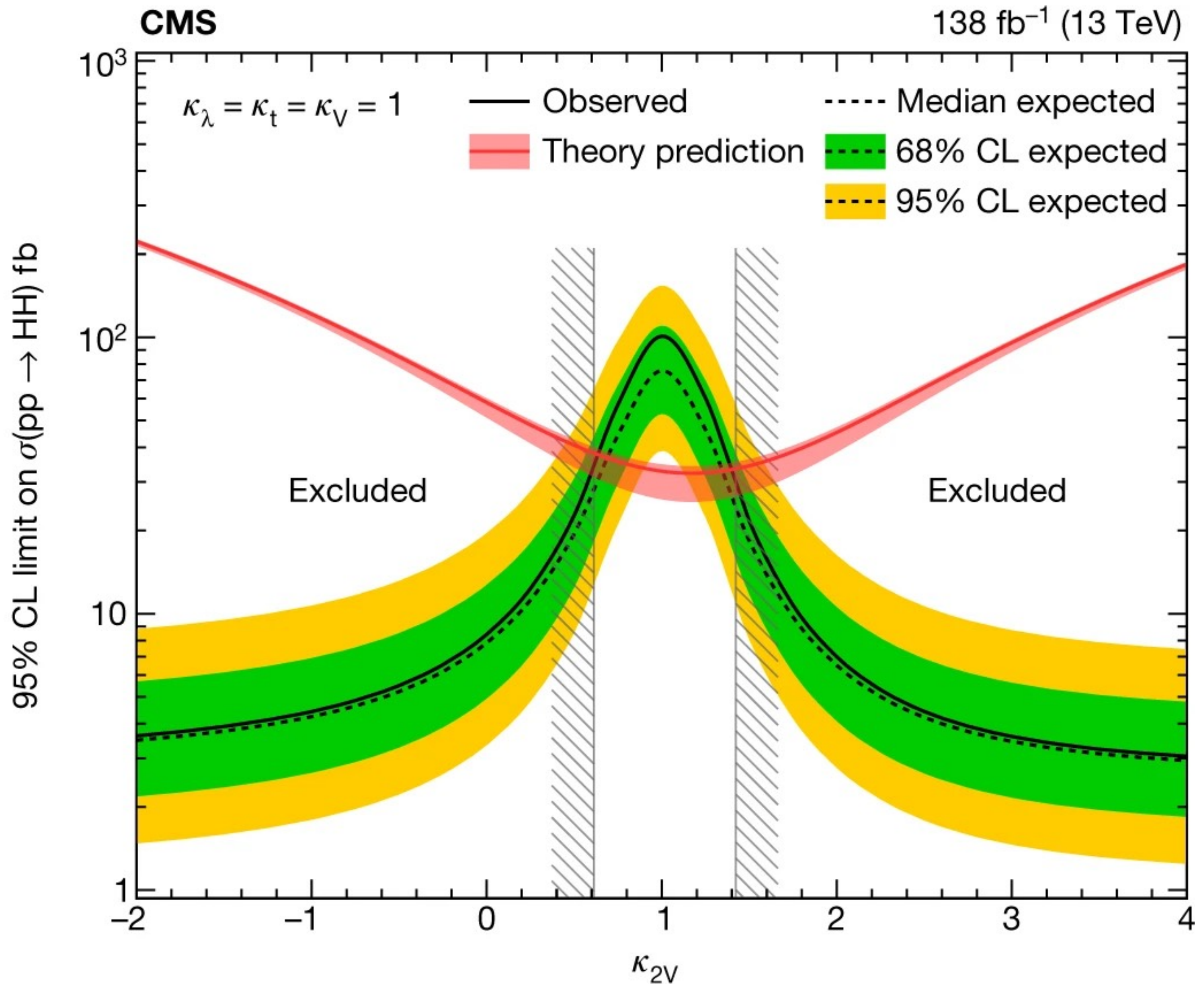
CMS HH Combination

Nature 607, 60–68(2022)



- Combination of published Run-2 analyses (2022 Higgs Anniversary):
  - Channels: bbbb, bb$\gamma\gamma$, bb$\tau\tau$, bbZZ and multilepton (4W,2W2τ,4τ)
  - bbbb (resolved+boosted) is the most sensitive channel
- SM constraint: Obs. (exp.) limit on σ/σ_{theory} is 3.4 (2.5)
- Constraints on anomalous Higgs couplings





κ_λ coupling

[-1.3, 6.9] ([-0.9,7.1])

κ_{2V} coupling

[0.6, 1.4] ([0.7,1.4])

More in additional material

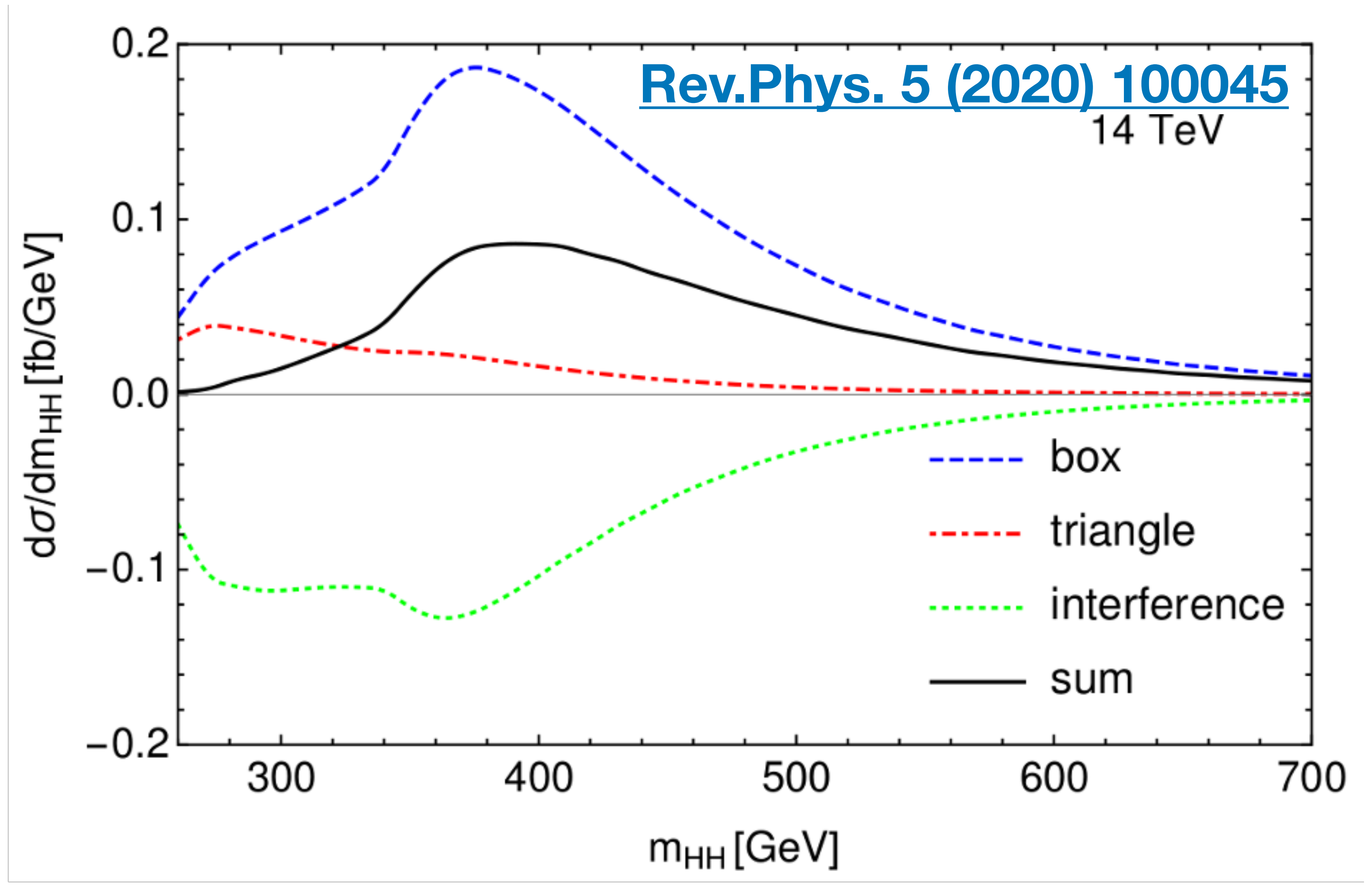
Summary and prospects

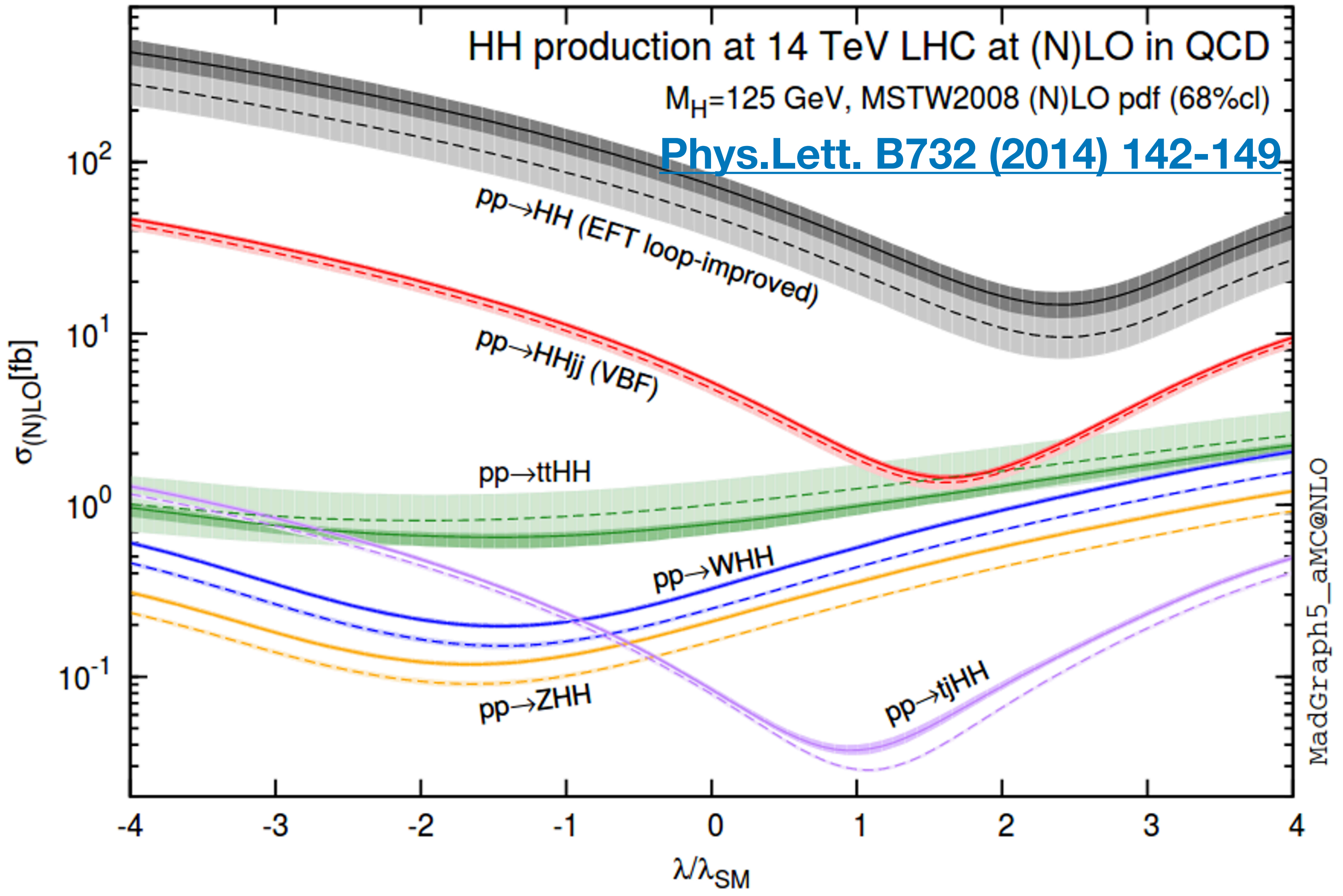
- The study of HH can shed light on the structure of the Higgs potential
- The CMS HH Run-2 program has been very successful
    - Many channels, modes, and combination (2022)
    - The bbbb is the most sensitive channel, leveraging in novel techniques
    - Tight constraints have been placed on SM and anomalous couplings
- The LHC Run-3 will bring improvements for the exploration of this process
    - New dedicated HH triggers, ID algorithms, and analysis techniques are foreseen

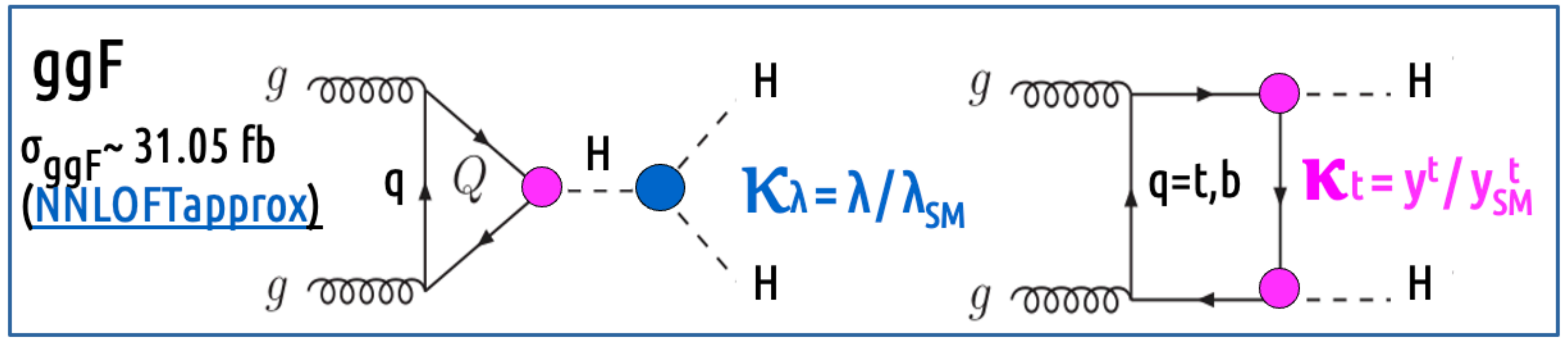
Very exciting times are ahead,
we are looking forward to carrying out analyses with Run-3 data,
and the future HL-LHC!

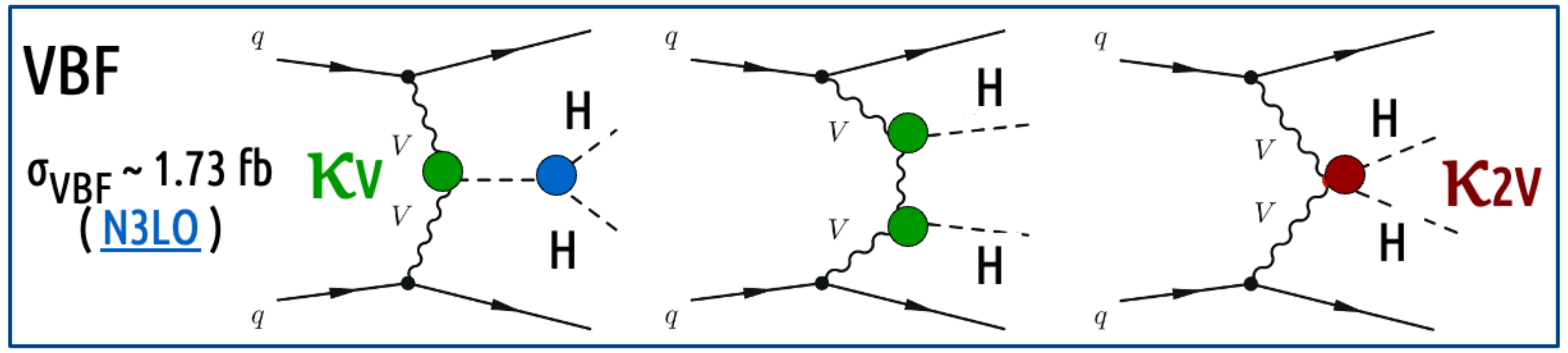
Additional Material

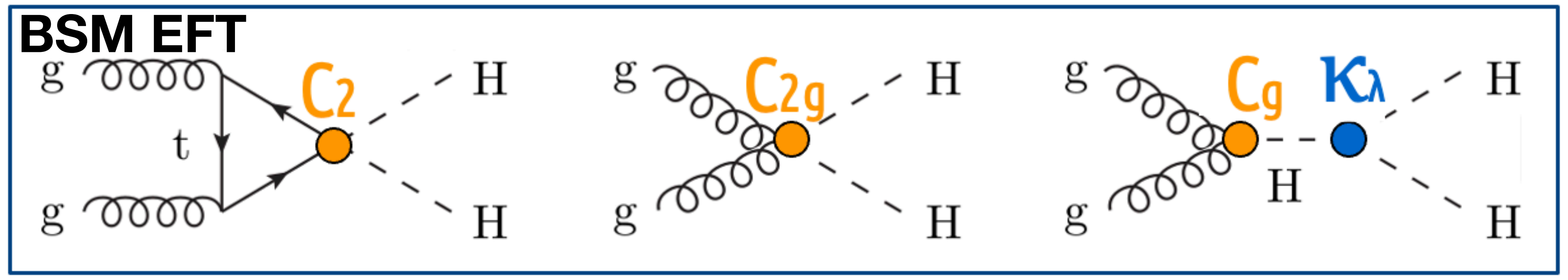
HH production









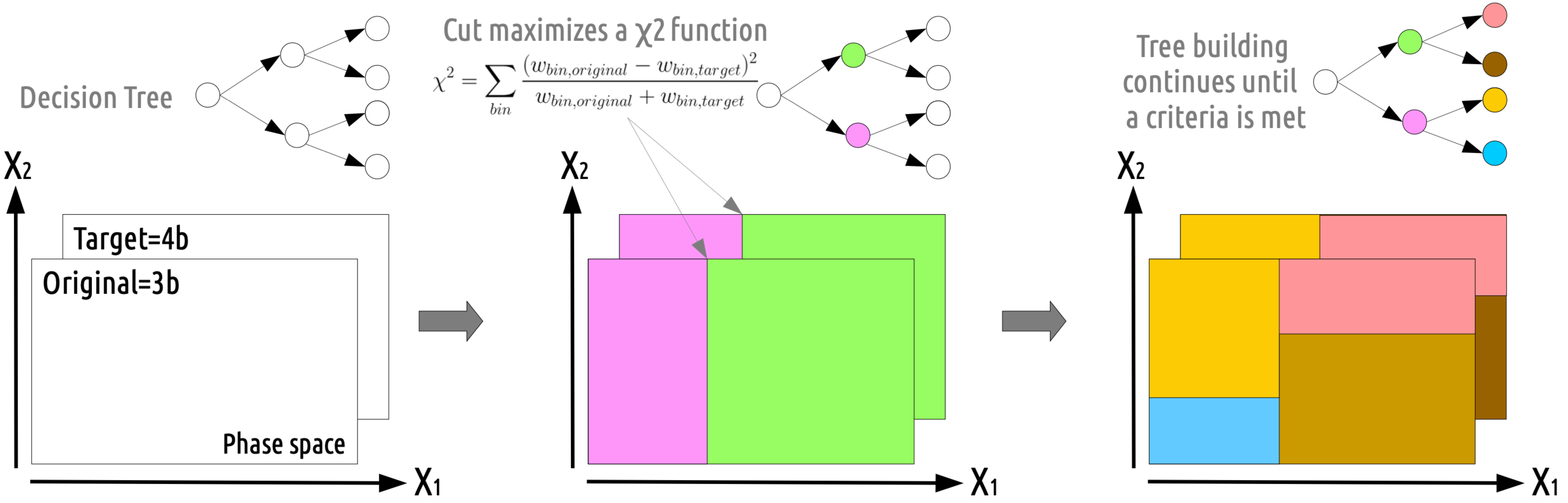


HH→bbbb (resolved): Extended

If we want to model '4b' bkg by re-weighting '3b' data, thus we find the $\rho(X) = f(X)_{4b}/f(X)_{3b}$

A `BDT-Reweighter' splits cleverly the multi-dimensional space X to find regions that need reweighing

BDT-Reweighter Training: Iteratively builds a tree and reweighs the '3b' data with tree weights (a.k.a. boosting)



The BDT-Reweighter score = a event-by-event reshaping weight, i.e. our $\rho(X)$ estimator

The algorithm is implement in hep_ml library. More information in the associated paper: arXiv:1608.05806.

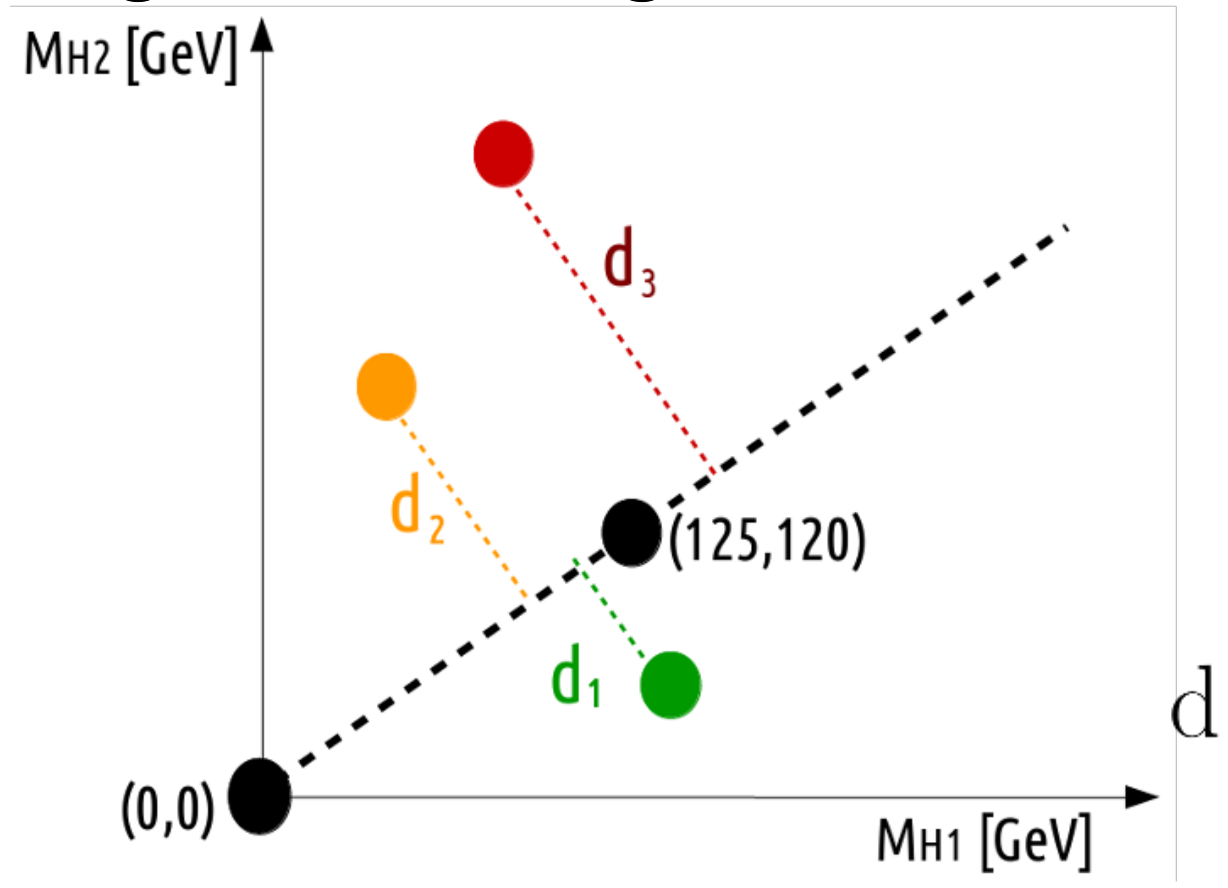
HH→bbbb (resolved): Extended

PRL 129, 081802

Pairing four b jets into two Higgs candidates → 3 pairing possibilities
The chosen method should avoid sculpting the background near the Higgs mass

H1, H2 identified by pairing algorithm:

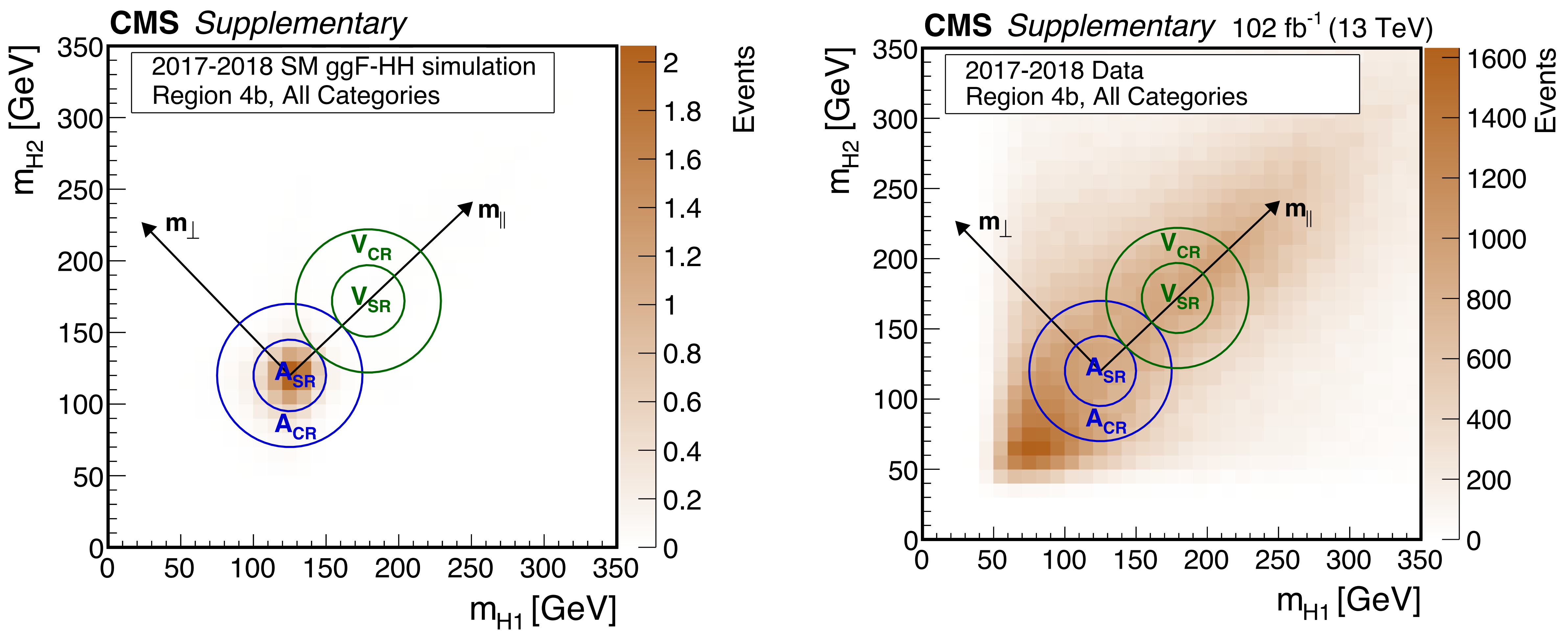
If |d1-d2| ≥ 30 GeV:
  d1, i.e. closest to diagonal
else:
  d1 or d2, largest PT(H) in 4-jet CM frame



PT (H1) > PT (H2)
d1 < d2 < d3
k=125/120=1.04

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

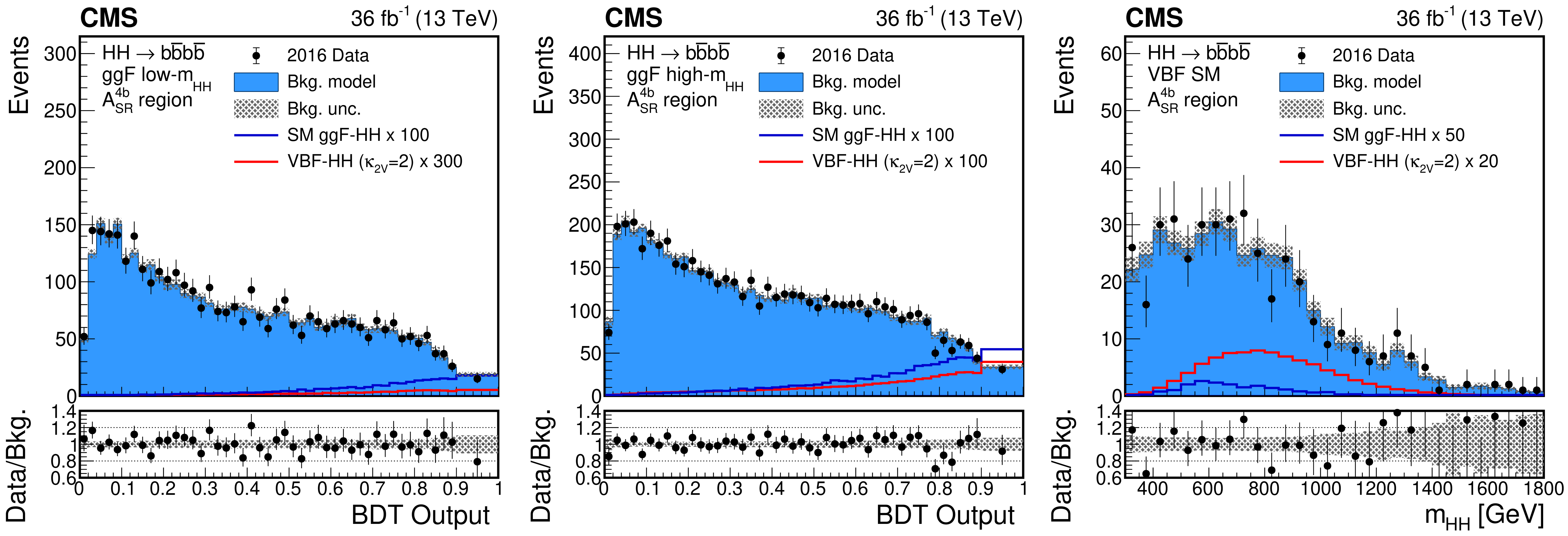
Correct pairing is 82-96% (91-98%) depending on ggF (VBF) hypotheses w/o bkg. sculpting near the Higgs mass



HH→bbbb (resolved): Extended

PRL 129, 081802

2016 discriminant observables



VBF Counting Experiment

2016 (2017– 2018) data set, 4 (13) events are observed for a total of 4.0+/-1.3 (15.0+/-3.4) background

HH→bbbb (resolved): Extended

PRL 129, 081802

Signal experimental uncertainties

- Luminosity in 2016 (1.2%), 2017 (2.3%) and 2018 (2.5%)
- Pile-up, L1 Pre-firing (2016, 2017)
- b-tagging and trigger efficiency
- Jet energy scale, jet energy resolution

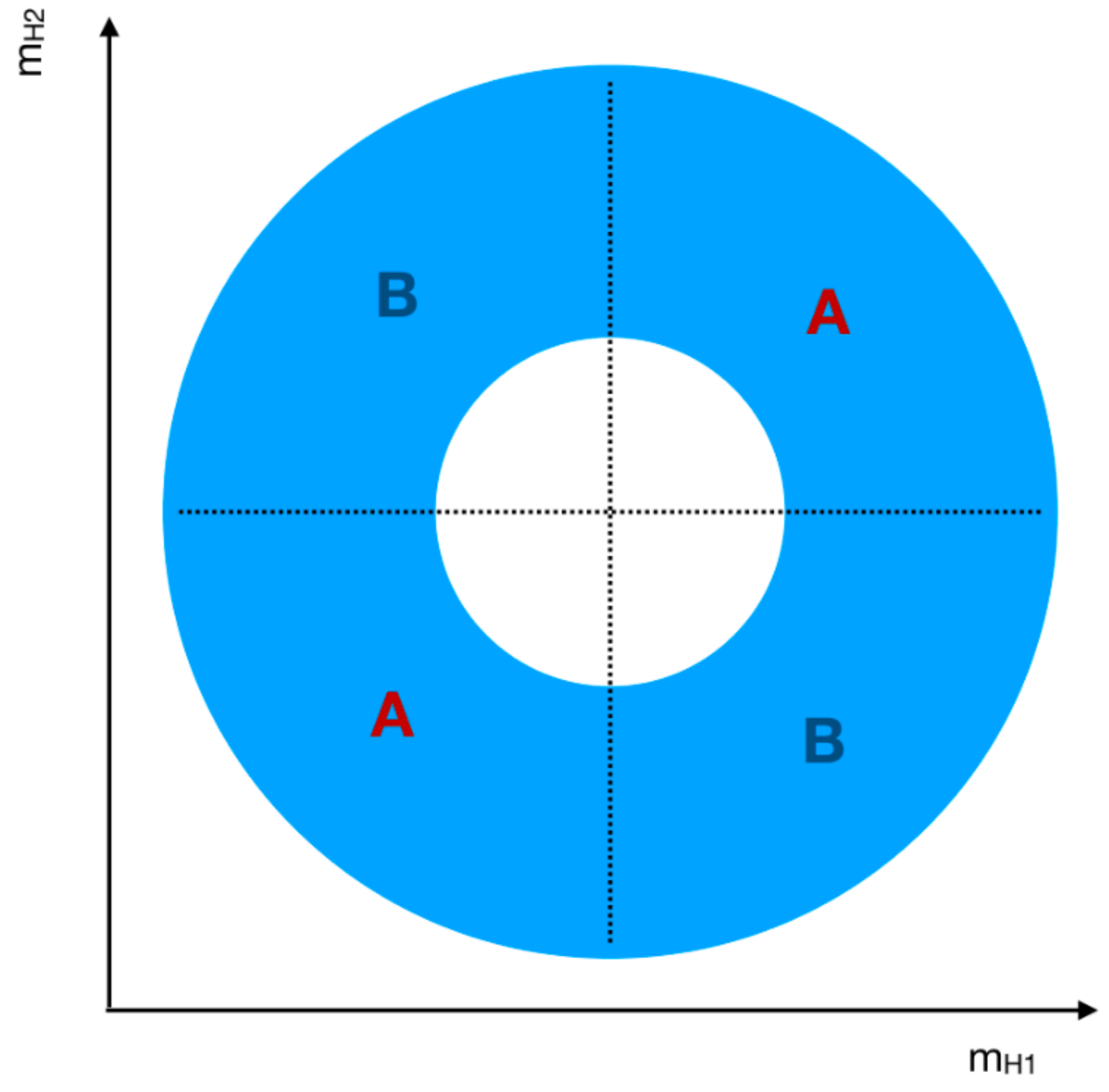
Signal generation and theory uncertainties

- Factorization scales, Parton-Shower (PS), and PDF
- Event migration due to PS ISR recoil scheme (only for VBF signals)
- Cross section and final state branching fraction

Background model uncertainties

- Bin-by-bin uncertainty to account for Poisson fluctuations of the SR(3b) data
- 3b→4b Transfer factor uncertainty from limited CR statistics
- Shape uncertainty:
  - ggF cat. 1,2: Alternative shape derived training BDTReweighter using alternative CR definition
  - VBF cat. 1: Linear fit to M(HH) data/bkg ratio in validation region
- Uncertainty due to V-SR(4b) statistical power with respect to A-SR(4b)
- Uncertainty on normalization closure in V-SR(4b): 1.5 - 4.7% depending on category/year

Alternative CR definition

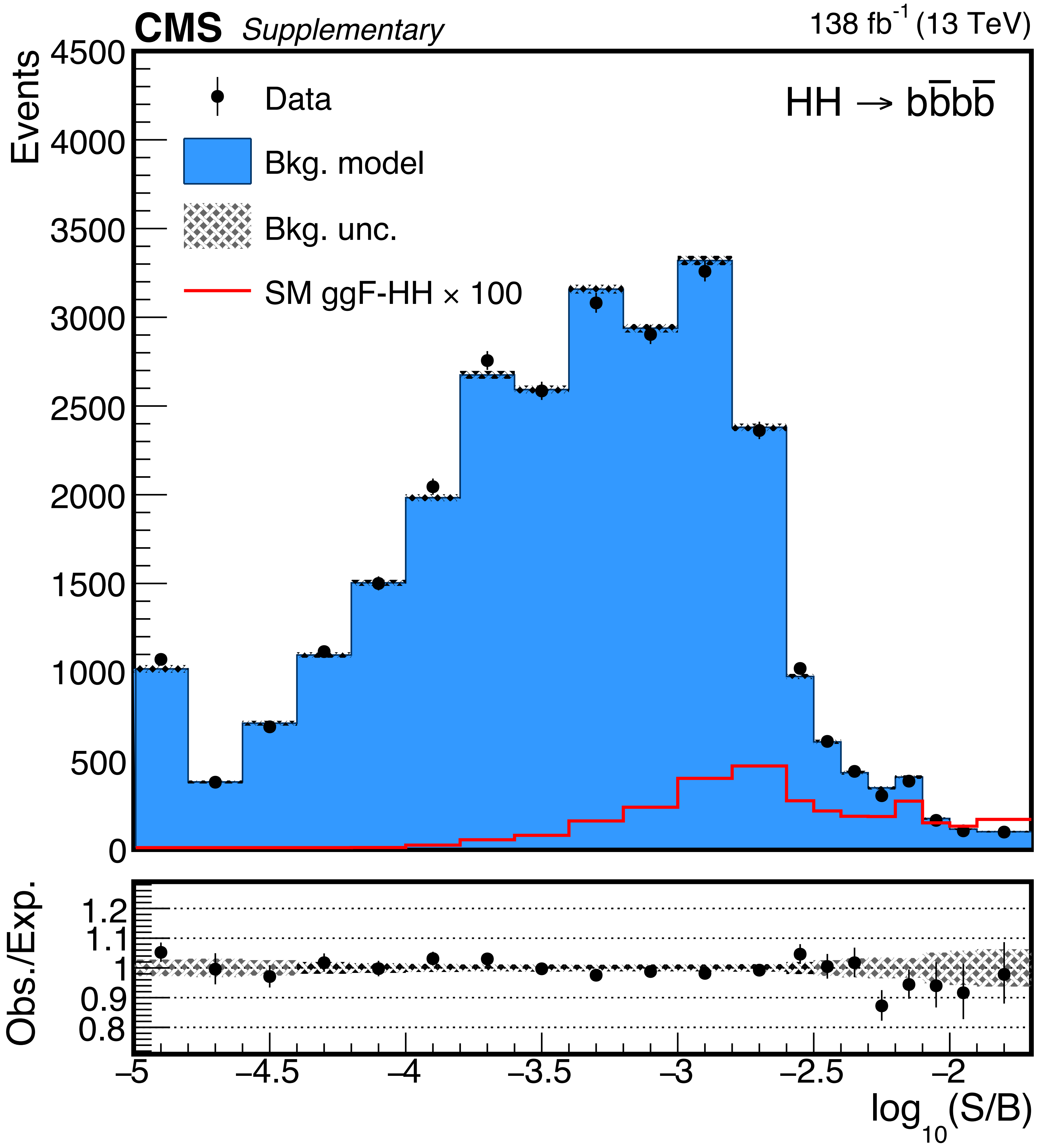


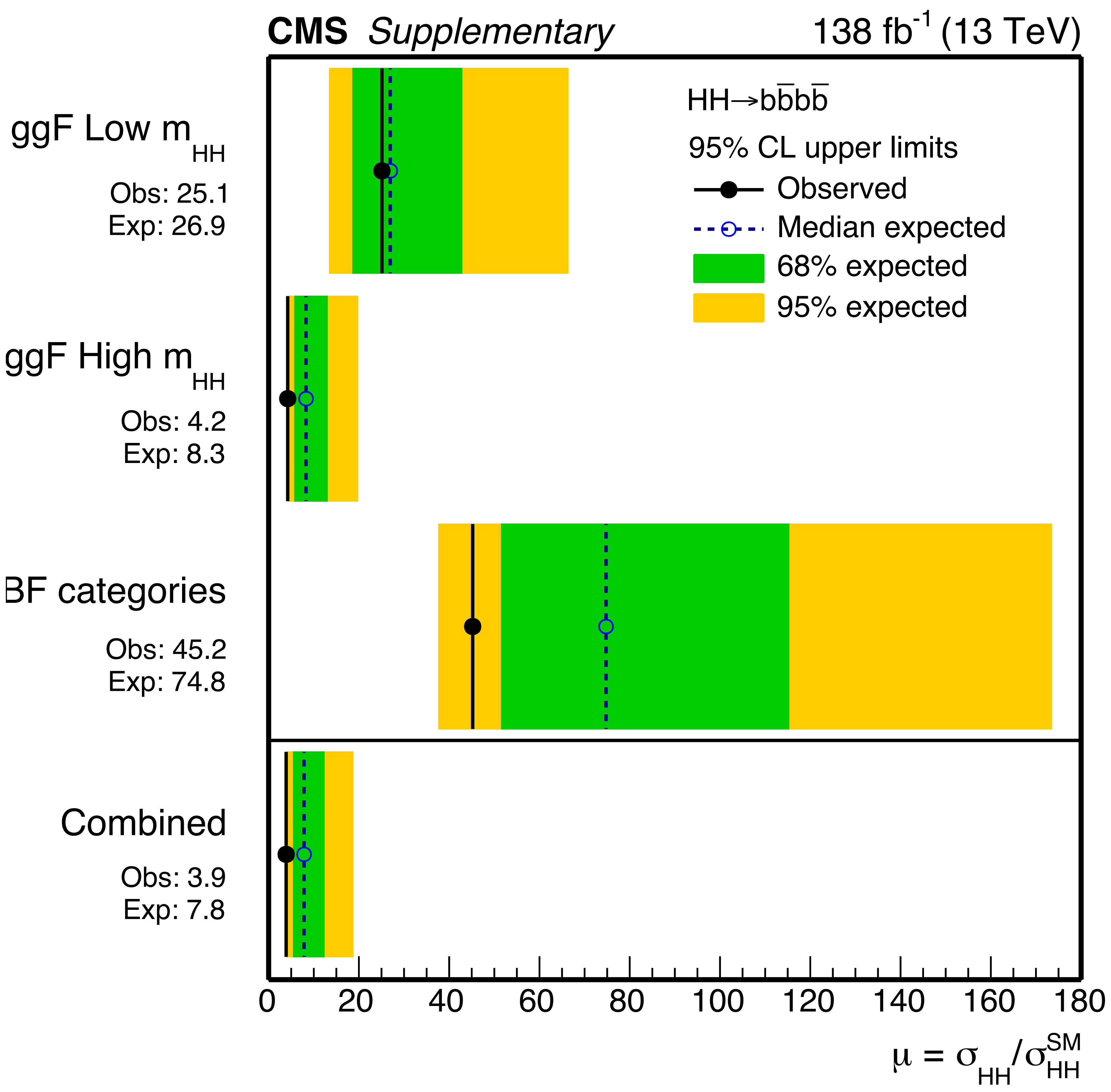
HH→bbbb (resolved): Extended

PRL 129, 081802

No excess of data is observed relative to the background expectation

95% CL upper limits are set on the SM and BSM production cross sections

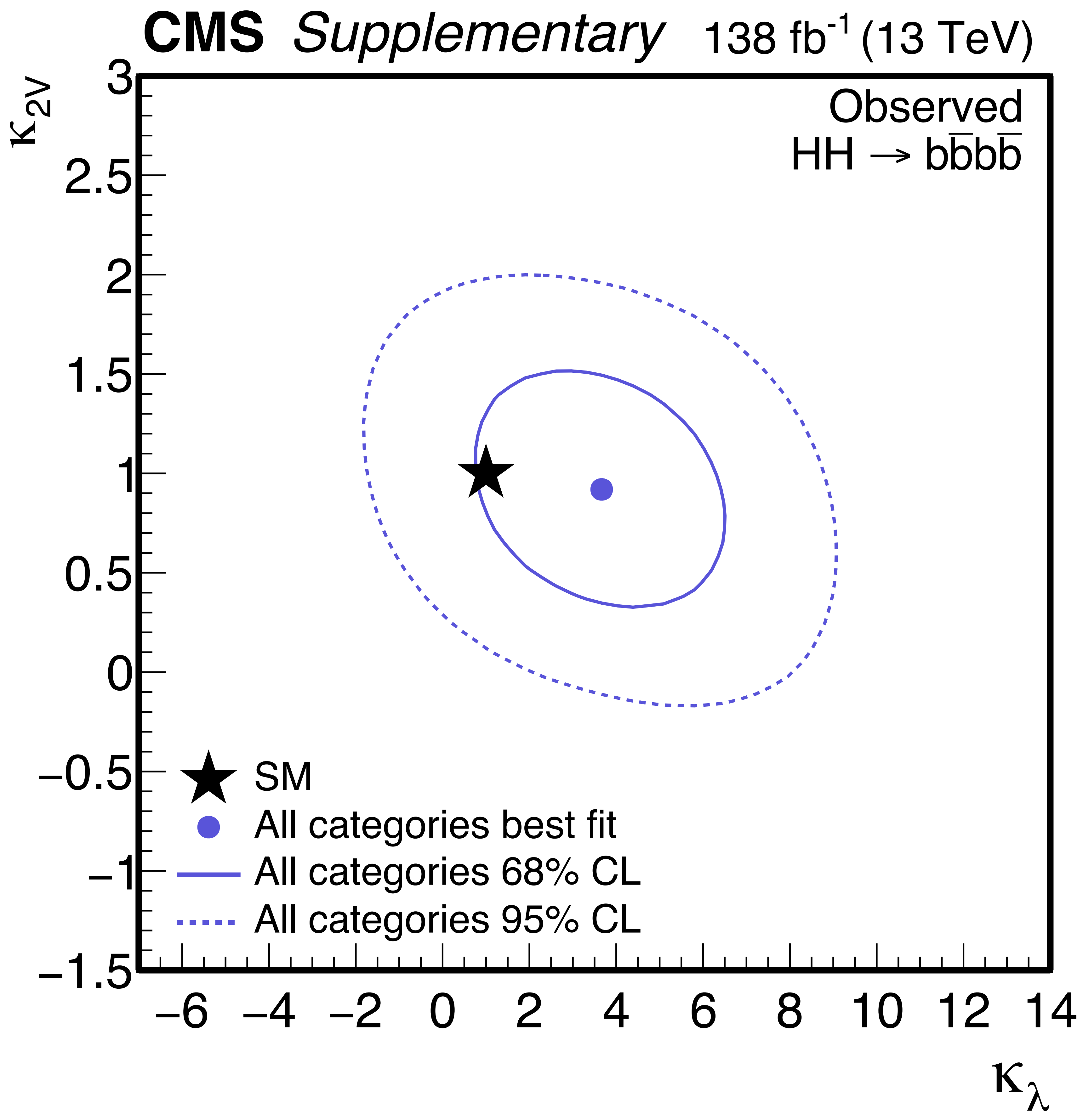


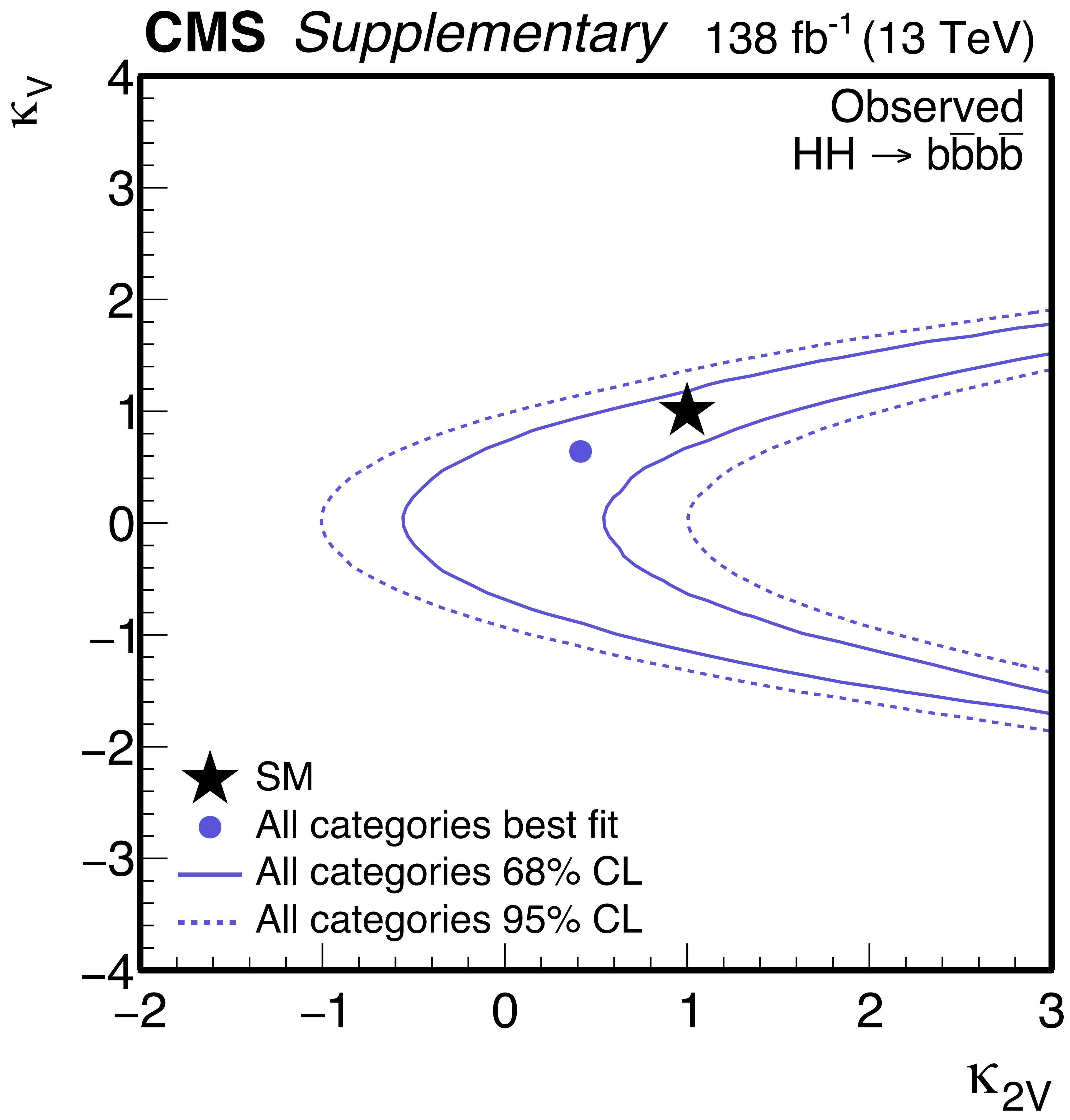


CMS HH→bbbb (resolved): Extended

PRL 129, 081802

Measurement of Higgs couplings via the negative log-likelihood scan
Simultaneous fit of the ggF and VBF signal contributions as a function of the couplings



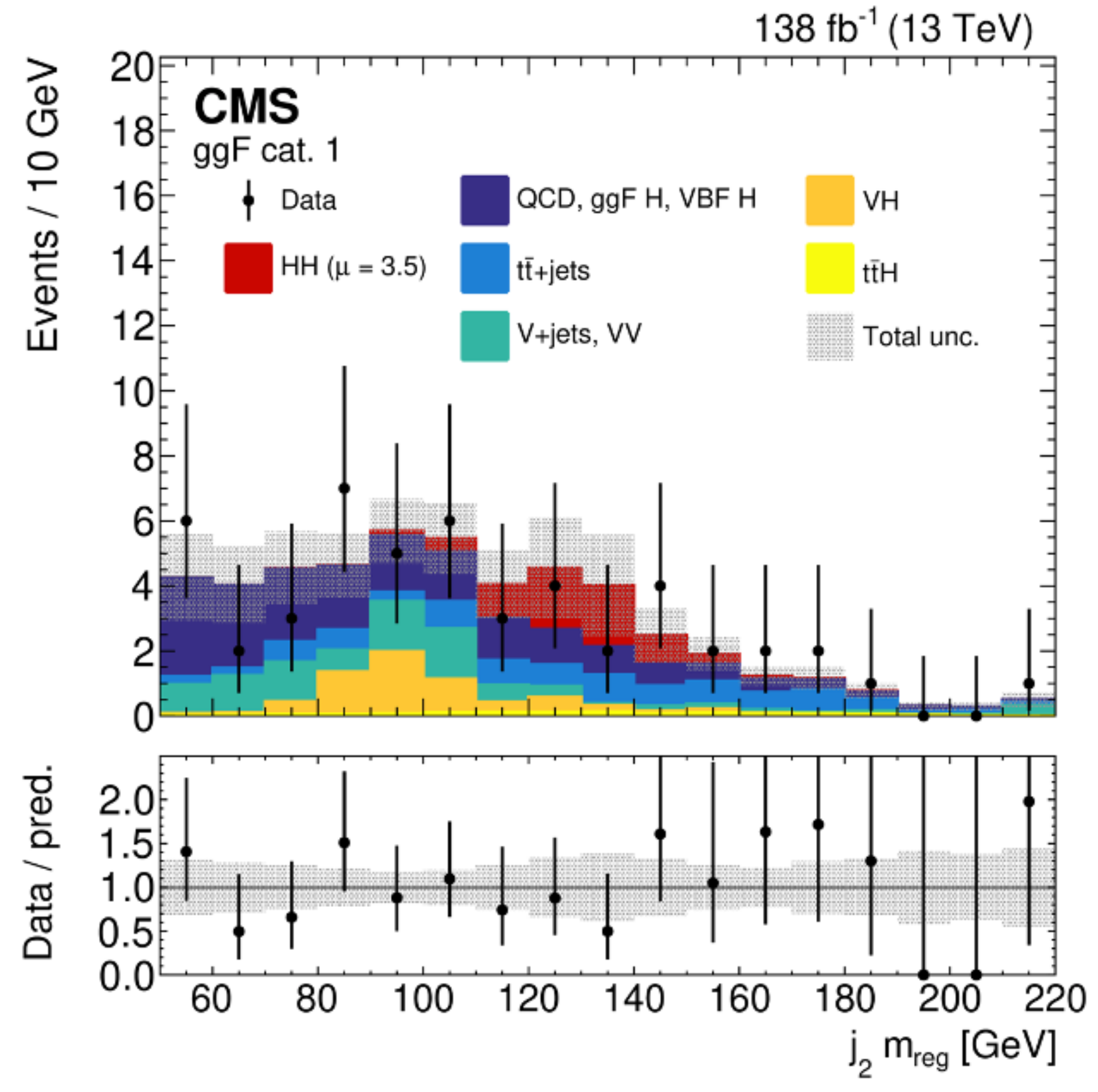


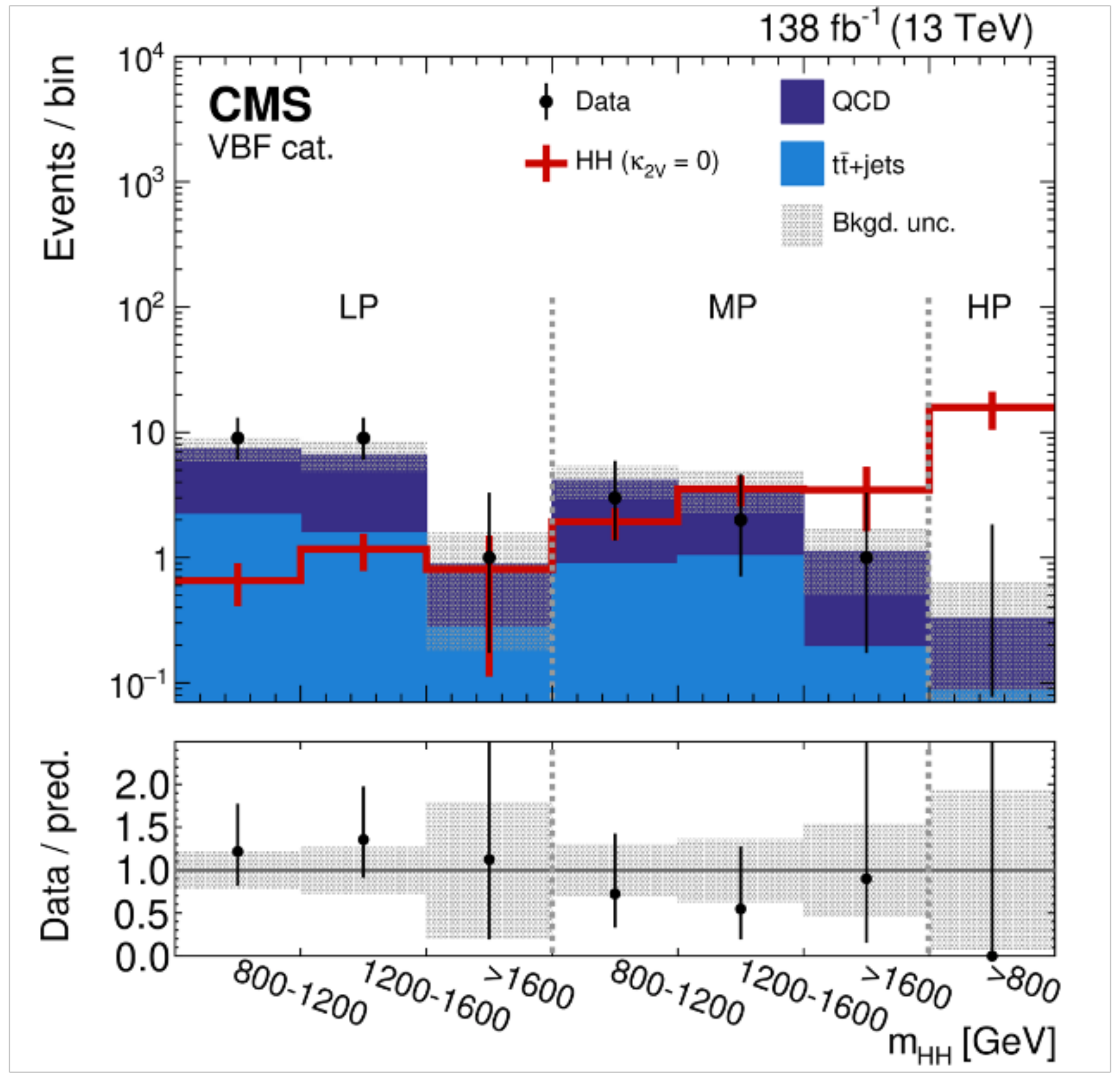
Observed results are compatible with SM at 95% CL

CMS HH→bbbb (boosted): Highlights

arXiv:2205.06667

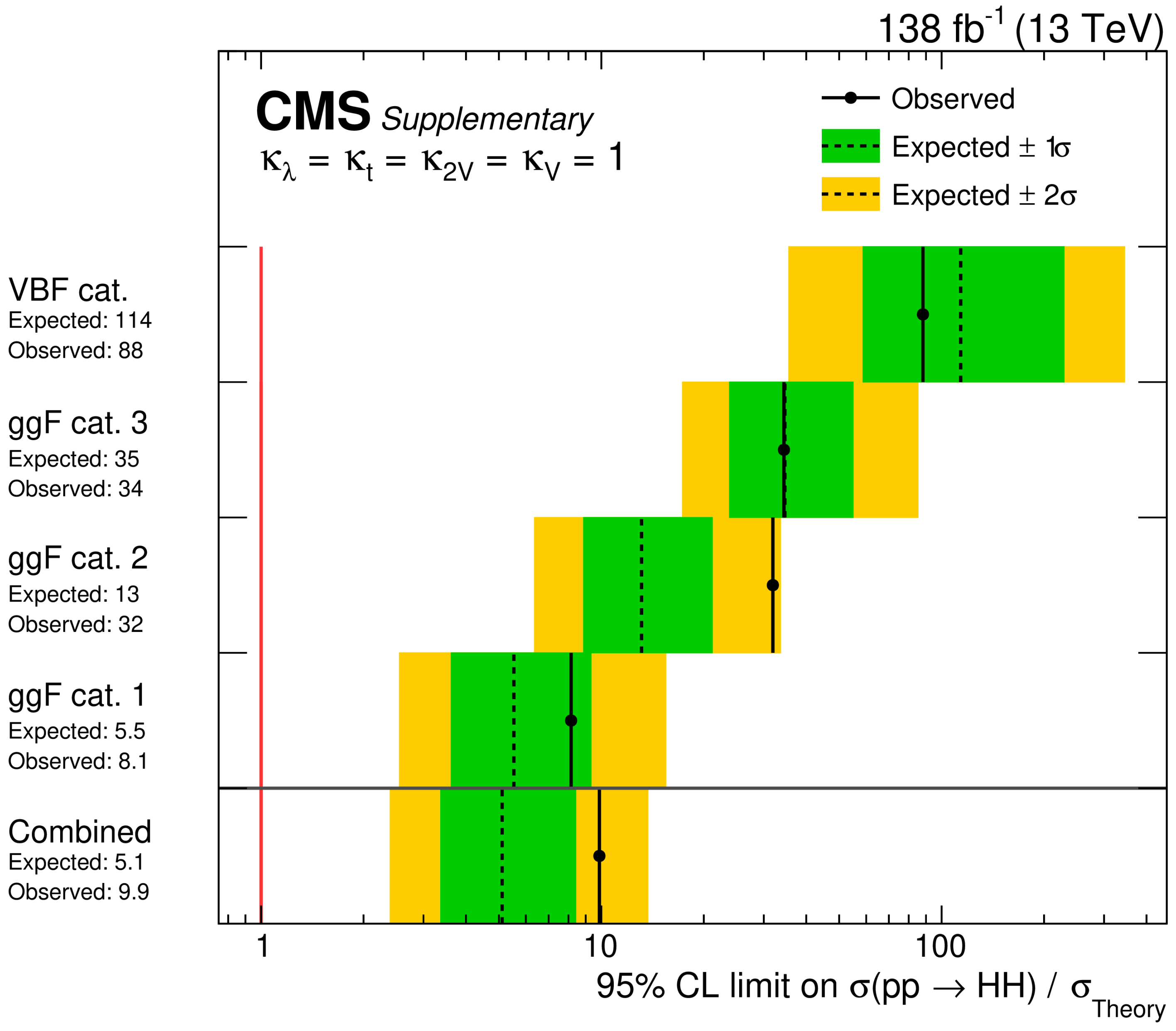
- It targets both ggF and VBF modes
- AK8 jet triggers
- 2 AK8 jets with pT > 300 GeV:
  - Two highest ParticleNet scores (D_{bb})
  - Mass regression is applied
- Event categorization using BDT
- Background model:
  - QCD in ggF categories: Parametric alphabet method
  - QCD in VBF categories: ABCD method
  - tt+others from MC simulation
- Signal extraction observables:
  - ggF categories: j2 regressed mass (m_{reg})
  - VBF categories: m_{HH} distribution (3 or 1 bin)



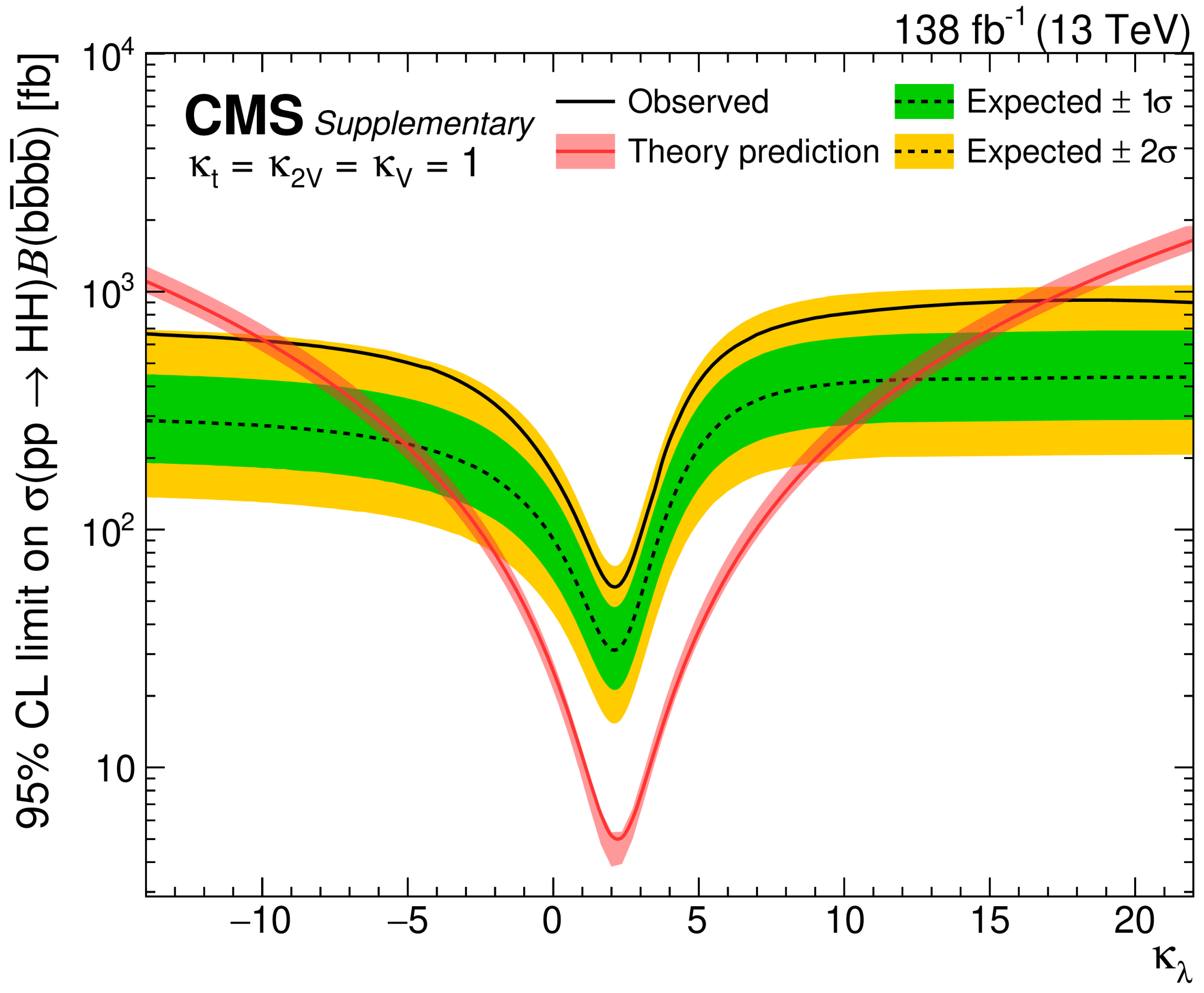


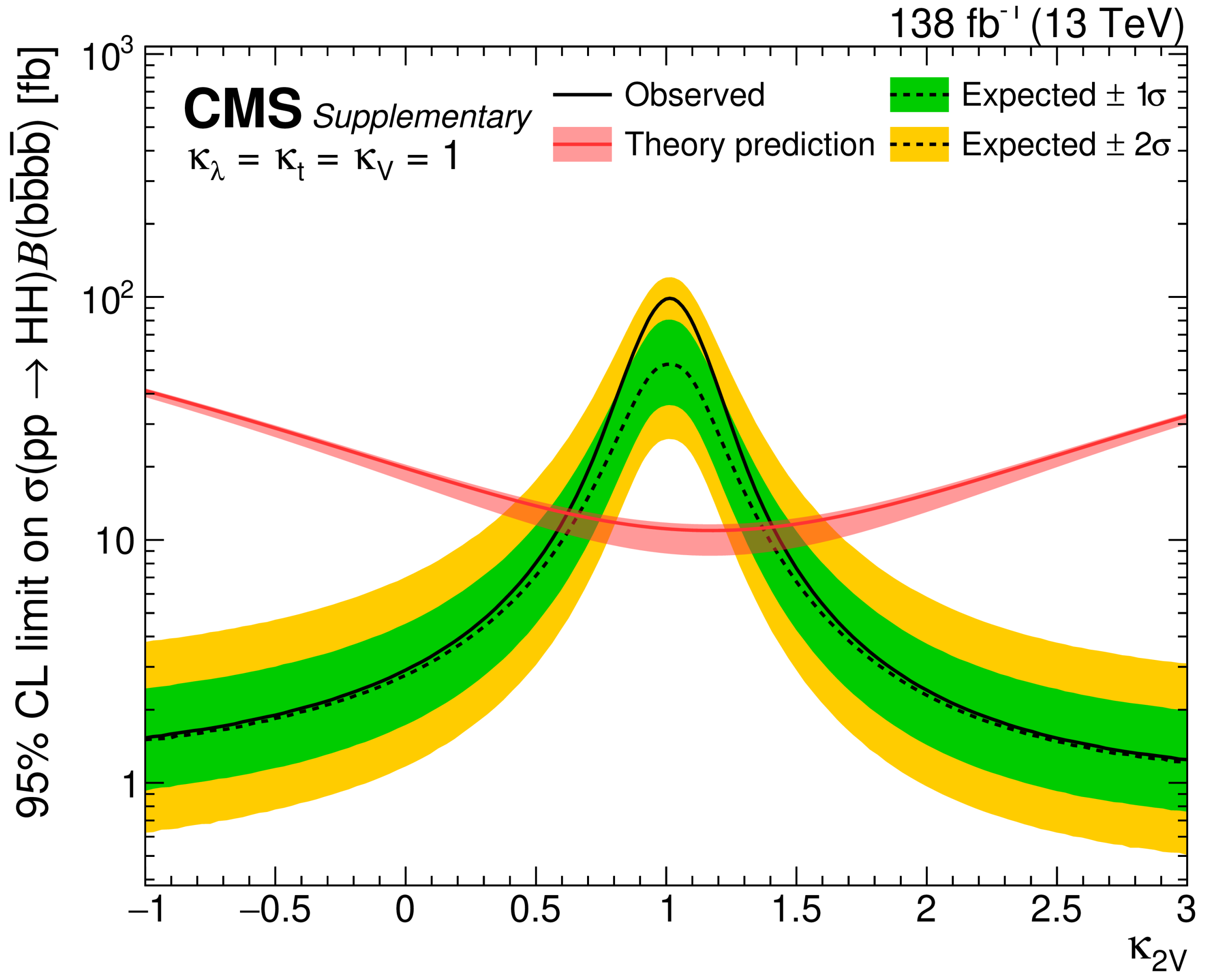
CMS HH→bbbb (boosted): Results

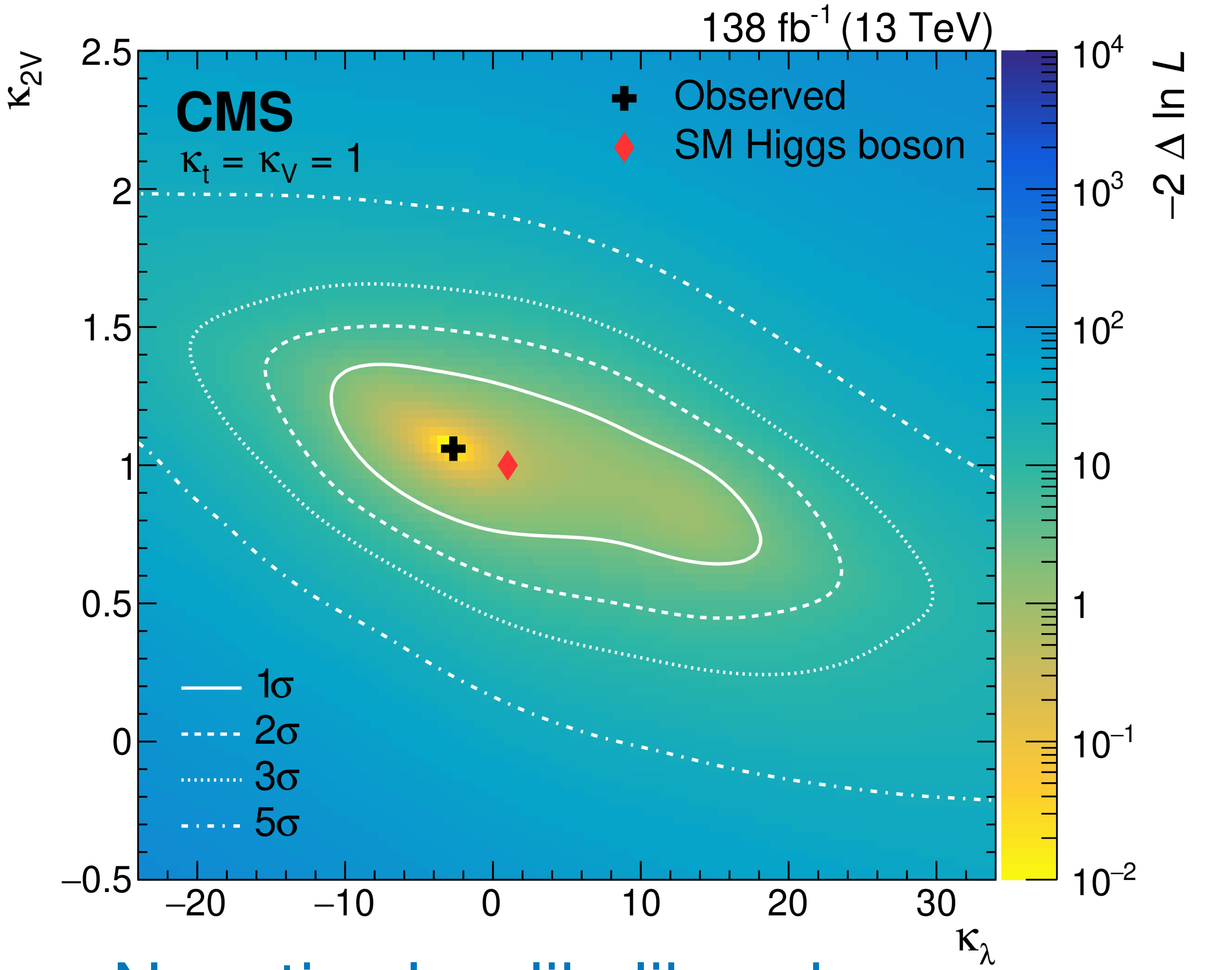
arXiv:2205.06667



SM constraint







Negative log-likelihood scan as a function of the couplings

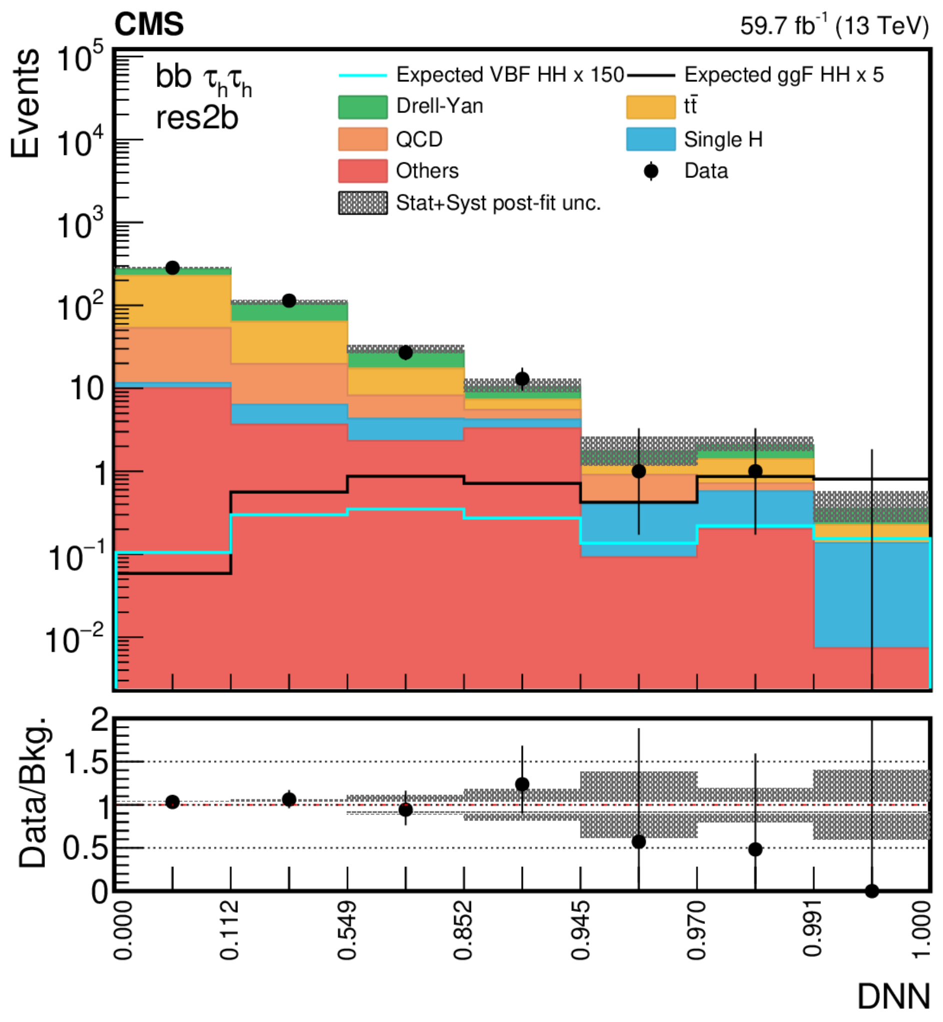
Anomalous couplings constraints

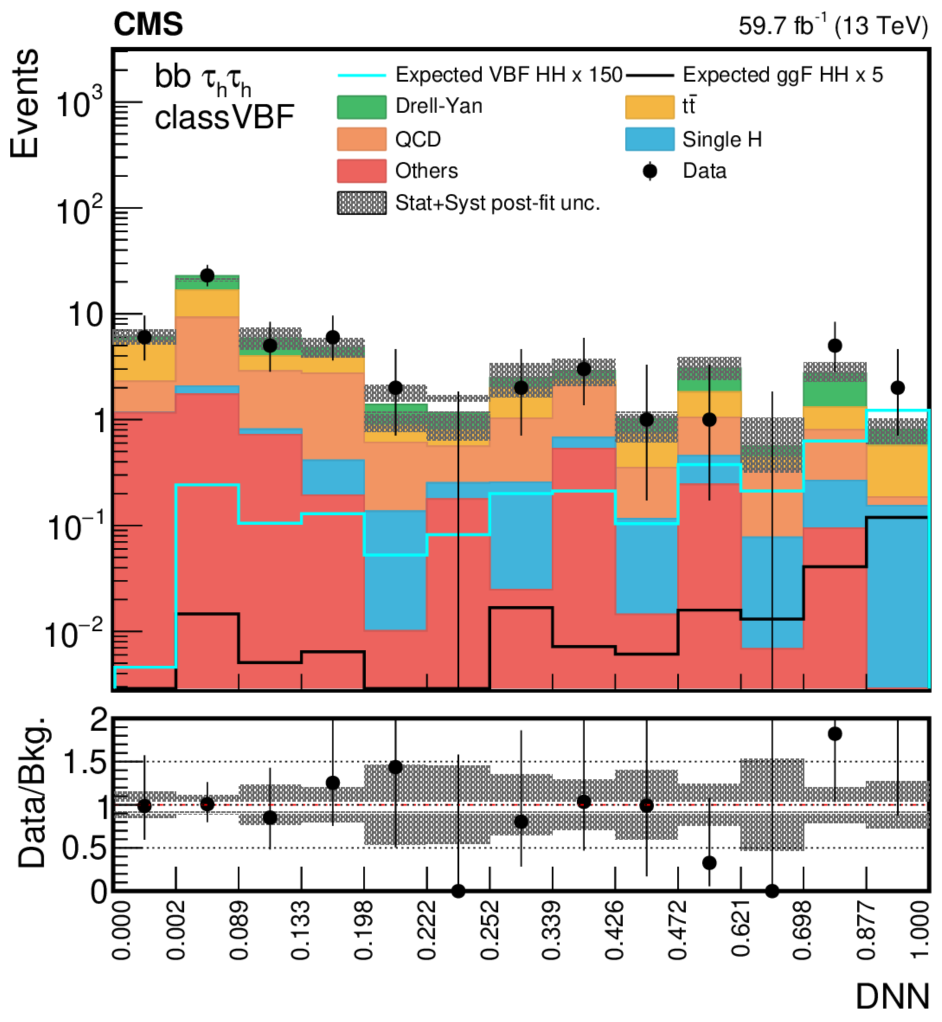
$\kappa_{2V} = 0$ hypothesis is excluded with 6.3 s.d. significance

CMS HH→bb$\tau\tau$: Highlights

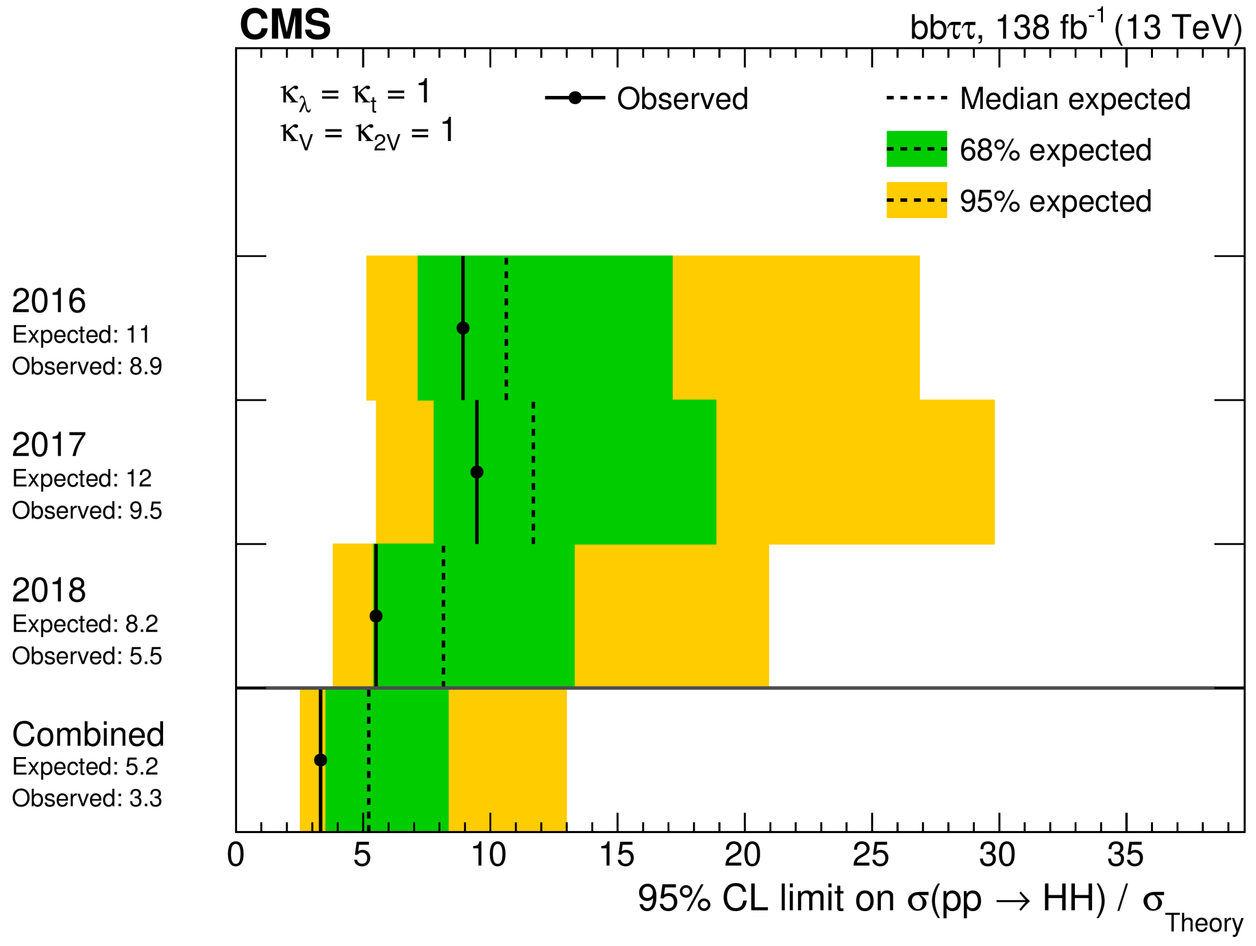
PLB 842 (2023) 137531

- It targets both ggF and VBF modes
- Triggers: 1 lepton ($l = e, \mu$), $l + \tau_h$, and $2\tau_h$
- 3 channels: $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$
- Background model
  - QCD is data-driven method
  - $t\bar{t}$ + DrellYan, others from MC simulation
- Signal extraction observables: DNN output





SM constraints



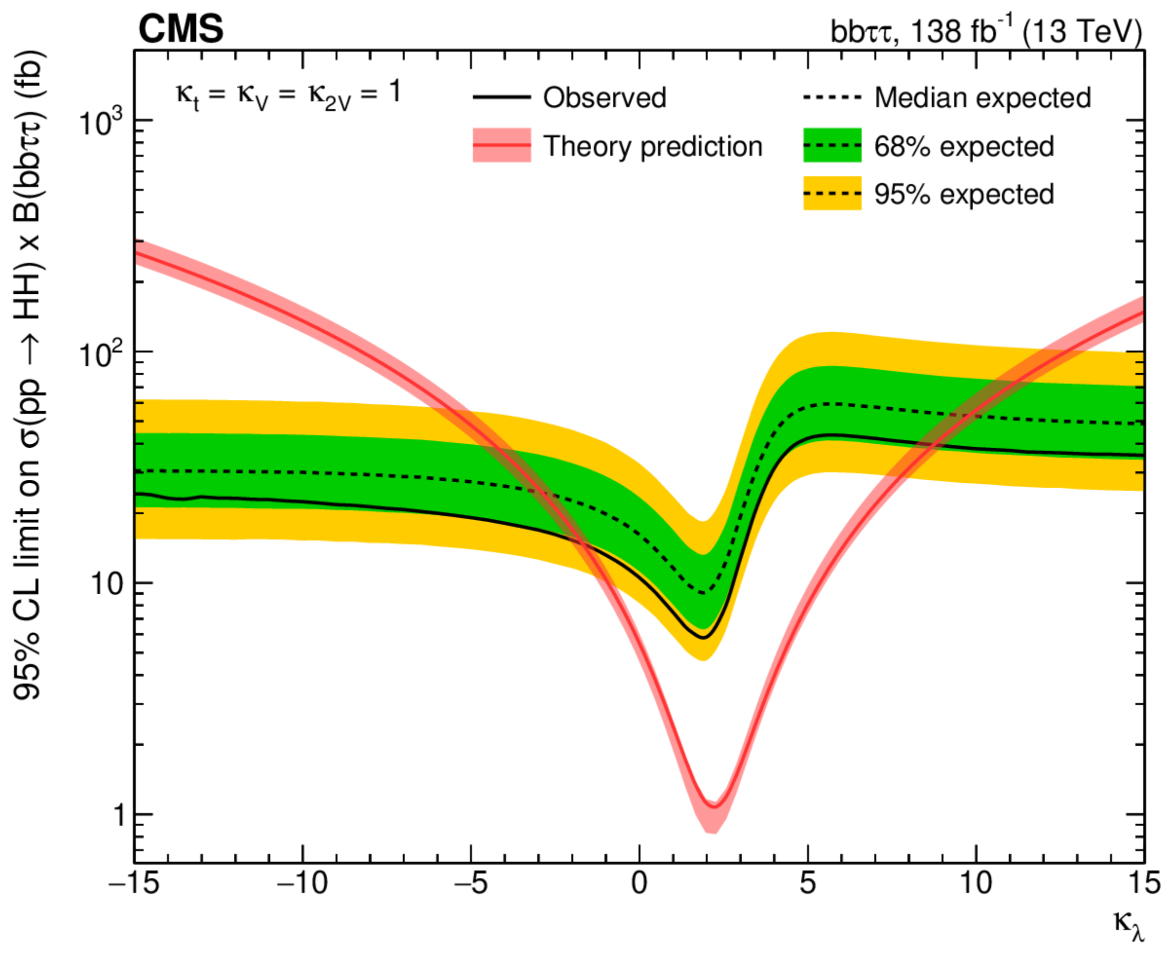
Anomalous coupling constraints

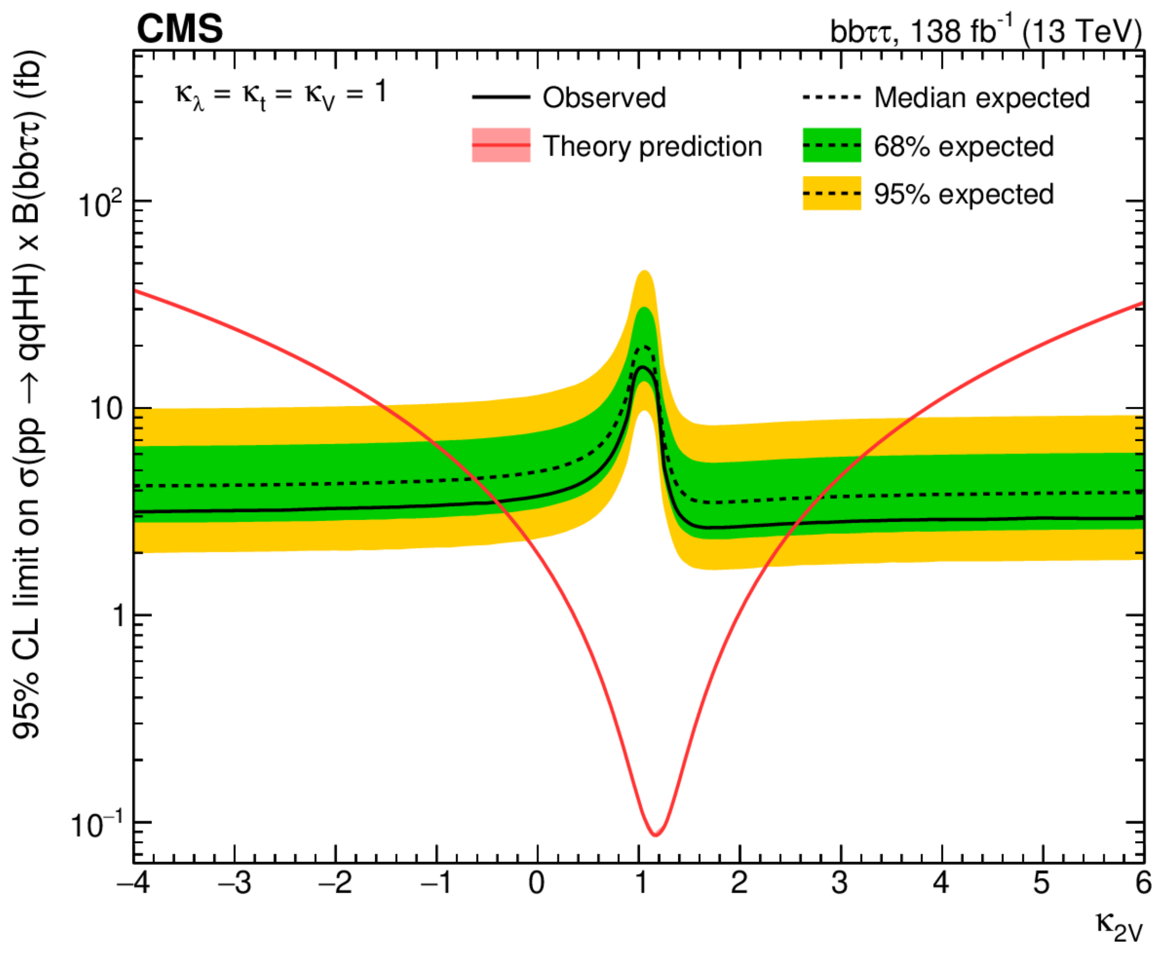
Obs. (exp.) κ_λ-constraint

-1.7 (-2.9) < κ_λ < 8.7 (9.8)

Obs. (exp.) κ_{2V}-constraint

-0.4 (-0.6) < κ_{2V} < 2.6 (2.8)

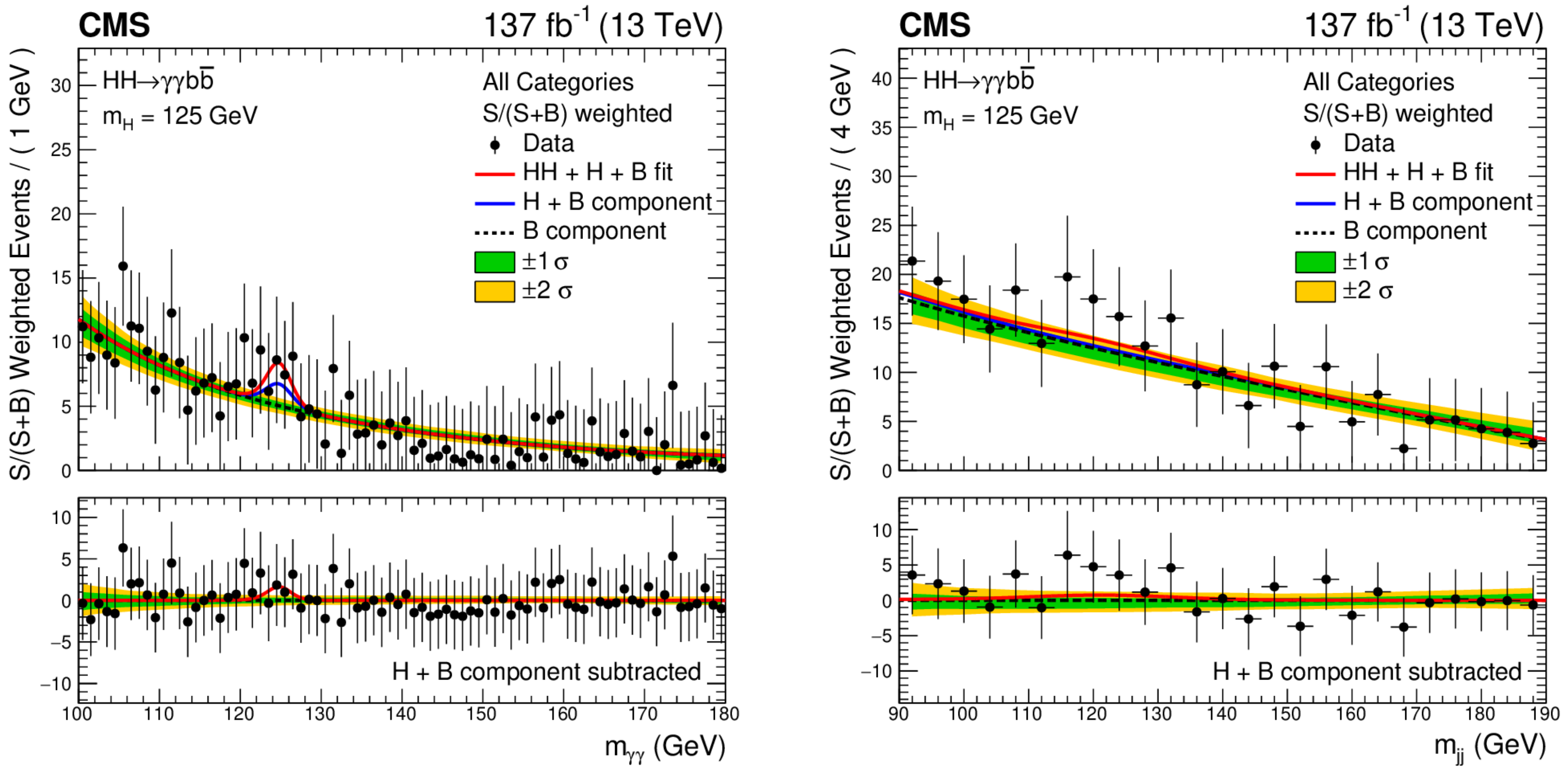




CMS HH→bb$\gamma\gamma$: Highlights

JHEP03(2021)257

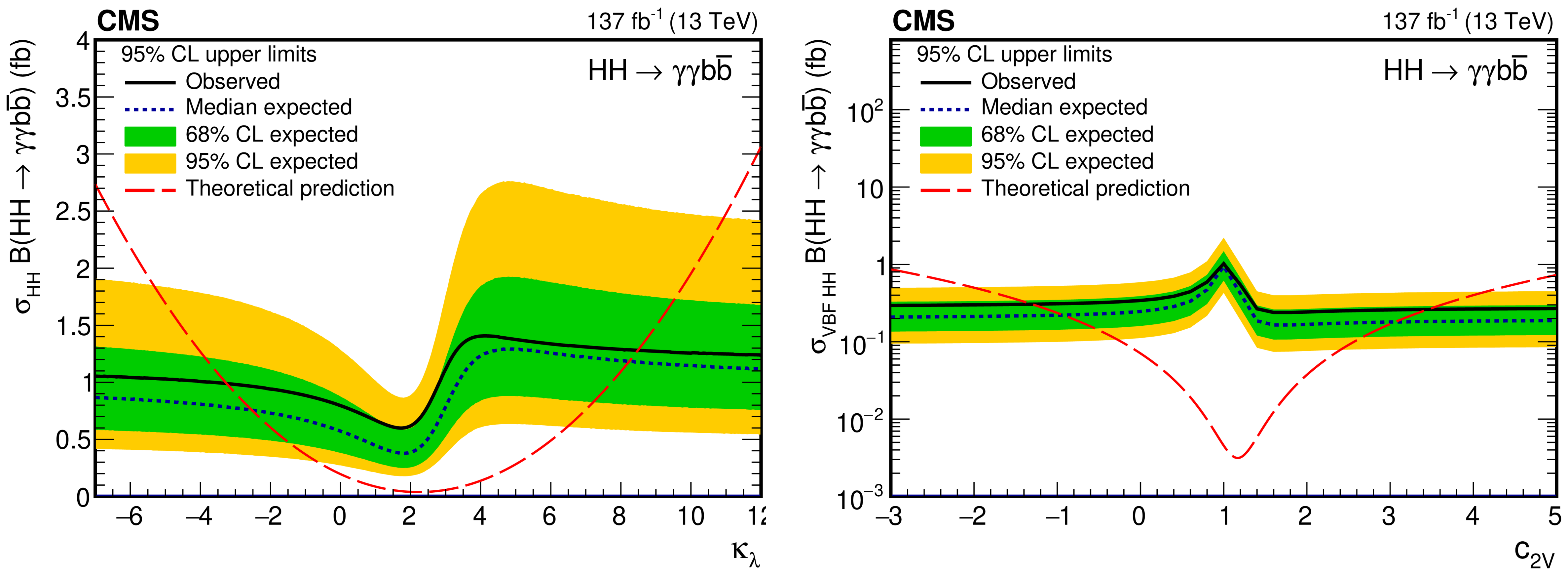
- It targets both ggF and VBF modes
- Photon triggers
- 2 γ's and 2 highest b-tagged jets:
  - DNN based b jet energy regression
  - di-jet mass regression
- Event categorization with MVA and m_{HH}
- Background model from parametric fit
- Signal extraction: 2D Fit to $m_{\gamma\gamma}$ and m_{bb} distribution



SM constraints

Obs. (exp.) limit on σ/σ_{theory} is 7.7 (5.2)

Anomalous coupling constraints



Obs. (exp.) κ_λ-constraint

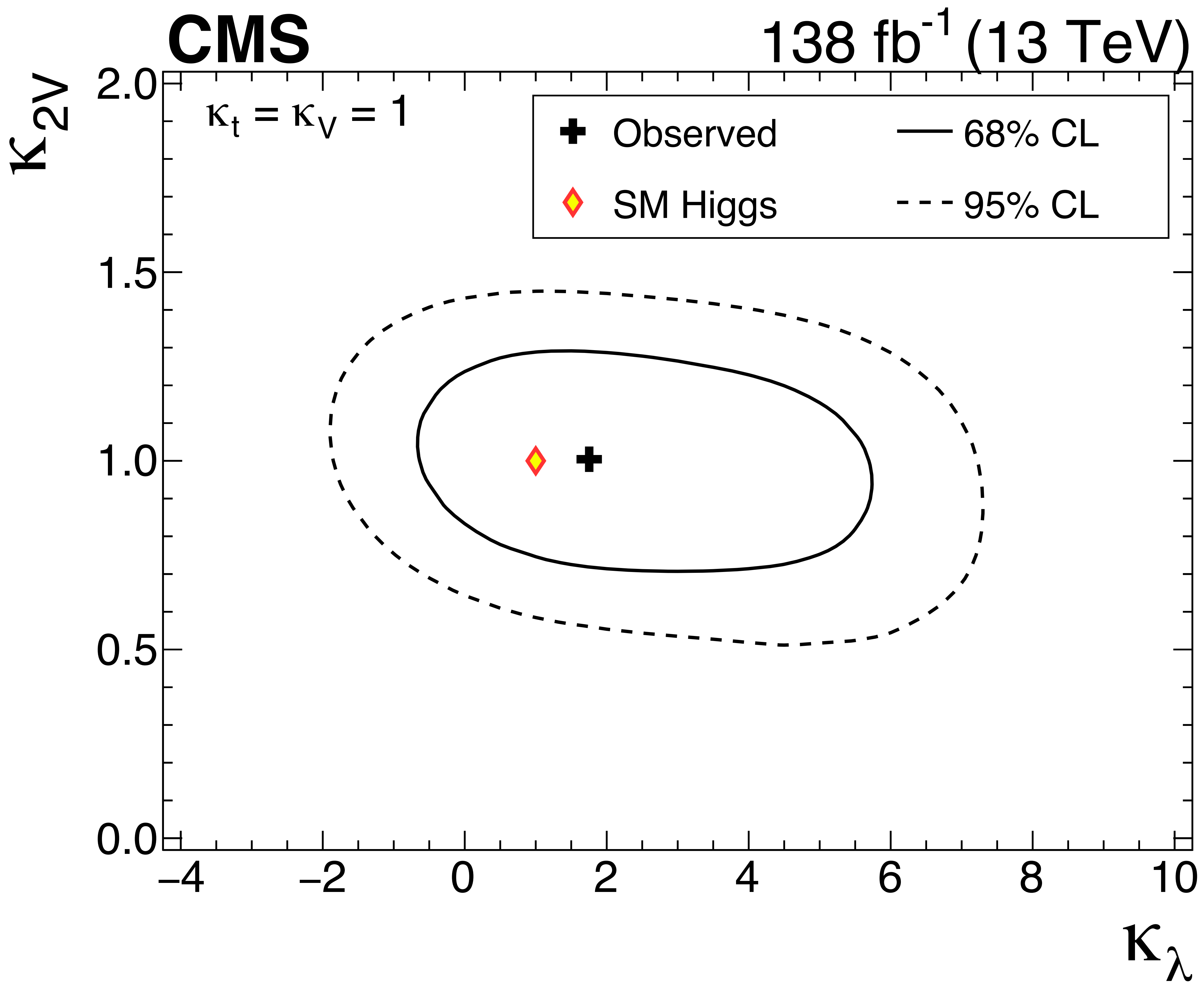
-3.3 (-2.5) < κ_λ < 8.5 (8.2)

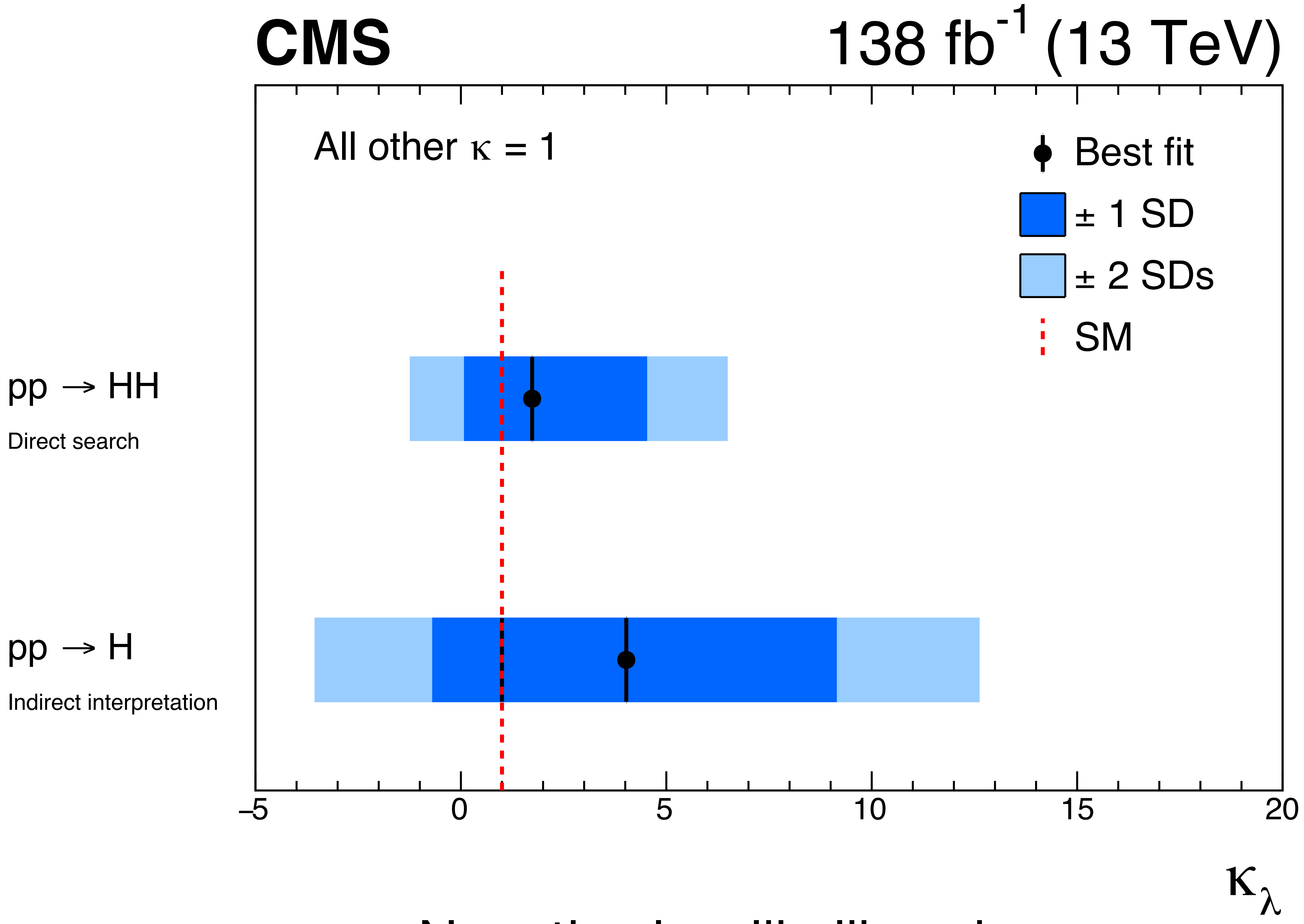
Obs. (exp.) κ_{2V}-constraint

-1.3 (-0.9) < κ_{2V} < 3.5 (3.1)

CMS HH Run-2 Combination: Extended

[Nature 607, 60–68(2022)]





Negative log-likelihood scan as a function of the κ_λ and κ_{2V} couplings

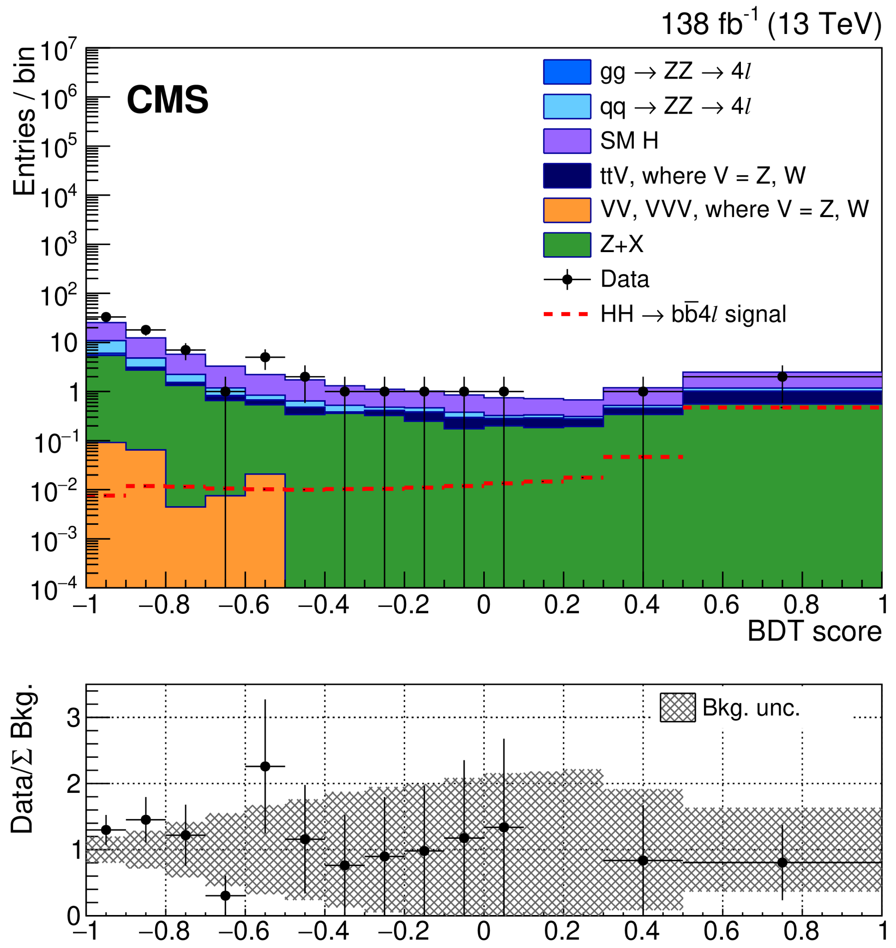
Negative log-likelihood scan as a function of the κ_λ

$\kappa_{2V} = 0$ hypothesis is excluded with 6.6 s.d. significance

CMS HH complementary results

bbZZ(4l)

JHEP 06 (2023) 130



Obs. (exp.) SM constraint

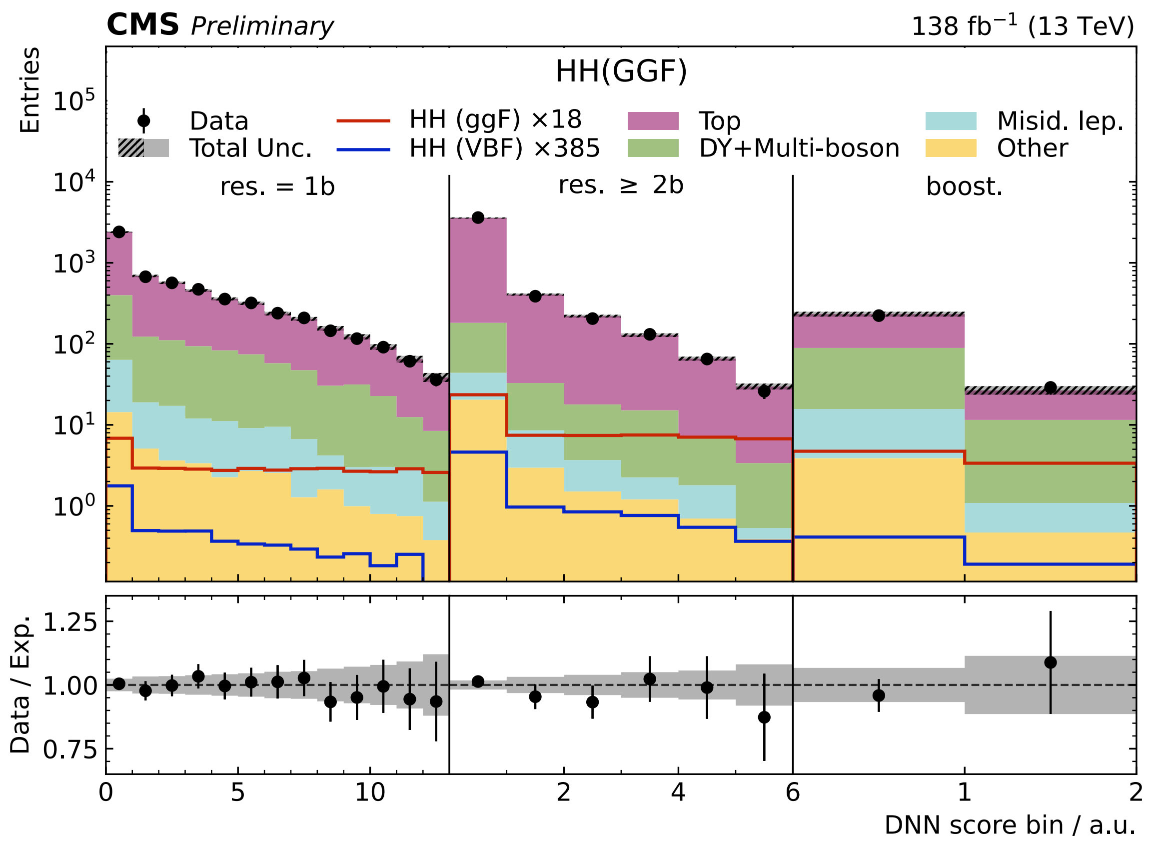
$\sigma/\sigma_{theory} < 32.4$

Obs. (exp.) κ_λ-constraint

$-8.8\ (-9.8) < \kappa_\lambda < 13.4\ (15.0)$

bbWW

CMS-PAS-HIG-21-014



Obs. (exp.) SM constraint

$\sigma/\sigma_{theory} < 14\ (18)$

Obs. (exp.) κ_λ-constraint

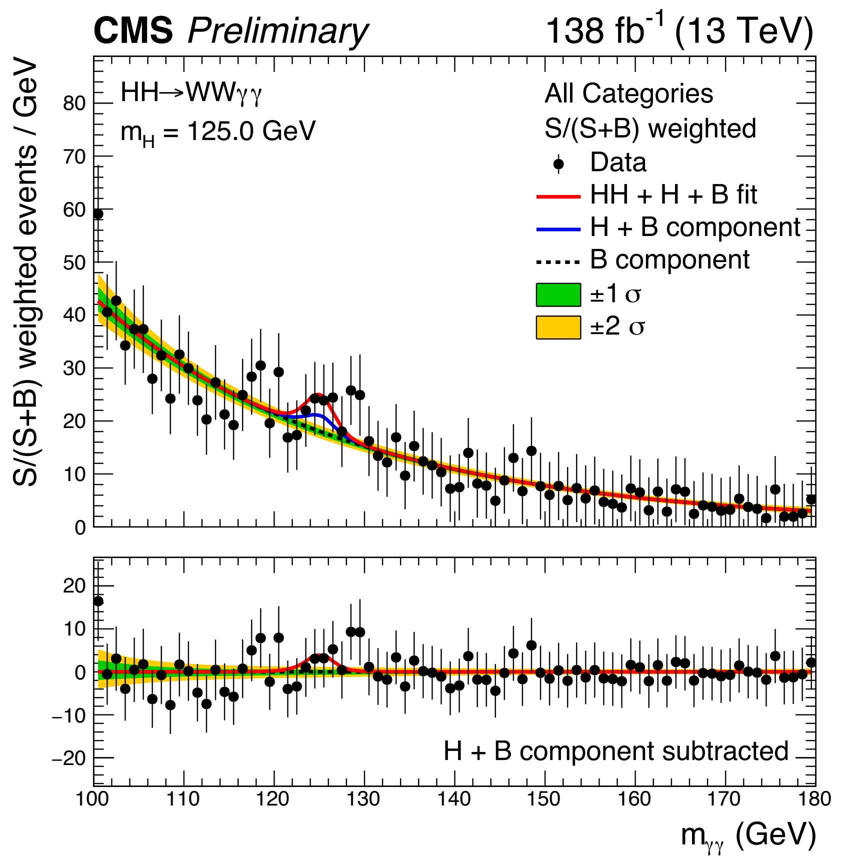
$-7.2\ (-13.8) < \kappa_\lambda < 13.8\ (15.2)$

Obs. (exp.) κ_{2V}-constraint

$-1.1\ (-1.4) < \kappa_{2V} < 3.2\ (3.5)$

WW$\gamma\gamma$

CMS-PAS-HIG-21-014



Obs. (exp.) SM constraint

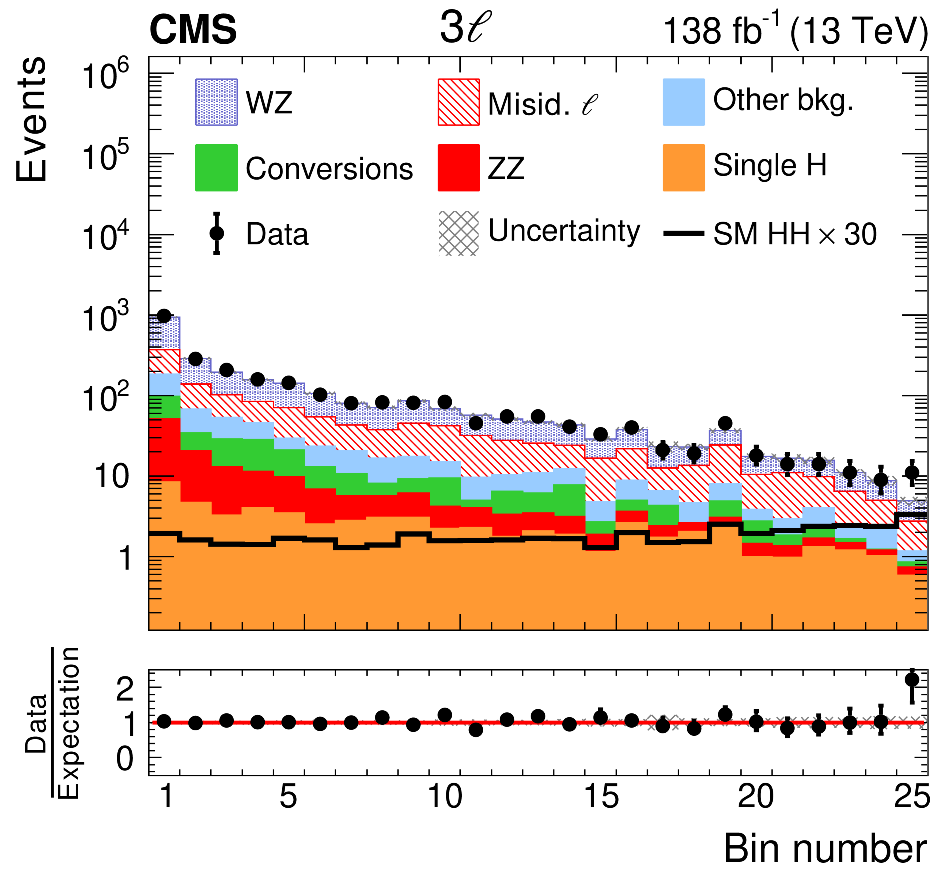
$\sigma/\sigma_{theory} < 97\ (53)$

Obs. (exp.) κ_λ-constraint

$-25.8\ (-14.4) < \kappa_\lambda < 24.1\ (18.3)$

Multilepton (4W,2W2τ,4τ)

arXiv:2206.10268



Obs. (exp.) SM constraint

$\sigma/\sigma_{theory} < 21.3\ (19.4)$

Obs. (exp.) κ_λ-constraint

$-6.9\ (-6.9) < \kappa_\lambda < 11.1\ (11.7)$

Also see the VHH(4b) result in **CMS-PAS-HIG-22-006**