HH→bbbb search with CMS

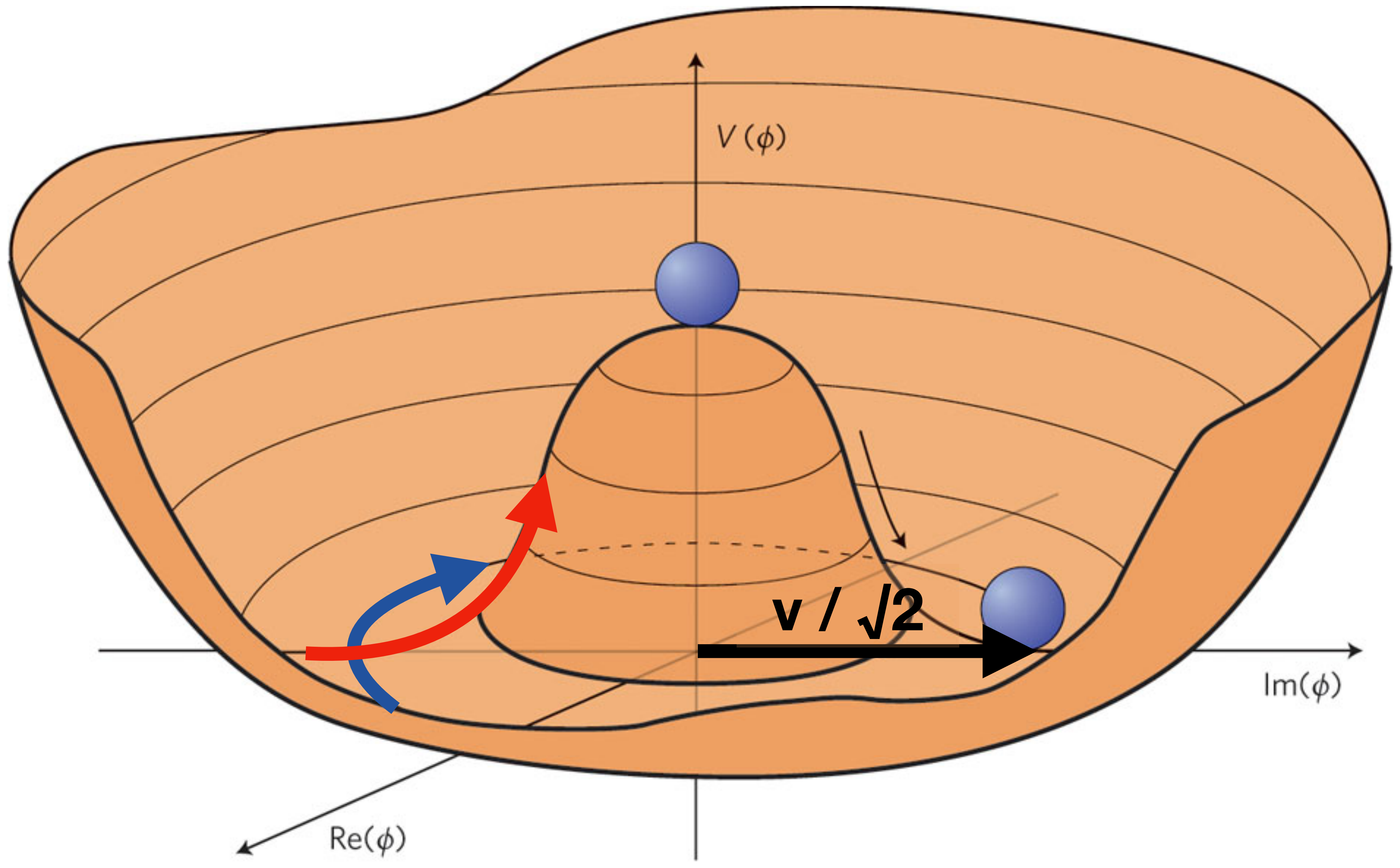
Based on CMS-PAS-HIG-20-005

Luca Cadamuro
on behalf of the CMS Collaboration

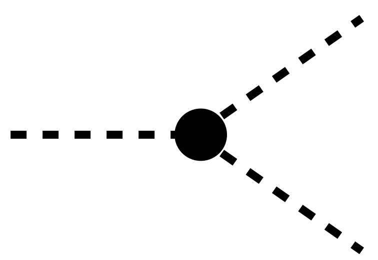
IDR Terascale, online
July 5th, 2021

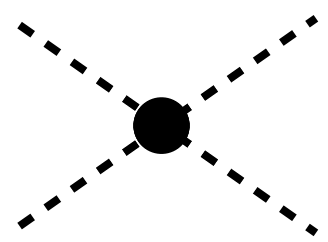
The scalar sector and the self-coupling

$$V(\Phi^\dagger\Phi) = -\mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2$$



$$V(H) = \frac{1}{2}m_{\mathrm{H}}^2 H^2 + \lambda_{\mathrm{HHH}} v H^3 + \frac{1}{4}\lambda_{\mathrm{HHHH}} H^4 - \frac{\lambda}{4}v^4$$





- The scalar sector: cornerstone of the SM
- Brout-Englert-Higgs mechanism: a scalar potential with a v.e.v. ≠ 0 originates a spontaneous breaking of the electroweak symmetry
- Properties of the scalar sector $\Longleftrightarrow$ BEH potential shape (λ) $\Longleftrightarrow$ self-coupling
- **Still largely unexplored at the LHC**

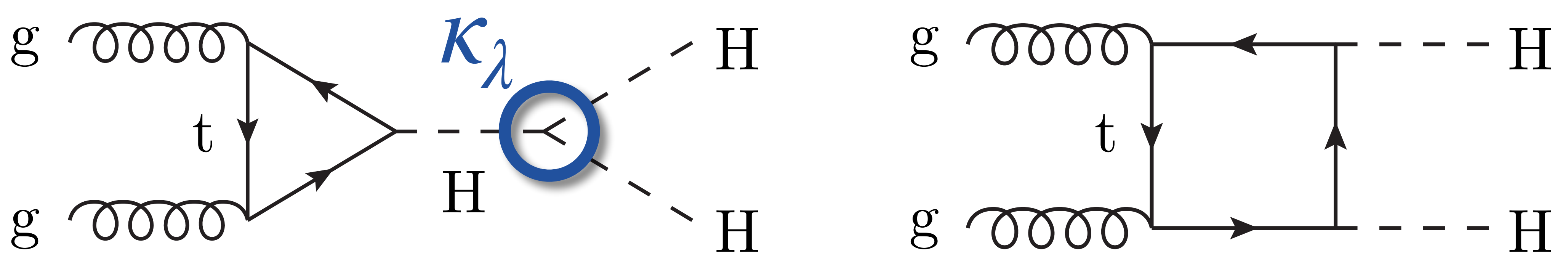
$$\lambda_{\mathrm{HHH}} = \lambda_{\mathrm{HHHH}} = \lambda = \frac{m_{\mathrm{H}}^2}{2v^2} \approx 0.13$$

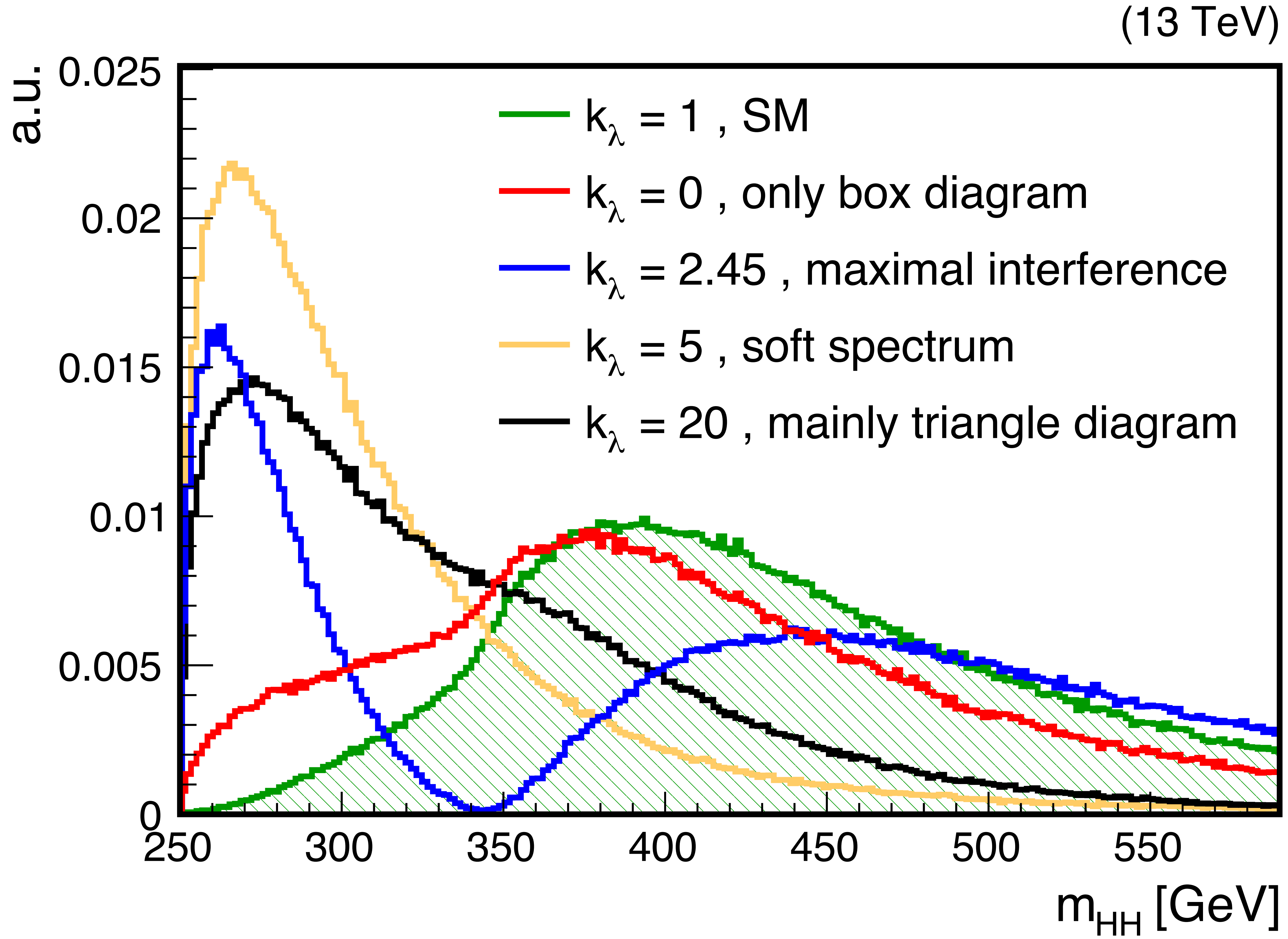
The Higgs boson self-coupling has a unique role in the SM

HH production: gluon fusion

NNLO FT-approx
JHEP 1805 (2018) 059

$$\sigma^{SM}_{ggF} = 31.05\ \text{fb}^{+6.7\%}_{-23.2\%}\ \text{(13 TeV)}$$





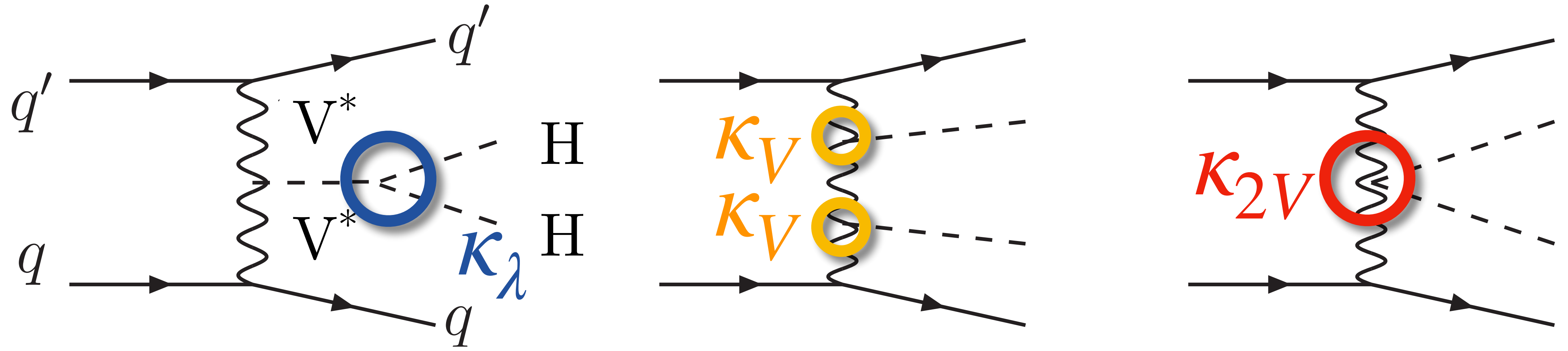
HH production ⟹ direct determination of Higgs trilinear coupling λ_{HHH}

- **Gluon fusion**: dominant production mode
- Large destructive interference
  ⟹ tiny cross section
- Self-coupling information both total and differential cross section (strong m_{HH} dependence on λ_{HHH})
  - analyses must be optimised for a variety of signal kinematics

HH production: vector boson fusion

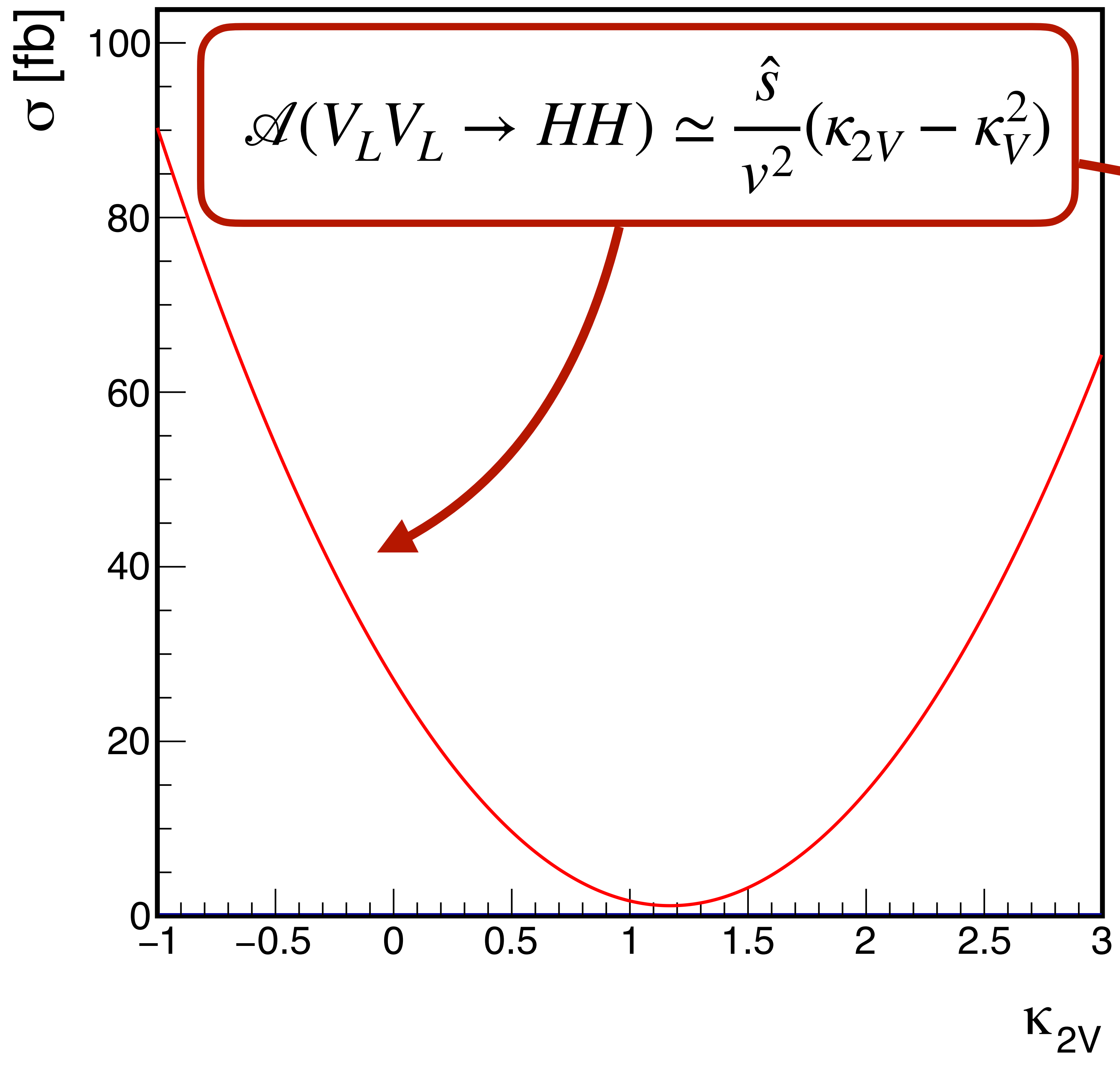
N³LO QCD
PRD 98, 114016 (2018)

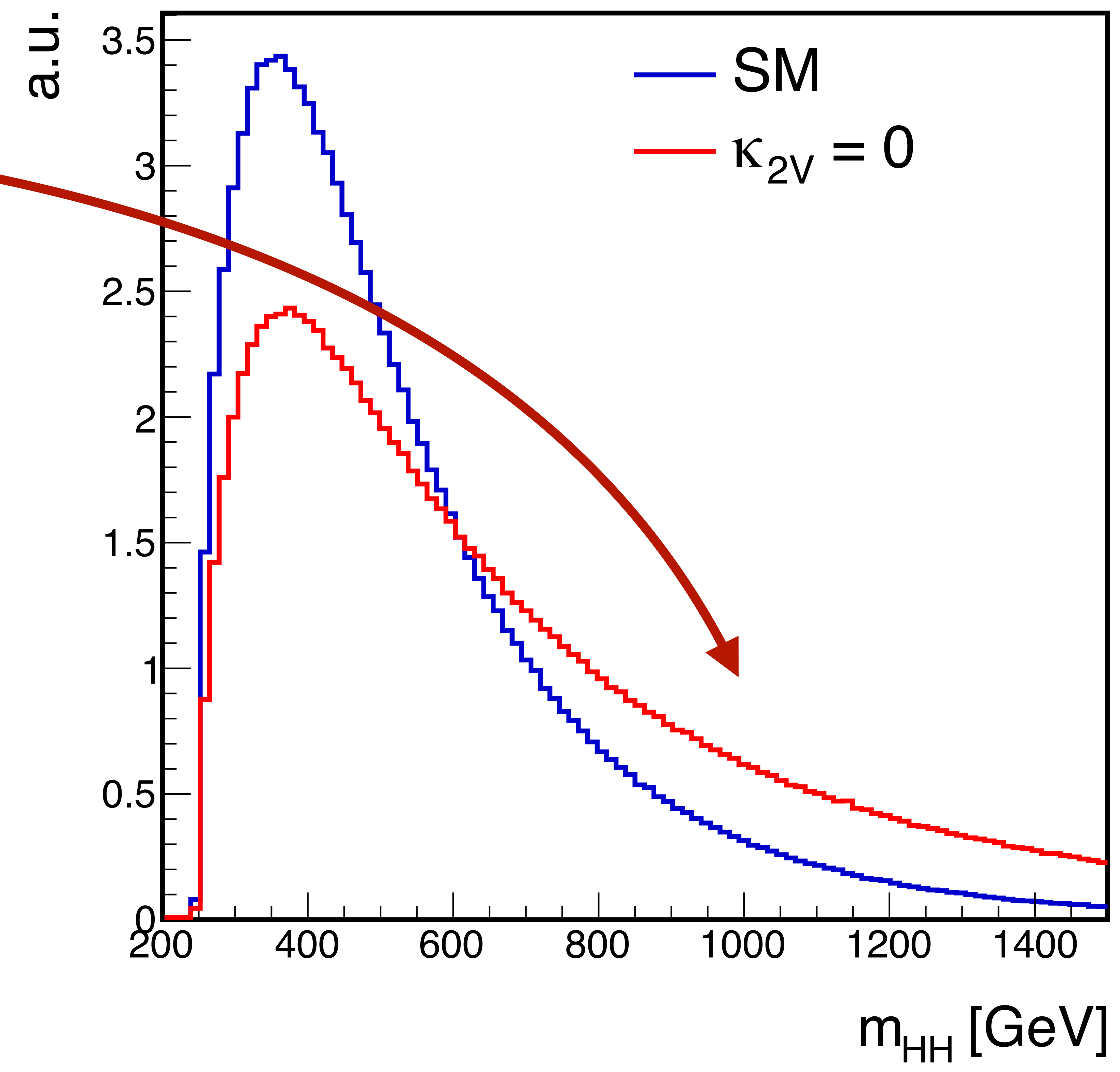
$$\sigma^{SM}_{VBF} = 1.73 \text{ fb} \pm 2.1\ \% \ (13 \text{ TeV})$$



- Very rare production mode
  - moderate sensitivity to λ
- Unique sensitivity to the VVHH interaction
  - $\kappa_{2V} \neq \kappa_V$ in e.g. composite Higgs models
  - longitudinal scattering opens when $\kappa_{2V} \neq \kappa_V$ → growth of xs at high m_{HH} values

$$\mathcal{A}(V_L V_L \to HH) \simeq \frac{\hat{s}}{v^2}(\kappa_{2V} - \kappa_V^2)$$

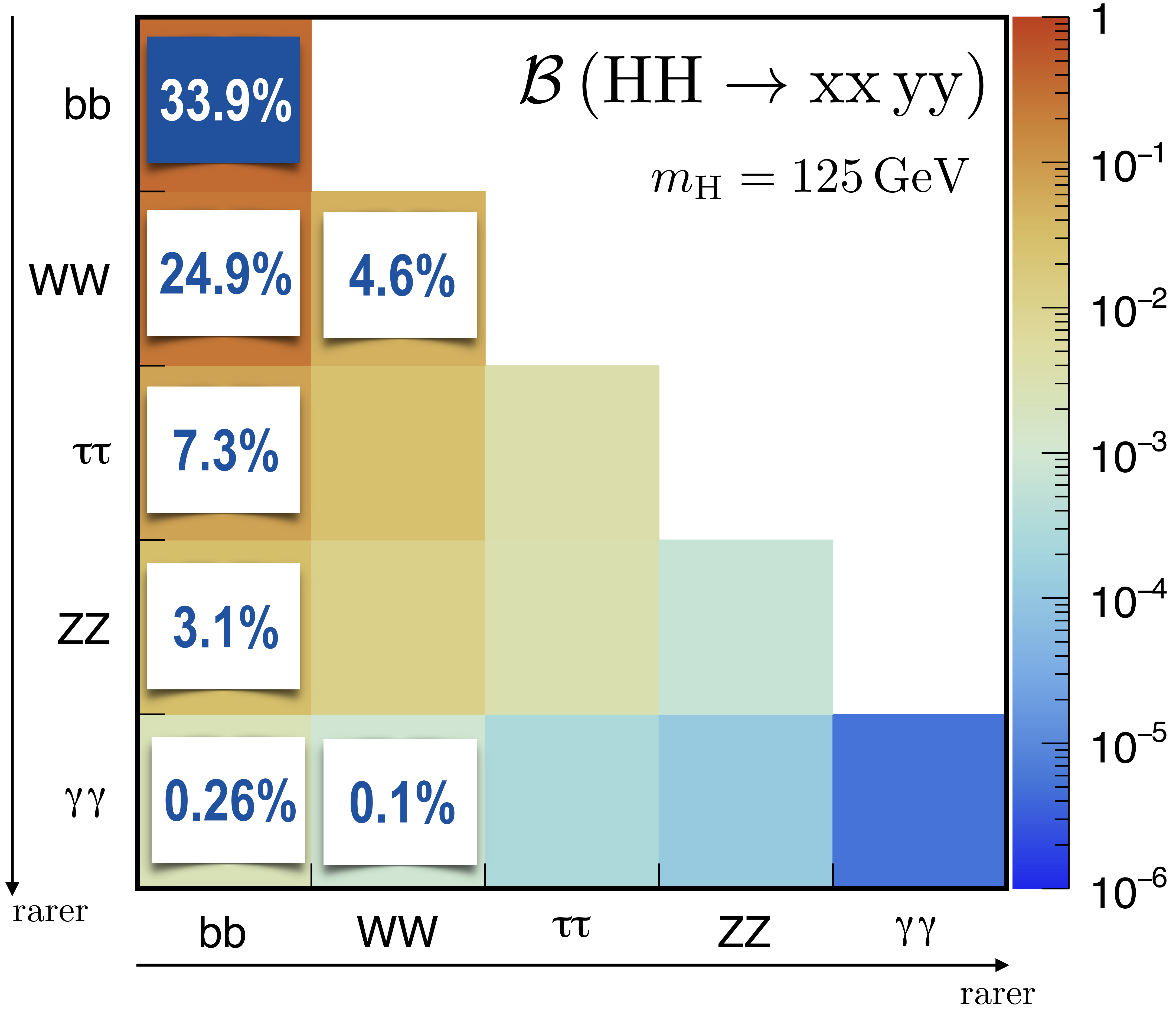




The 4b channel

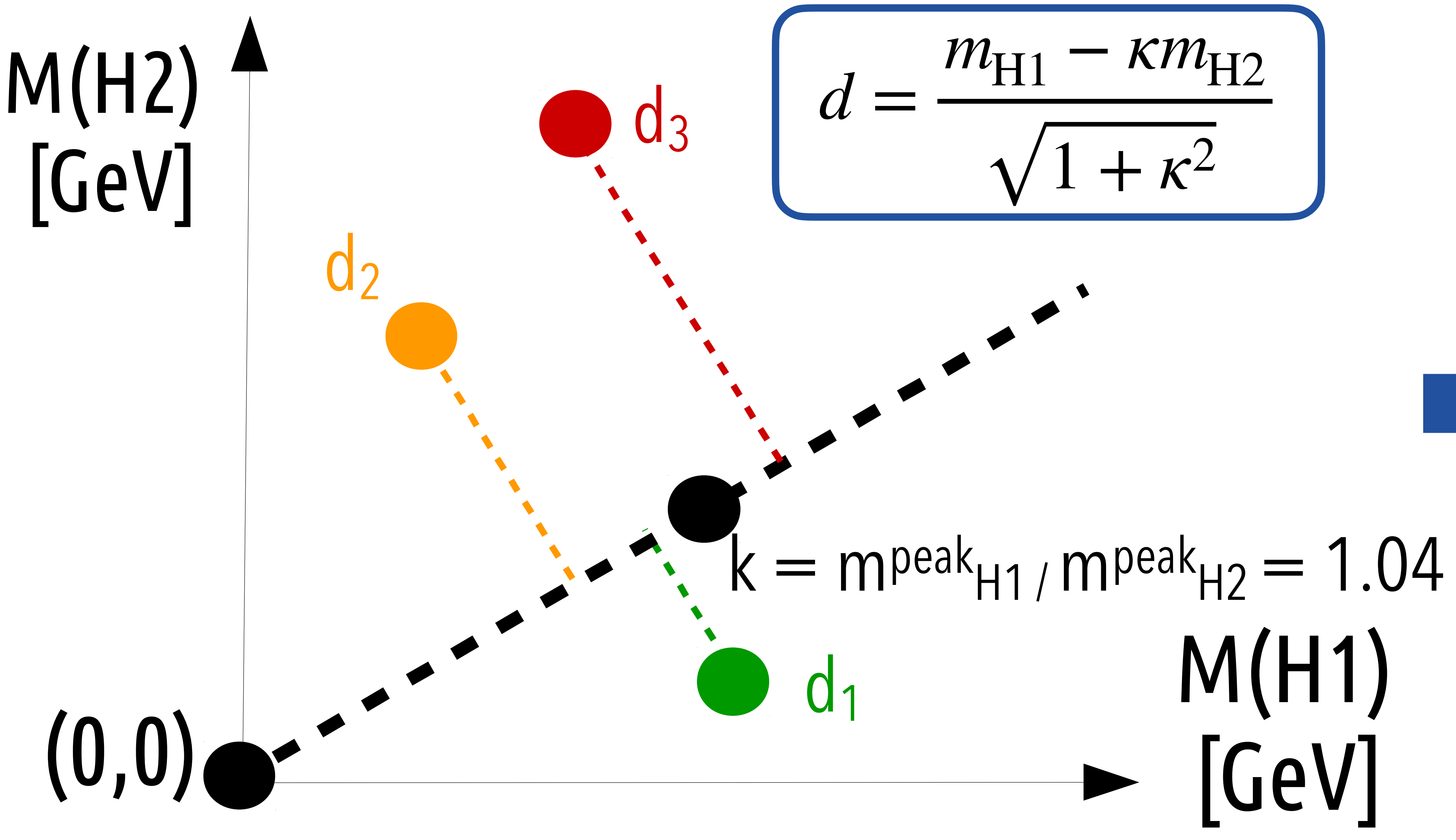
- ✔ Largest HH branching fraction
  - about 1500 HH→bbbb events expected in the Run 2 CMS dataset
- ⚠ Challenging multijet background
  - requires precise estimation and powerful rejection
- LHC 4b results
  - **CMS 2016 ggF** (*JHEP 04 (2019) 112*) — *36 fb^{-1}*
    *obs (exp) U.L. of **75 (37) x σ_{ggF}^{SM}***
  - **ATLAS 2016 ggF** (*JHEP 01 (2019) 030*) — *27 fb^{-1}*
    obs (exp) U.L. of **13 (21) x σ_{ggF}^{SM}**
  - **ATLAS Run 2 VBF** (*JHEP 07 (2020) 108*) — *126 fb^{-1}*
    840 (540) x σ_{VBF}^{SM}
    $-0.8 < \kappa_{2V} < 2.9$ ($-0.9 < \kappa_{2V} < 3.1$)
- **Today**: latest full Run 2 CMS result (CMS-PAS-HIG-20-005) — ***137 fb^{-1}***

XX % : current public results at sqrt(s) = 13 TeV

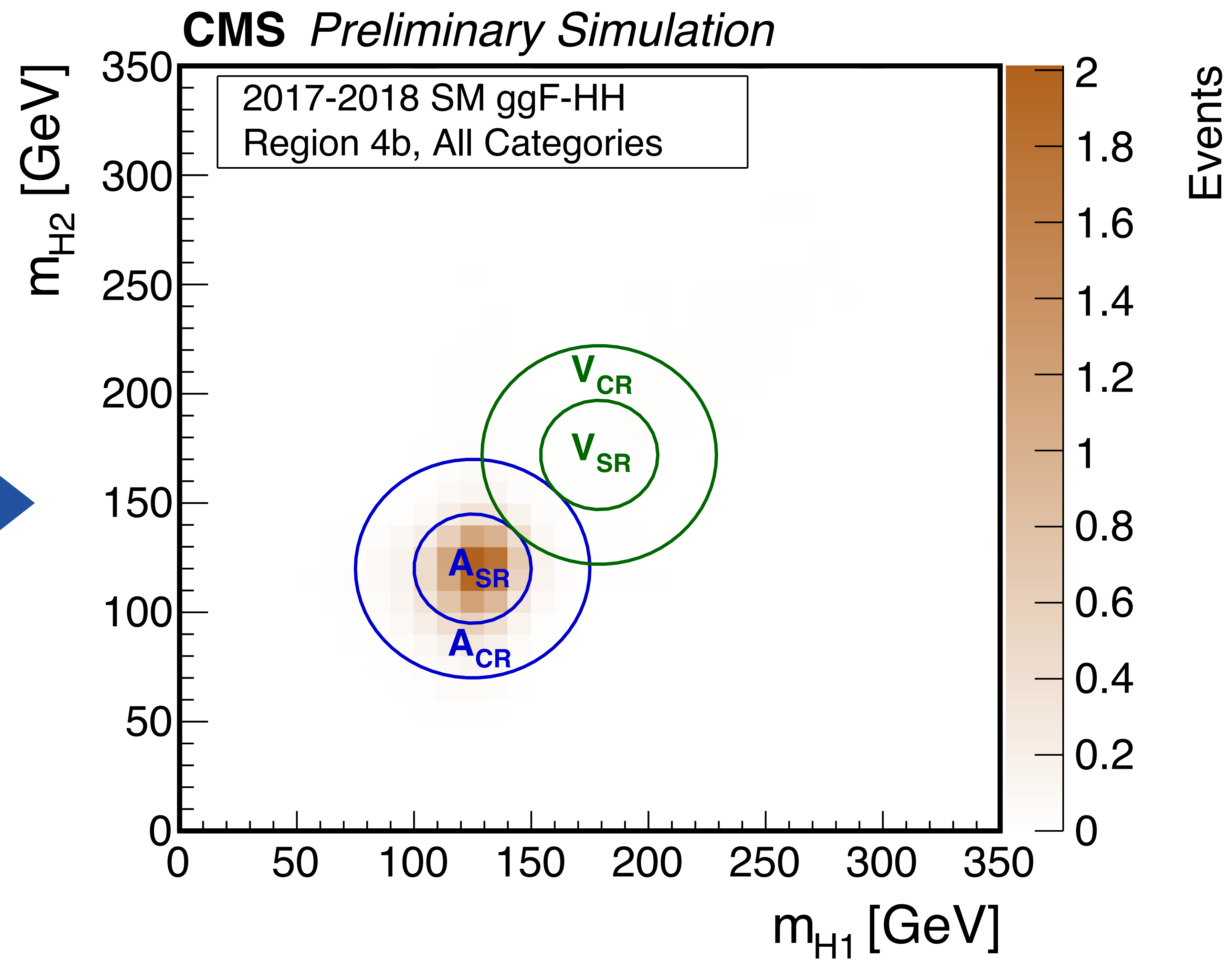


Reconstructing the H candidates

- 4jet+3b trigger, offline preselection of the four jets with the highest b-tag score (≥3 b tagged)
- Three possible pairings of the four b jets exist $\Longrightarrow$ exploit the "equal-mass" hypothesis
  - if $\Delta d = d_2 - d_1 > 30$ GeV, select d_1 pair
  - otherwise, select between pairs d_1 and d_2 the one giving the highest $p_T(H)$ in the 4b rest frame
- Achieve correct pairing in 82-98% of events

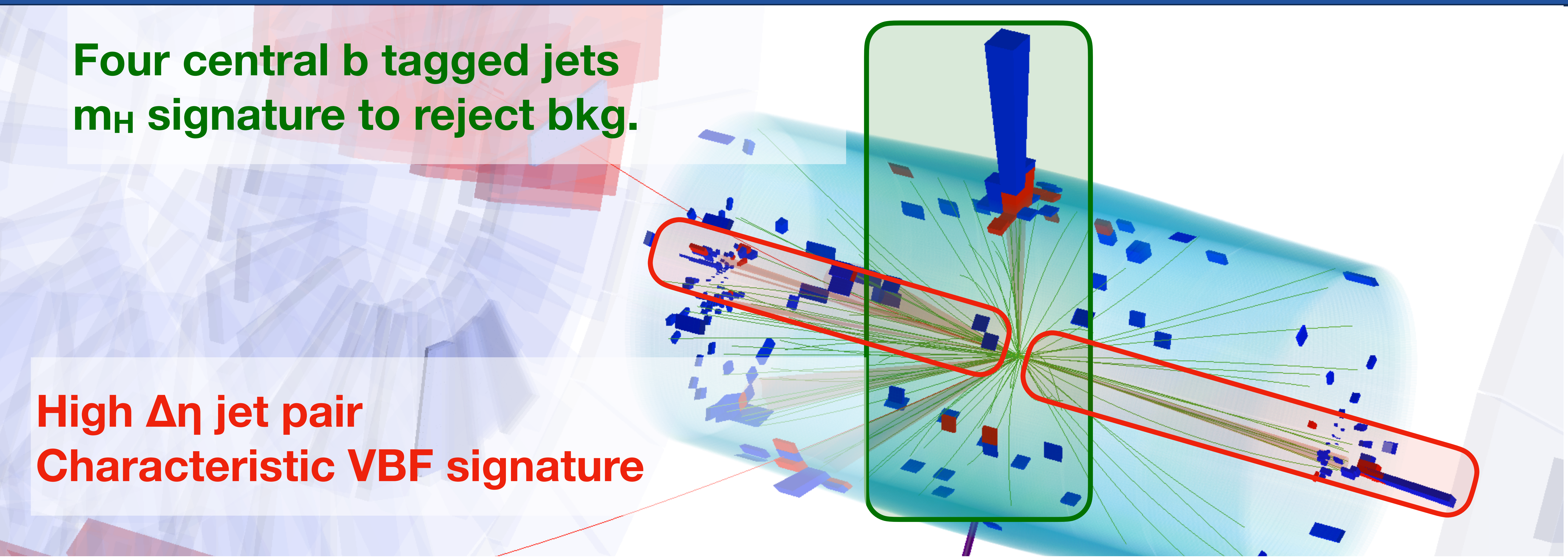


$$d = \frac{m_{H1} - \kappa m_{H2}}{\sqrt{1+\kappa^2}}$$



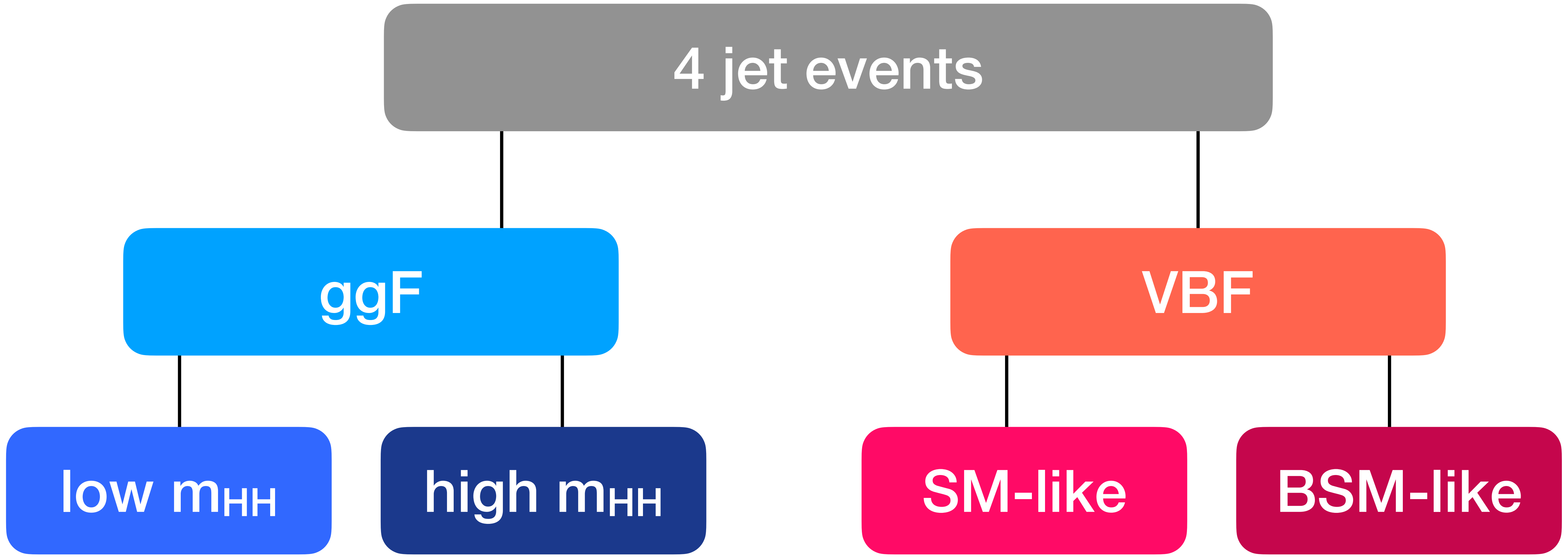
Optimal performance without biasing the bkg events

Event categories



- VBF events contain two additional jets with $\eta_1 \times \eta_2 < 0$
- A BDT is trained to separate misclassified ggF + 2 jets events
  - use kinematic properties of jet and H candidates
  - 97% of ggF events correctly classified

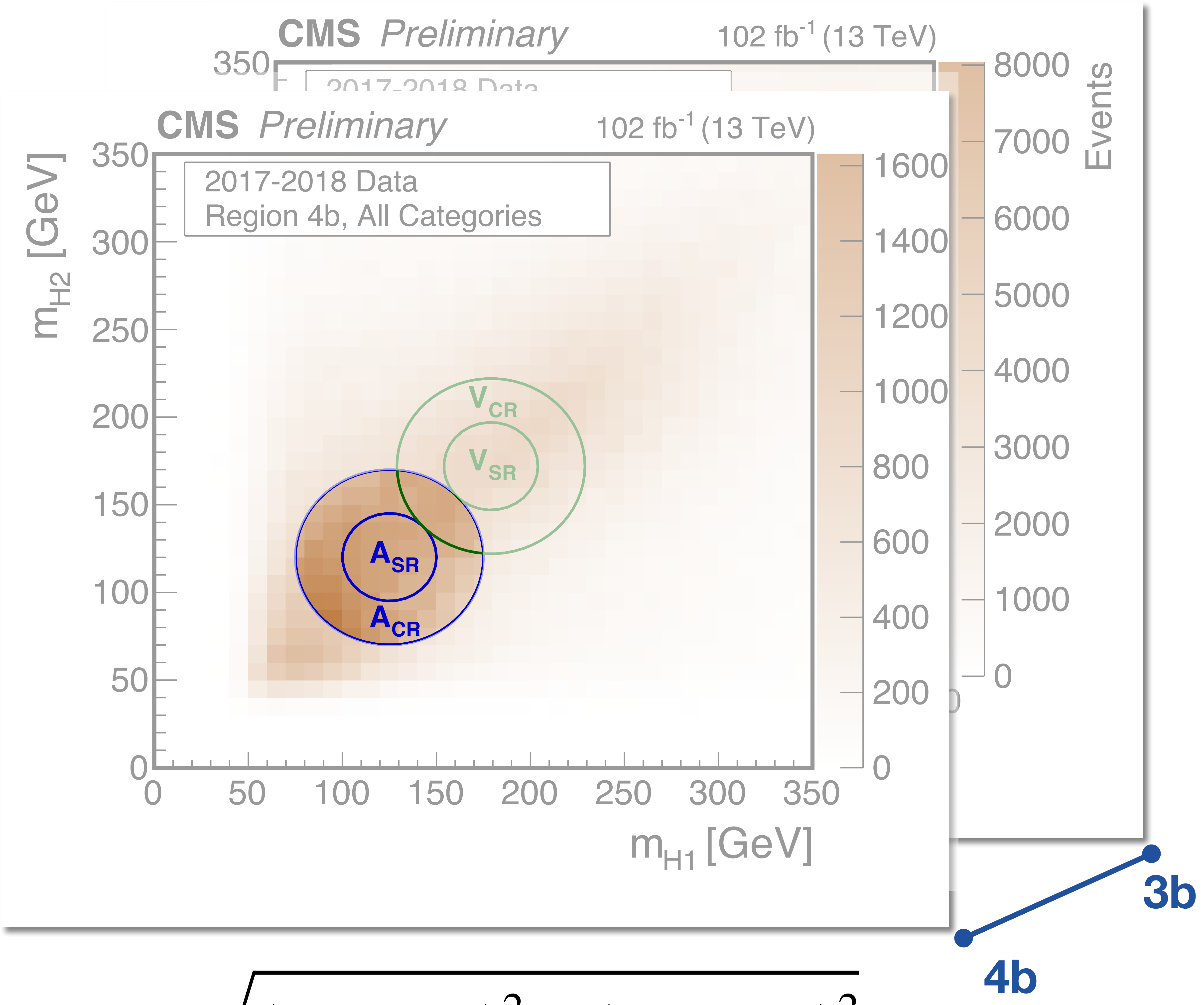
- ggF events split in low- and high-m_{HH} (450 GeV) to capture κ_λ dependence
- VBF events split in SM-like and BSM-like based on BDT score to enhance anomalous κ_{2V} contribution



Background normalization

- Signal region (SR): $\chi < 25\,\mathrm{GeV}$
- Control region (CR): $25 < \chi < 50\,\mathrm{GeV}$
- Data are divided into a 3b and a 4b sample
  - 5-10× more data in 3b w.r.t. 4b
- Background yield = $N_{CR}^{4b}/N_{CR}^{3b} \times N_{SR}^{3b}$

Background yield determined from data

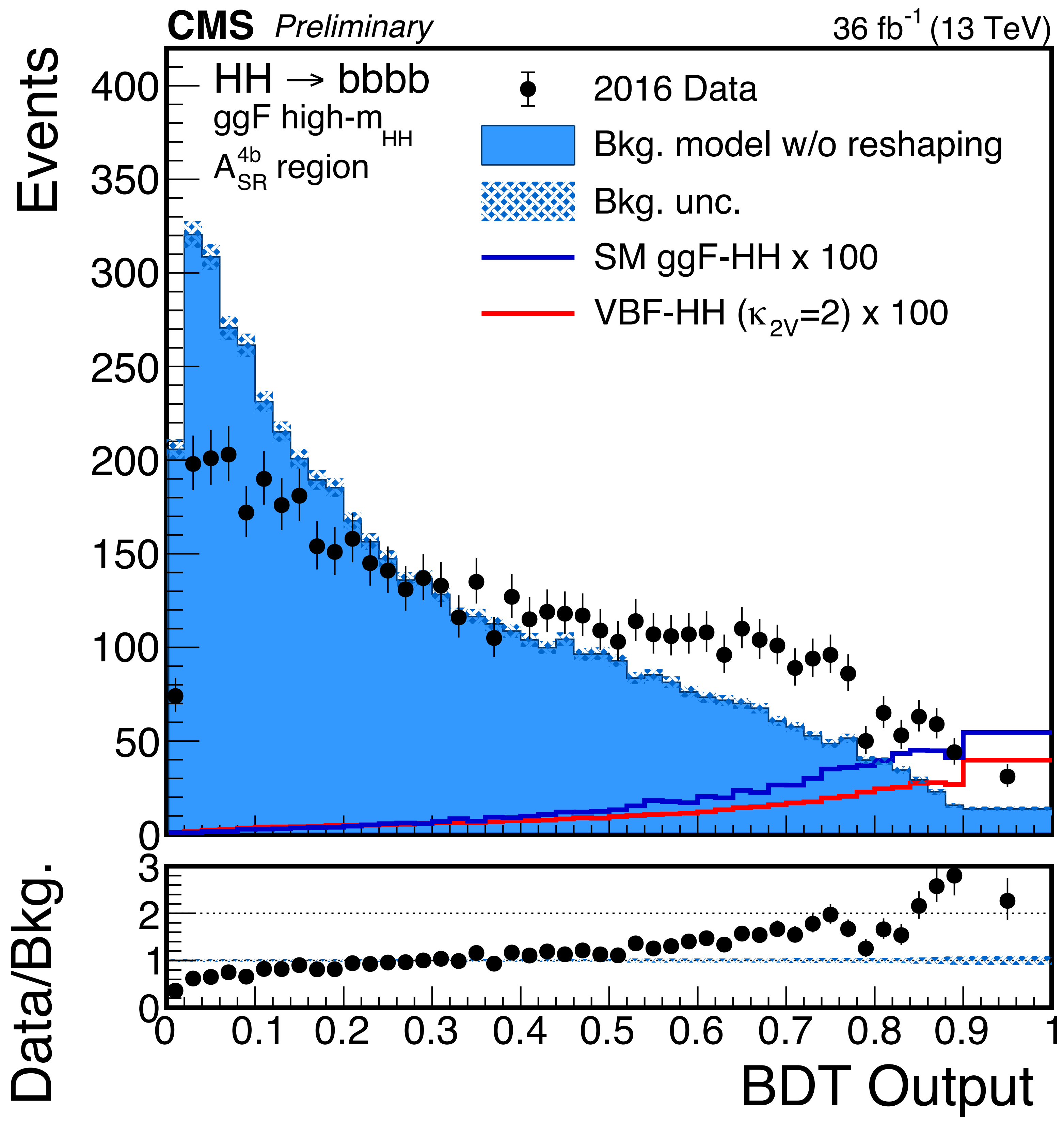


$$\chi = \sqrt{(m_{H1} - c_1)^2 + (m_{H2} - c_2)^2}$$

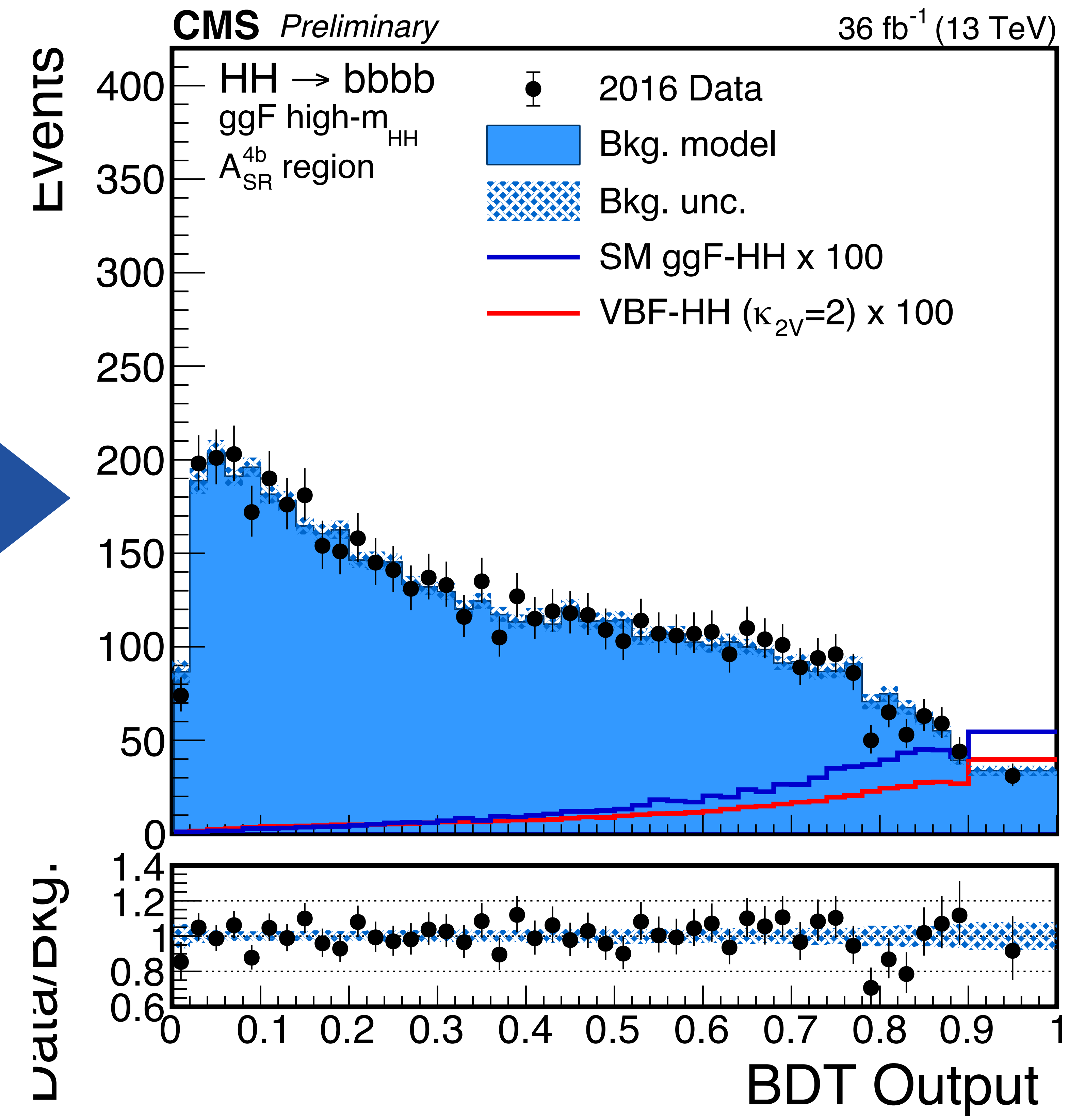
Modelling the background shape

- BDT with dedicated metric (J. Phys.: Conf. Ser. 762 012036) trained to separate CR^{4b} from CR^{3b} data
  - inputs: kinematic and object quality
  - stat. test + test discriminator used as metrics in hyperparameters tuning until no separation achieved (closure check)
- Score used to reweight SR^{3b} data to model SR^{4b}

Leverage on ML techniques to achieve multidimensional data-driven estimate



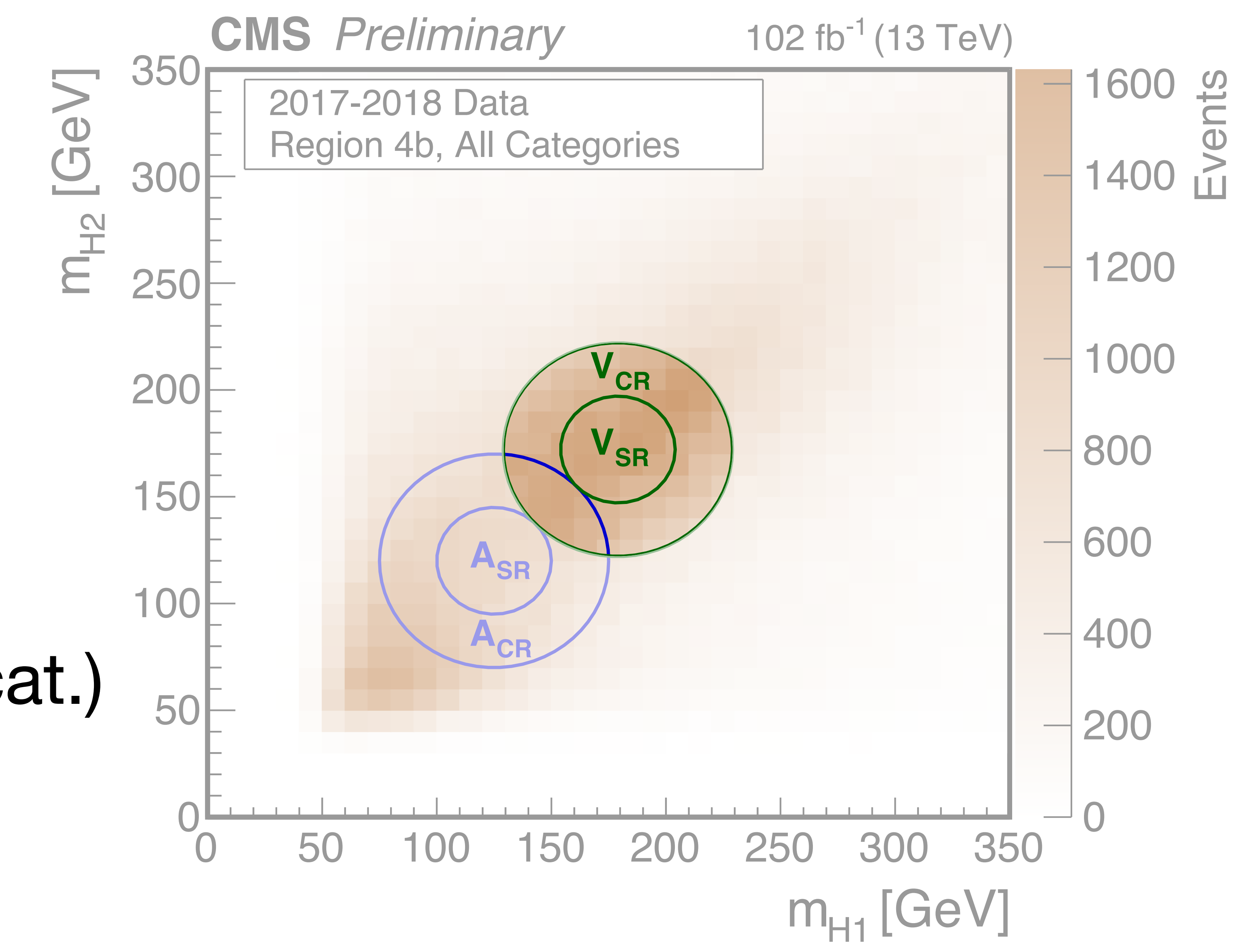
Before reshaping
(data directly from the 3b region)

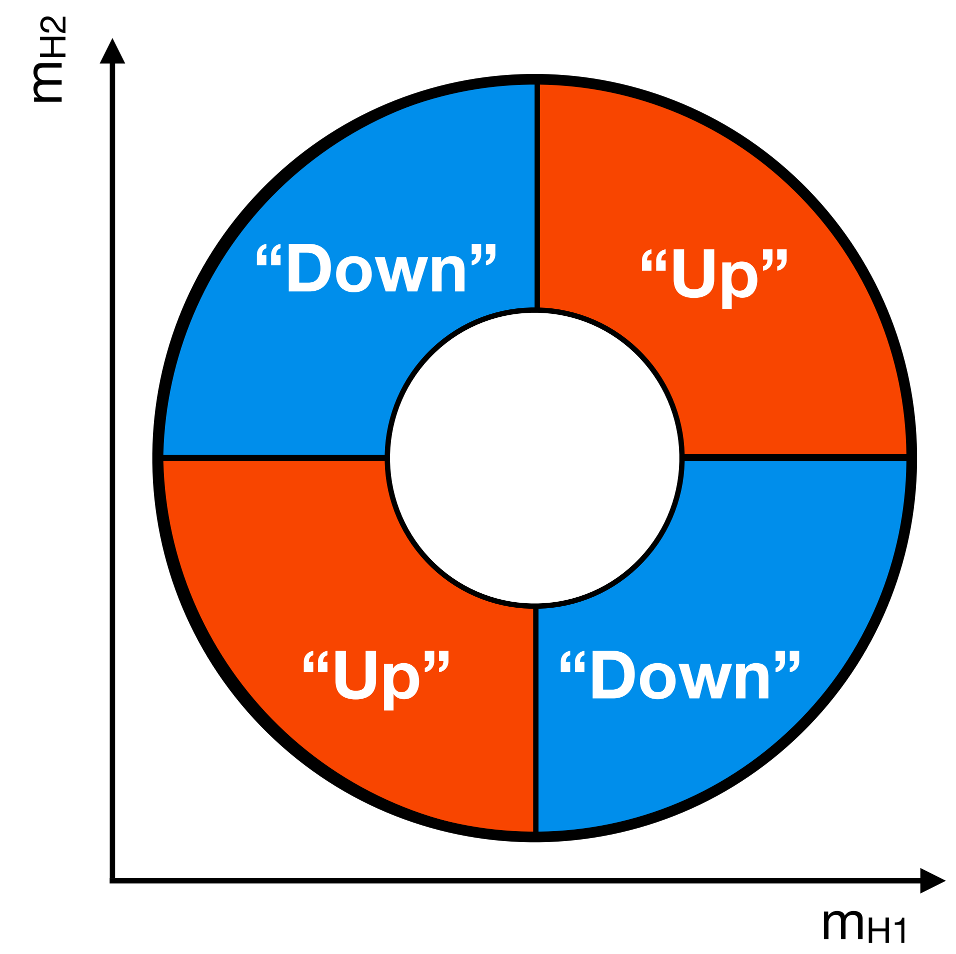


After reshaping

Validating the background model

- Signal-free validation region (VR) used
  - apply same methods as in the SR
  - VR shifted along the (m_{H1}, m_{H2}) diagonal → no bias from H reconstruction
- Good statistical agreement for all variables observed in VR
  - add uncertainties for total yields non-closures (1.5-4.7%)
  - uncertainties for the validation vs analysis region statistics (3-30% for VBF cat.)
- Additional SR uncertainties considered on the background templates
  - bin-bin-bin template variations (poisson counts in 3b data)
  - CR statistical uncertainties
  - alternative bkg. templates from trainings in sub-portions of the CR



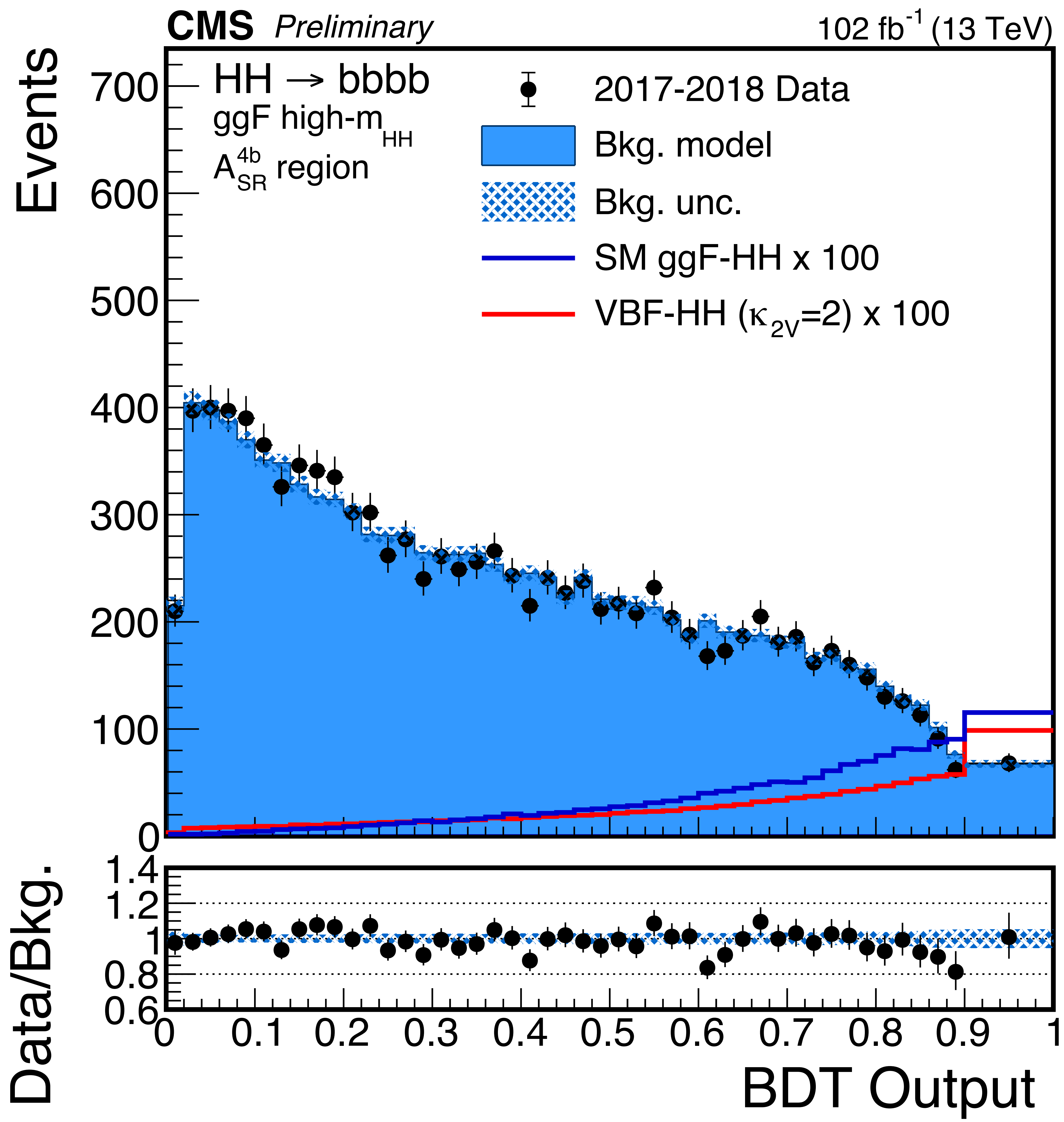


Good performance of bkg estimation method validated with data

Suppressing the multijet background in ggF

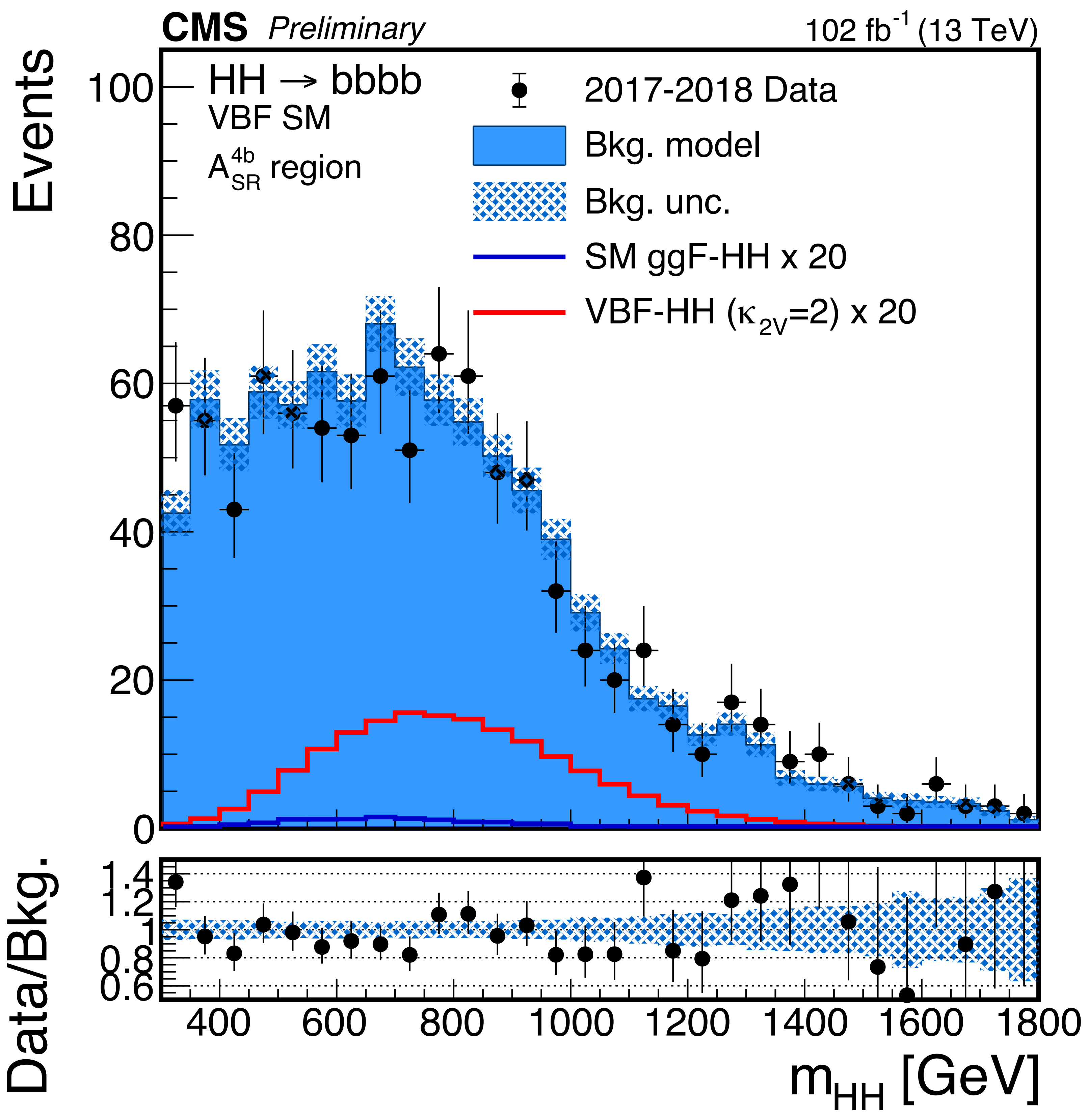
- BDT to separate the multijet background in ggF
- Training done on 3b reweighted data
  - 2 trainings performed with same parameters on 50% of the dataset and applied to the other 50%
  - separate training for every year/category
- Input features
  - kinematic variables: H_1, H_2, HH p_T and mass, 4 jet p_T sum
  - topological quantities: $\Delta\eta$(HH), ΔR(b,b) in a H candidate, min ΔR(4b), max $\Delta\eta$(4b), angles of b and H in the 4 jet rest frame
  - object quality variables: number of tight b jets, sum of the resolution estimator of the 3 jets

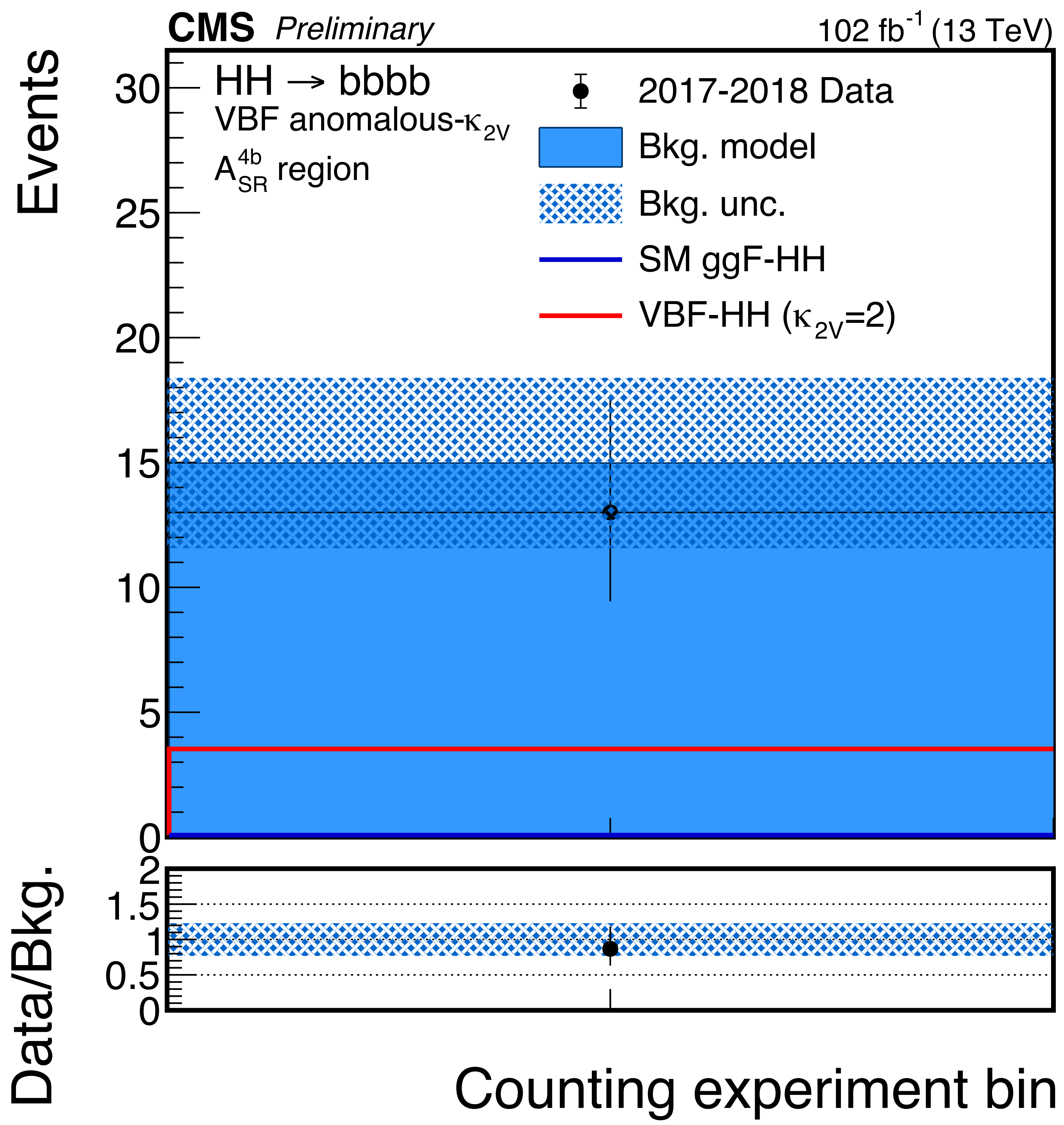
Effective separation of the multijet background



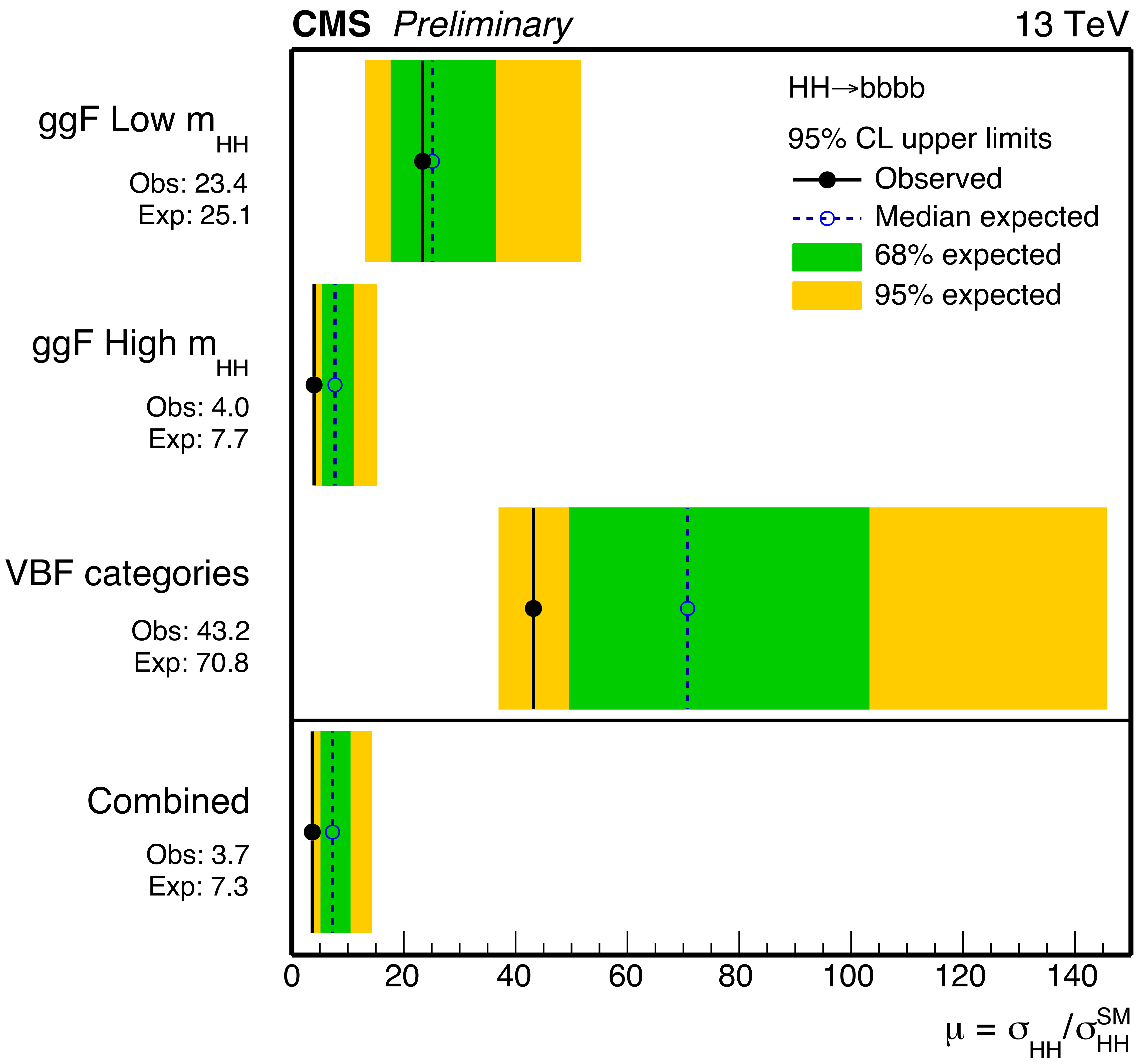
Signal extraction in VBF categories

- Fit m_{HH} in the SM-like category
  - statistics is too small to train a dedicated discriminant
- Counting experiment in the BSM-like category
  - high S/B for anomalous κ_{2V} events, O(10) bkg events





Results - SM production

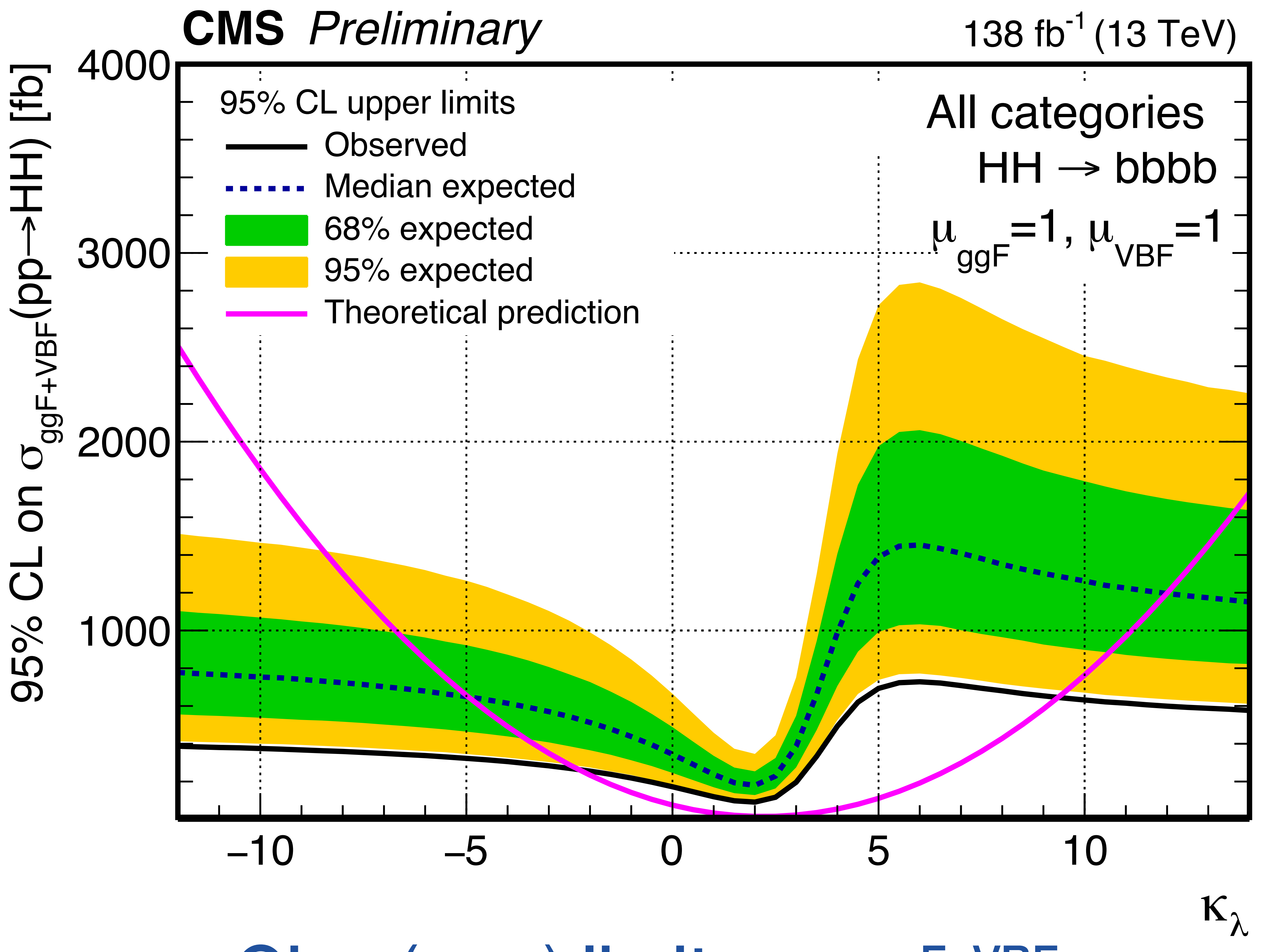


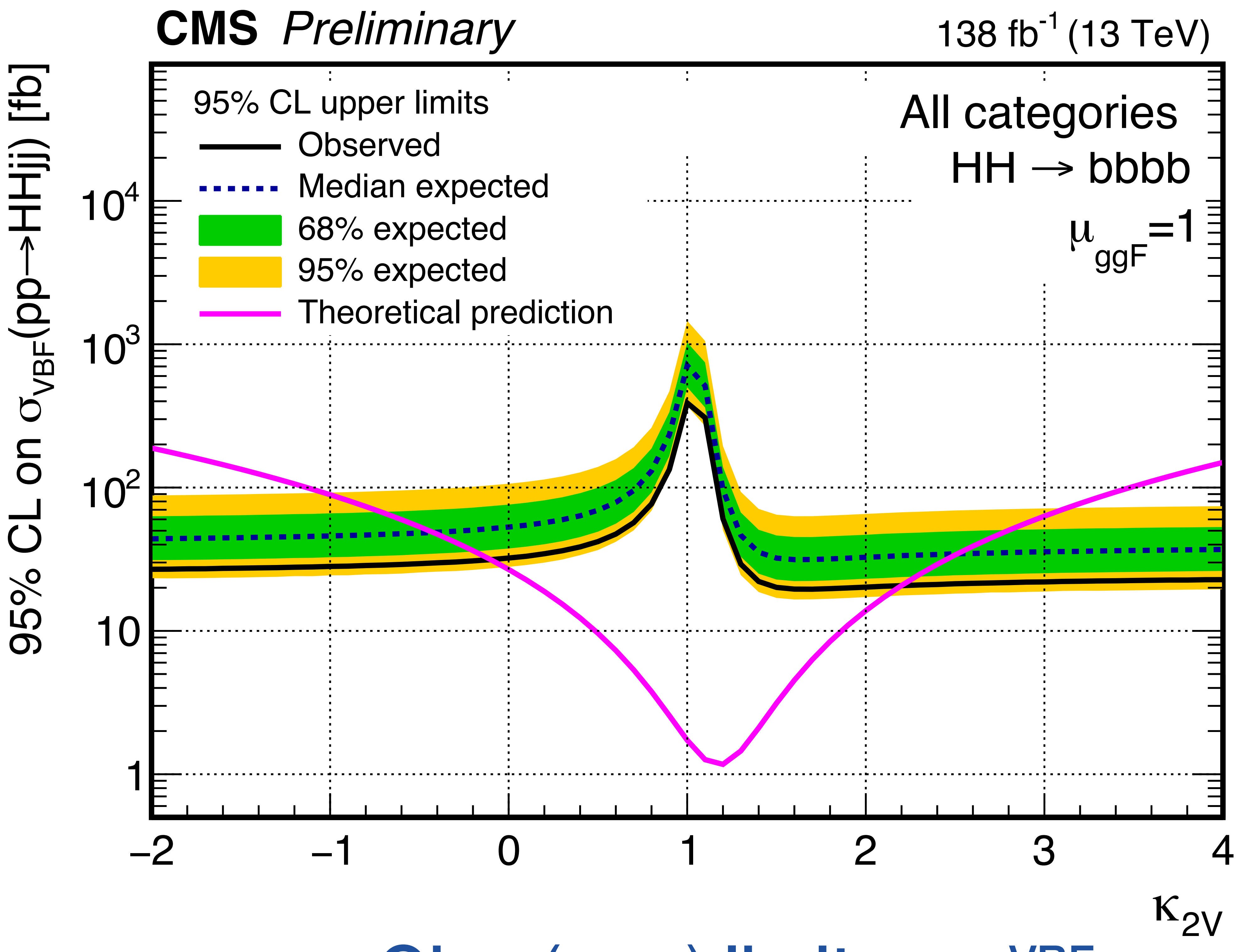
- The high m_{HH} ggF category leads the sensitivity to SM production
  - low m_{HH} ggF contributes to constrain anomalous κ_λ
  - VBF categories constrain the κ_{2V} coupling
  - note: a common signal strength to ggF and VBF is assumed here

Observed (expected) 95% CL UL
3.6 (7.3) × SM

Best constraint to date on the SM HH production

Results - couplings



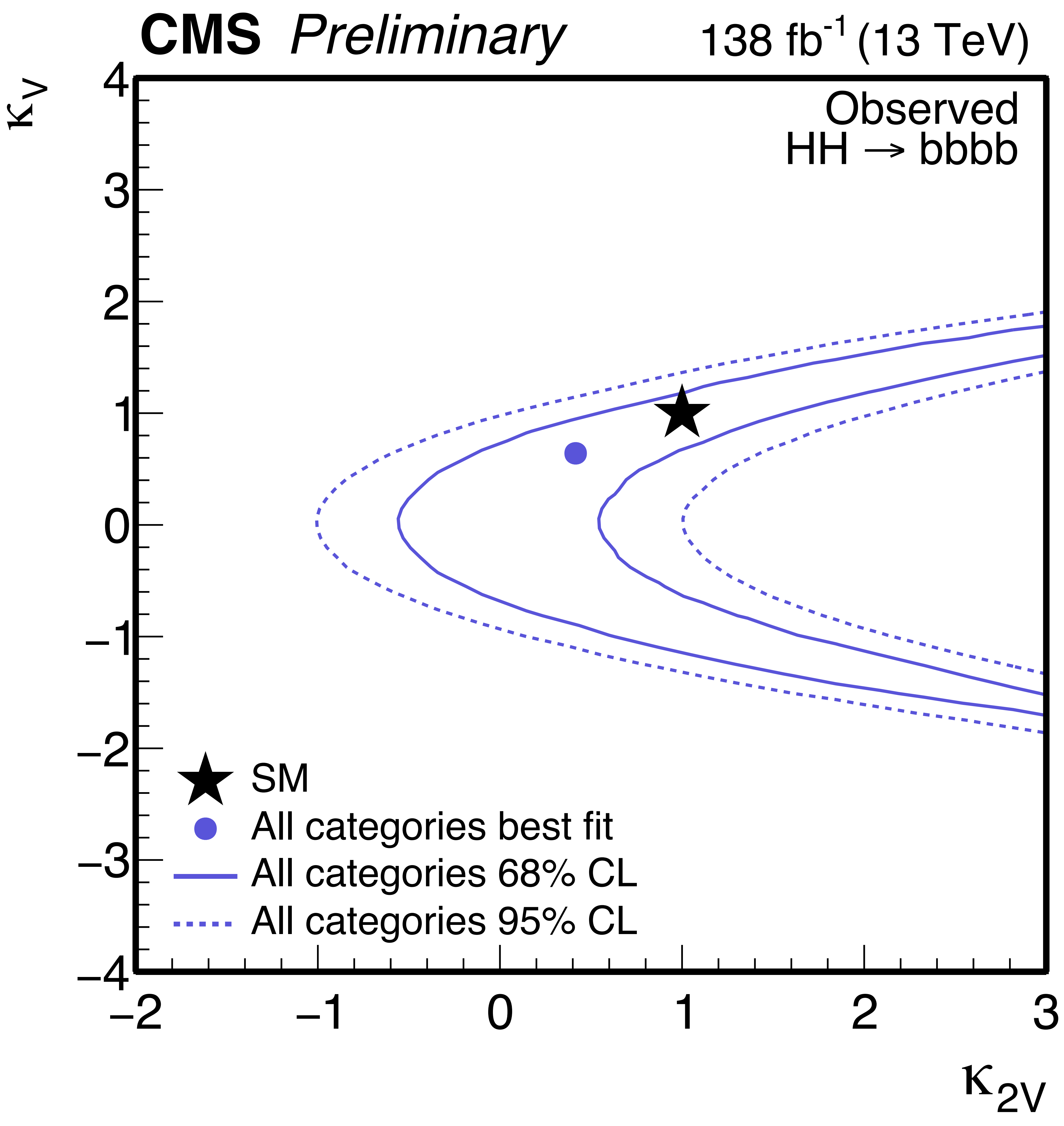


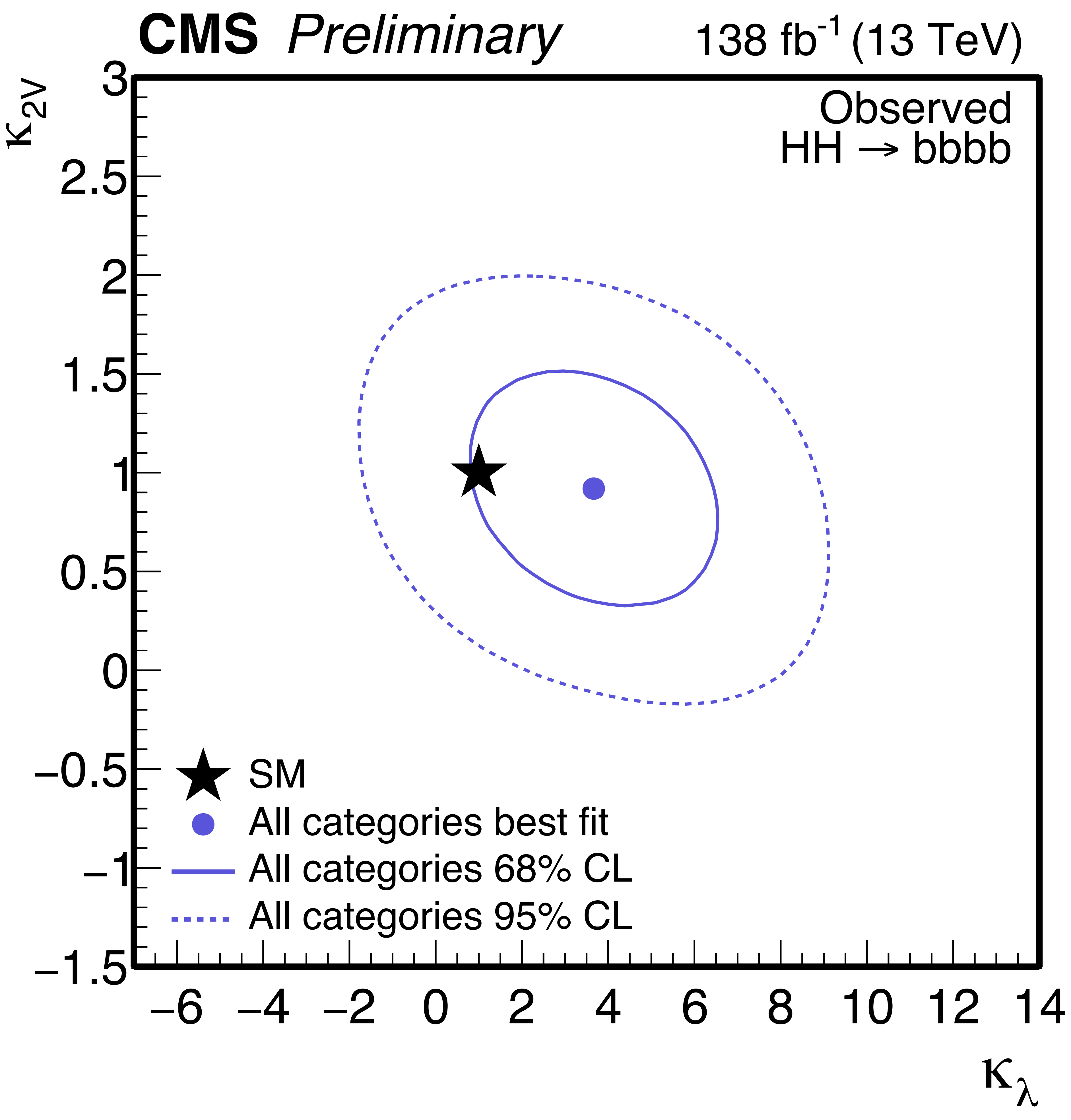
Obs. (exp.) **limit on $\sigma^{ggF+VBF}$**
3.6 (7.3) × **SM**
$-2.3 < \kappa_\lambda < 9.4$ ($-5 < \kappa_\lambda < 12$)

Obs. (exp.) **limit on σ^{VBF}**
226 (412) × **SM**
$-0.1 < \kappa_{2V} < 2.2$ ($-0.4 < \kappa_{2V} < 2.5$)

Best limits to date on κ_{2V}

Studying the couplings

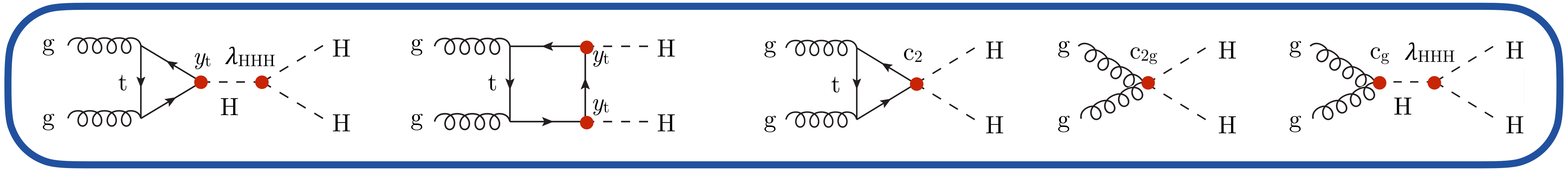


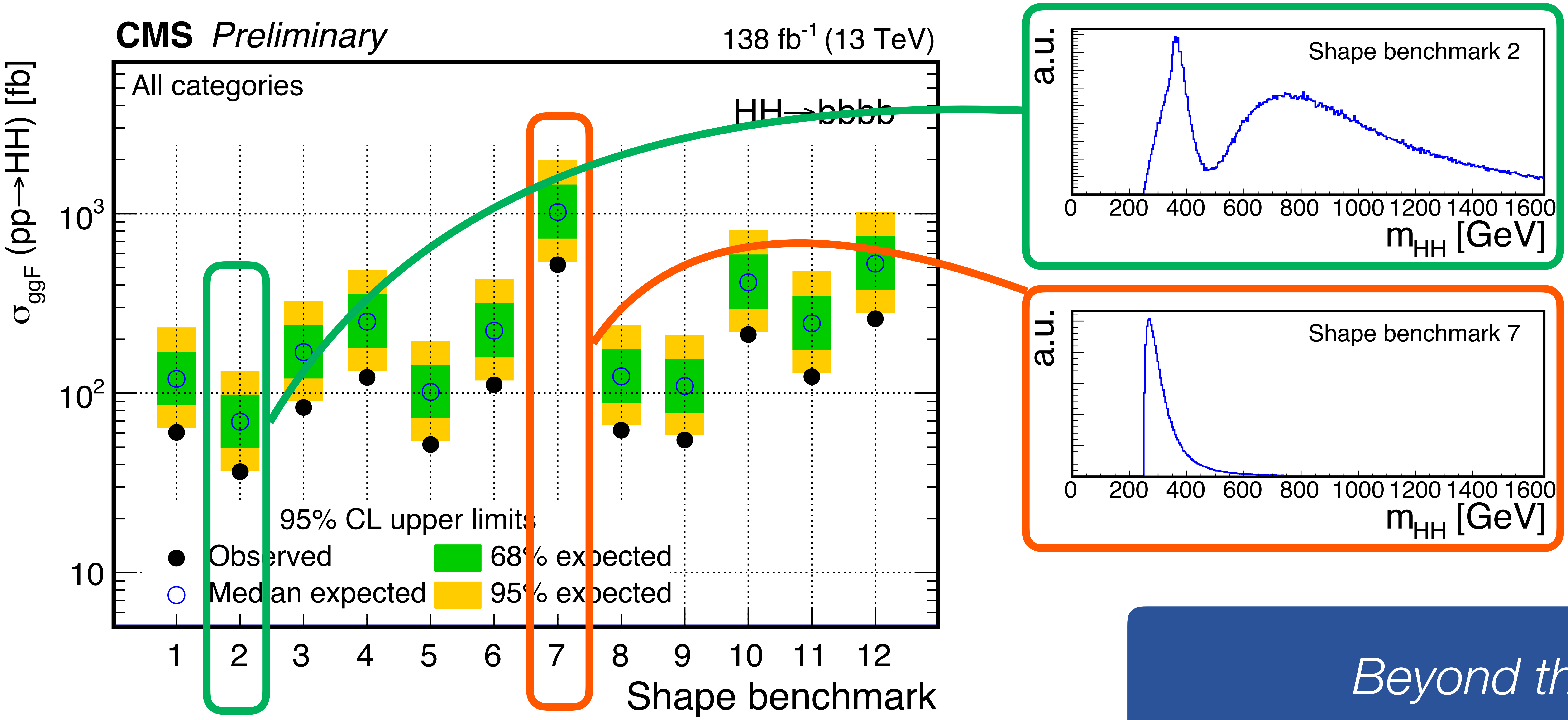


- Simultaneous fits of the couplings performed
  - couplings not displayed are kept to their SM value

A broader BSM picture

$$\mathcal{L} = \mathcal{L}_{\rm SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^6 + \cdots$$





- 5D parameter space in a generic EFT description
  - sampled with benchmark points with characteristic kinematic properties
- Ongoing work to define an approach for complete EFT scan in HH analyses

Beyond the self-coupling (only):
HH as a probe of high energy BSM effects

Conclusions ?

- **HH: a key topic** in the exploration of the scalar sector
  - direct access to the self-coupling
  - probe of high-energy new physics effects in anomalous couplings
- HH→bbbb: **largest BR**, but highly challenging **multijet background**
- CMS developed a **new full Run 2 analysis** to optimally benefit of the recorded dataset
  - ggF + VBF production modes
  - optimised selection and categorization of events
  - accurate data-driven estimation method
  - powerful discriminants to reject the multijet background
- Best constraints to date: obs. (exp.) **3.6 (7.3)** × **SM**
  - quickly approaching the SM sensitivity with upcoming combinations!

Conclusions

HH→bbbb was considered hopeless just a few years ago

A few examples
Papers from 6-8 years ago

In total, inclusive dihiggs production with decay to four b quarks has a signal-over-background ratio S/B which is too bad to be a suitable search channel, al-

As concerns the various decay channels, although the $4b$ final state is the dominant one, it suffers from huge QCD background. The most promising channel at the LHC is thought to be the rare decay

Channel	Significance Stat. + syst.	Significance Stat. only	95% CL limit on $\sigma_{HH}/\sigma_{HH}^{SM}$ Stat. + syst.	95% CL limit on $\sigma_{HH}/\sigma_{HH}^{SM}$ Stat. only
bbbb	0.95	1.2	2.1	1.6
$bb\tau\tau$	1.4	1.6	1.4	1.3
$bbWW(\ell\nu\ell\nu)$	0.56	0.59	3.5	3.3
$bb\gamma\gamma$	1.8	1.8	1.1	1.1
$bbZZ(\ell\ell\ell\ell)$	0.37	0.37	6.6	6.5
Combination	2.6	2.8	0.77	0.71

◀ *HL-LHC YR projections*

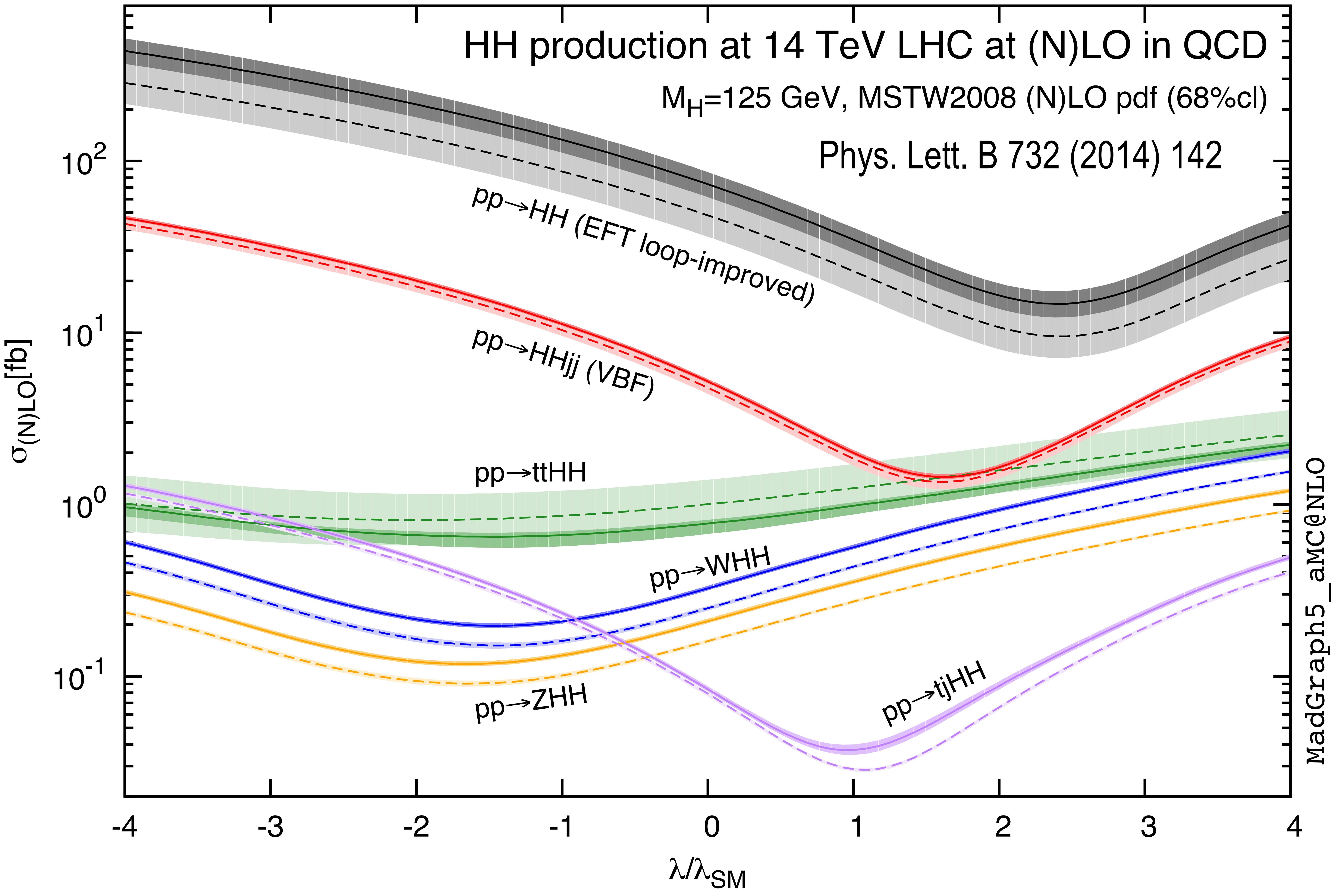
The CMS full Run 2 result improved by 5× over previous 2016 CMS results

Analyses improve much faster than the luminosity!

Excellent prospects for Run 3 and HL-LHC

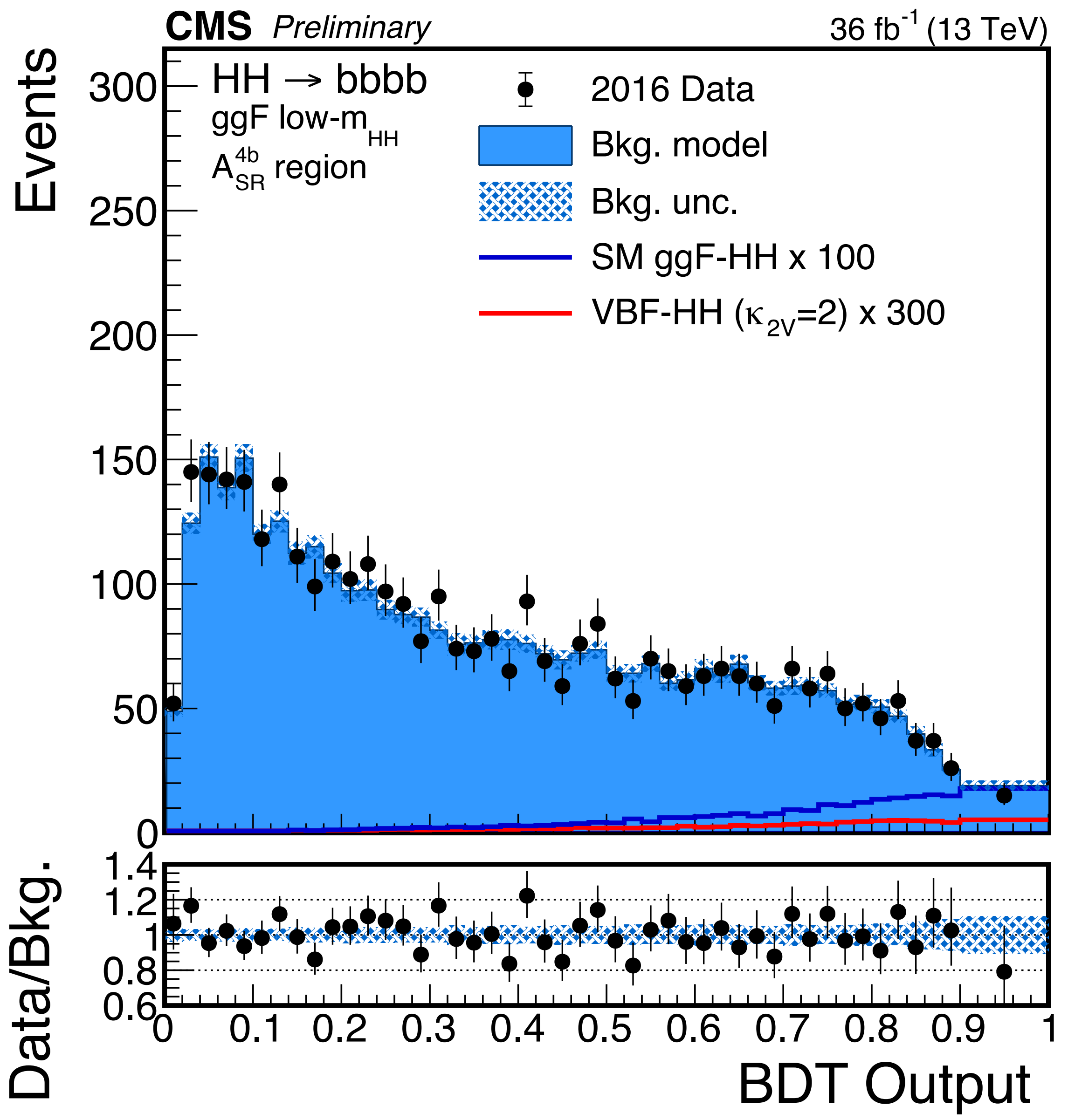
Additional material

HH cross-section dependence on κλ

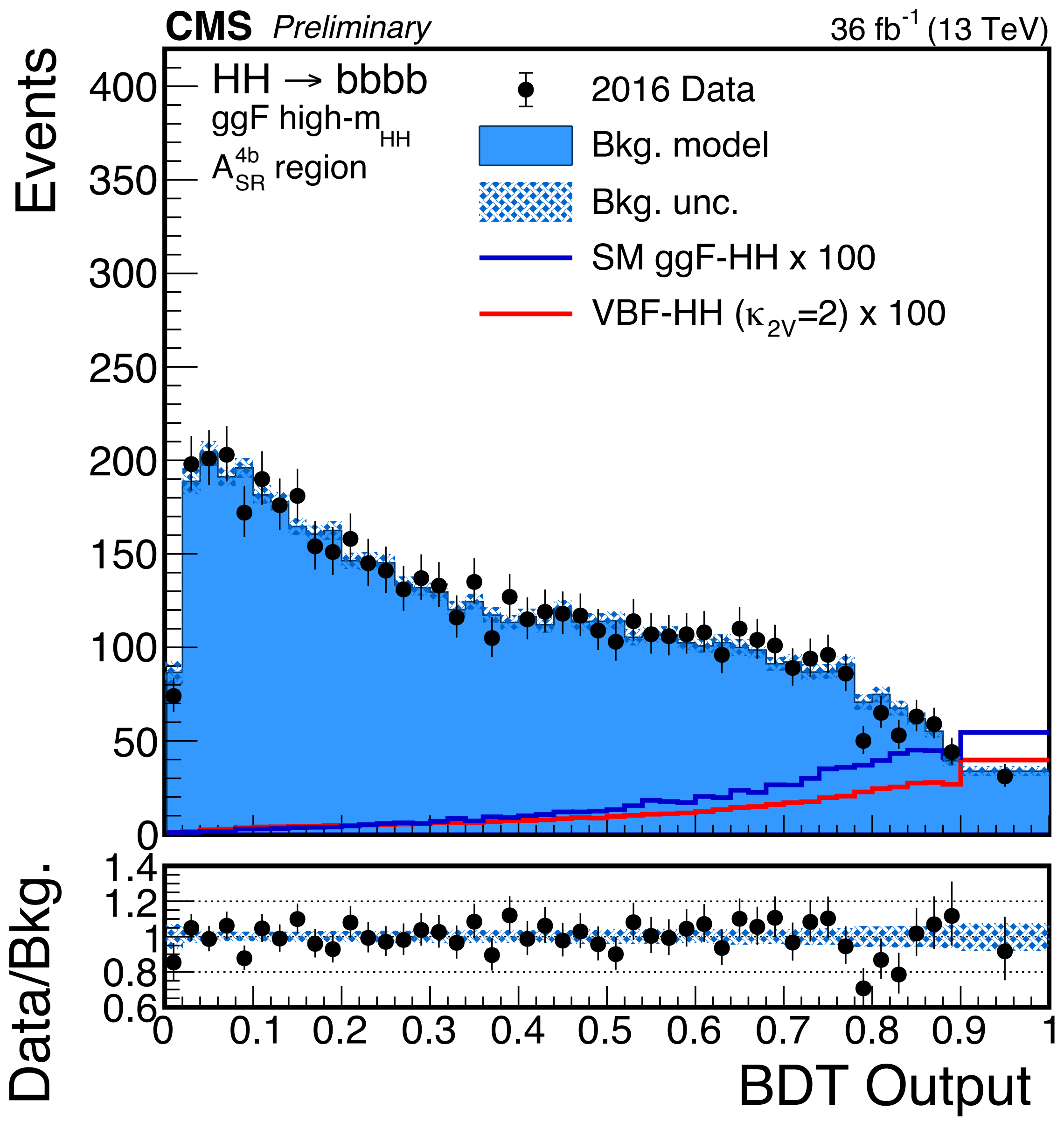


- Quadratic dependence of the xs on κλ
- Interference effects → minimum xs for κλ ≠ 1
  - minimum at κλ = 2.45 for ggF

2016 data

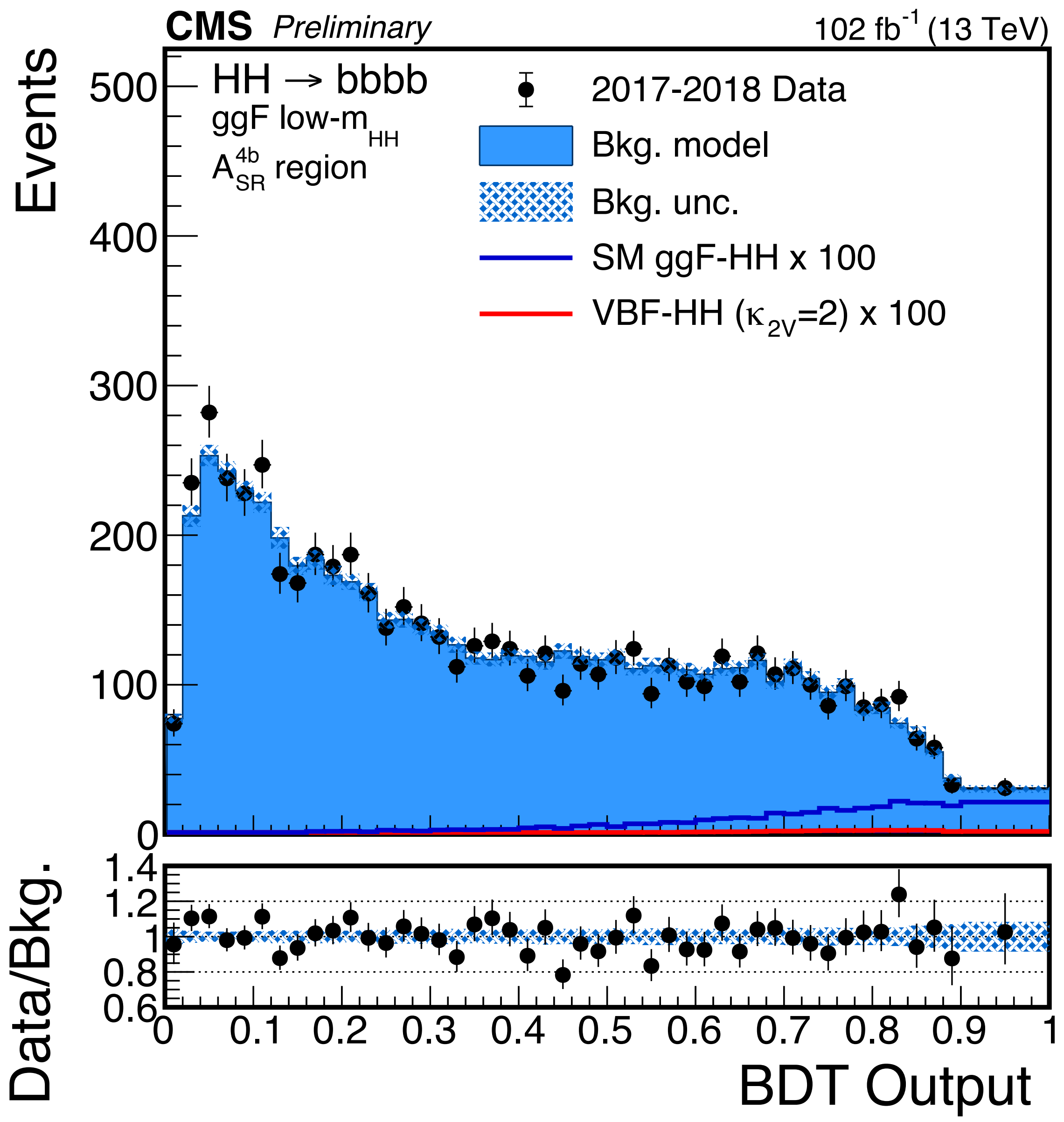


Low m_{HH}

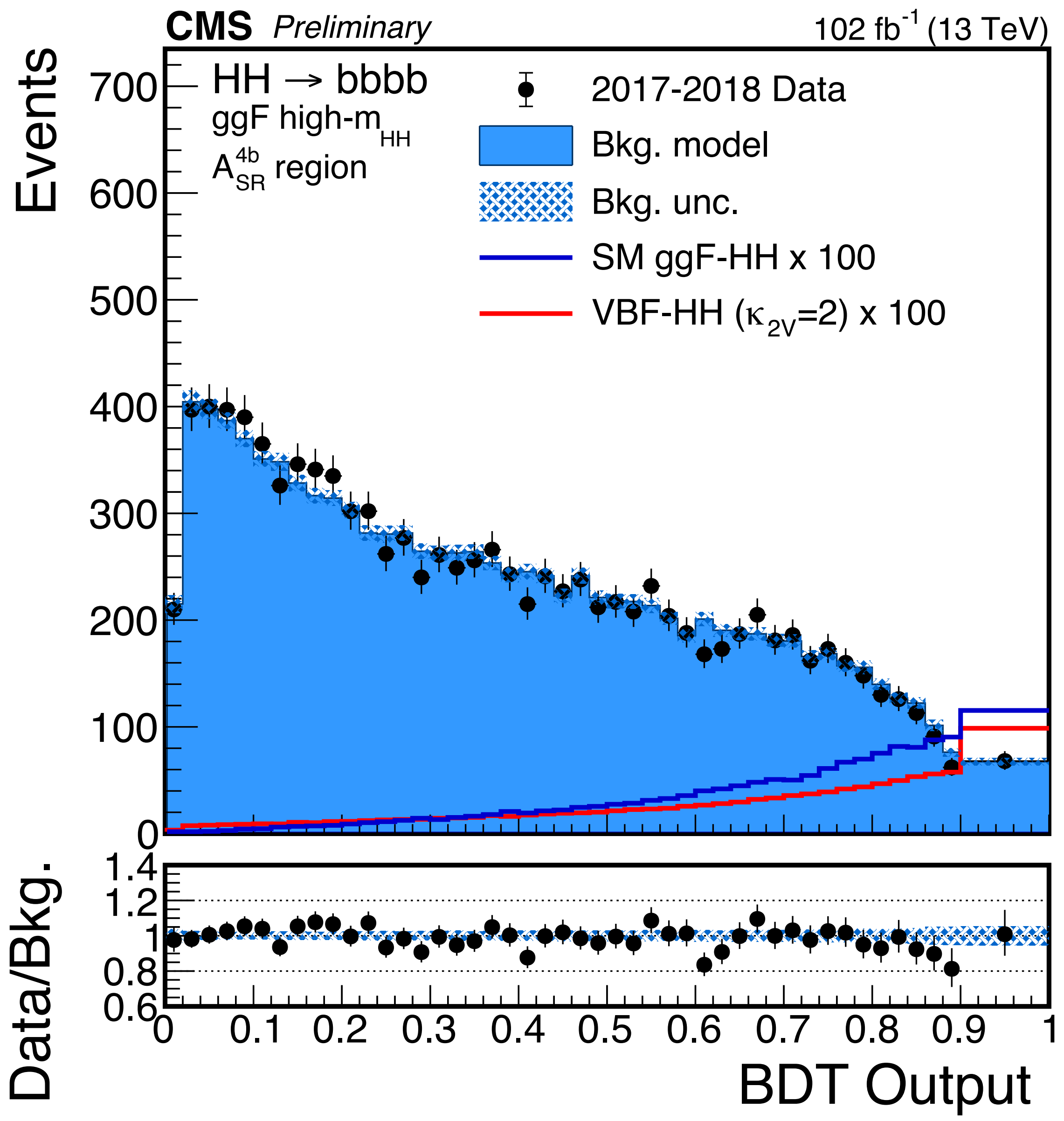


High m_{HH}

2017 + 2018 data



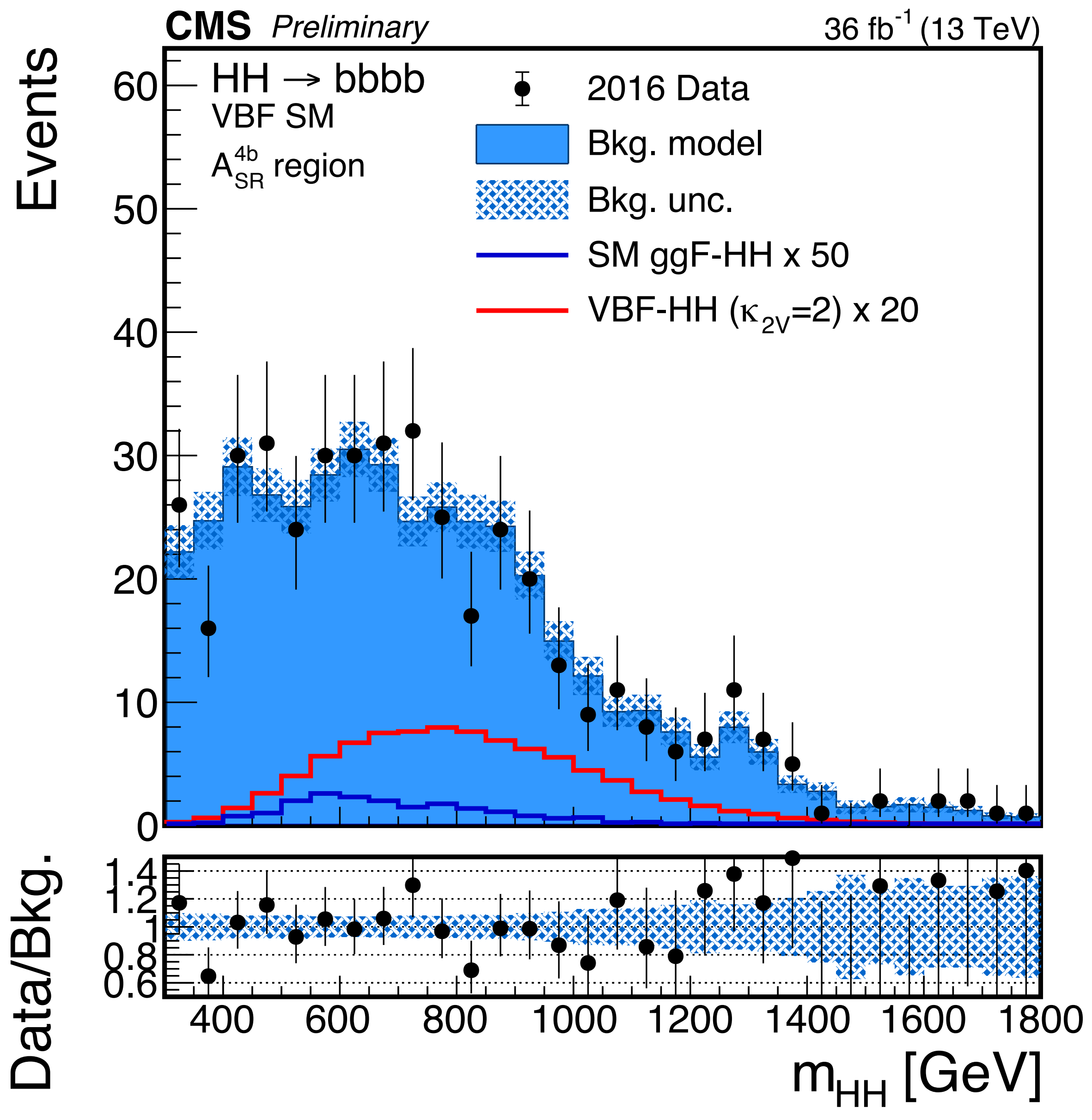
Low m_{HH}



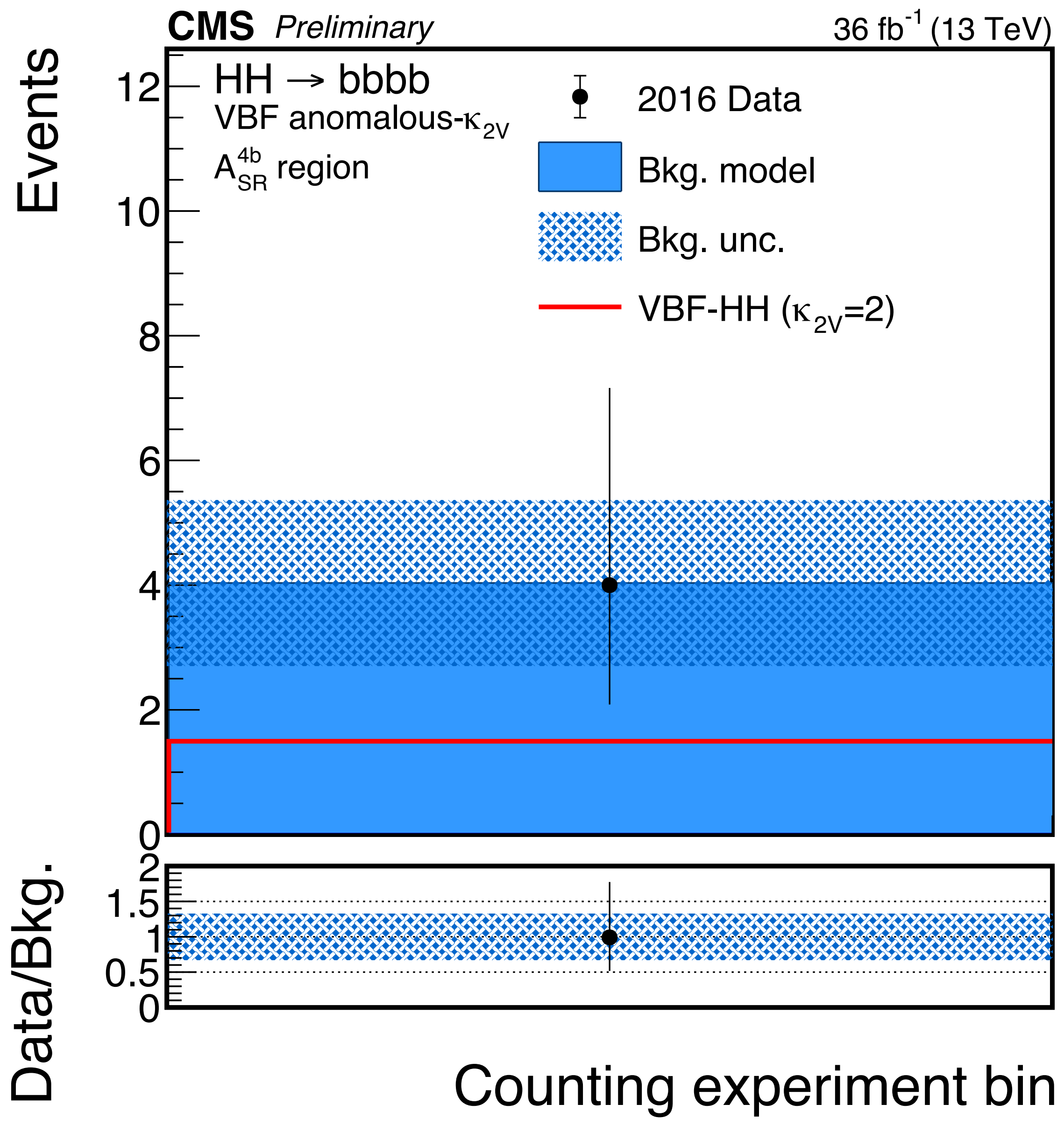
High m_{HH}

VBF discriminants

2016 data

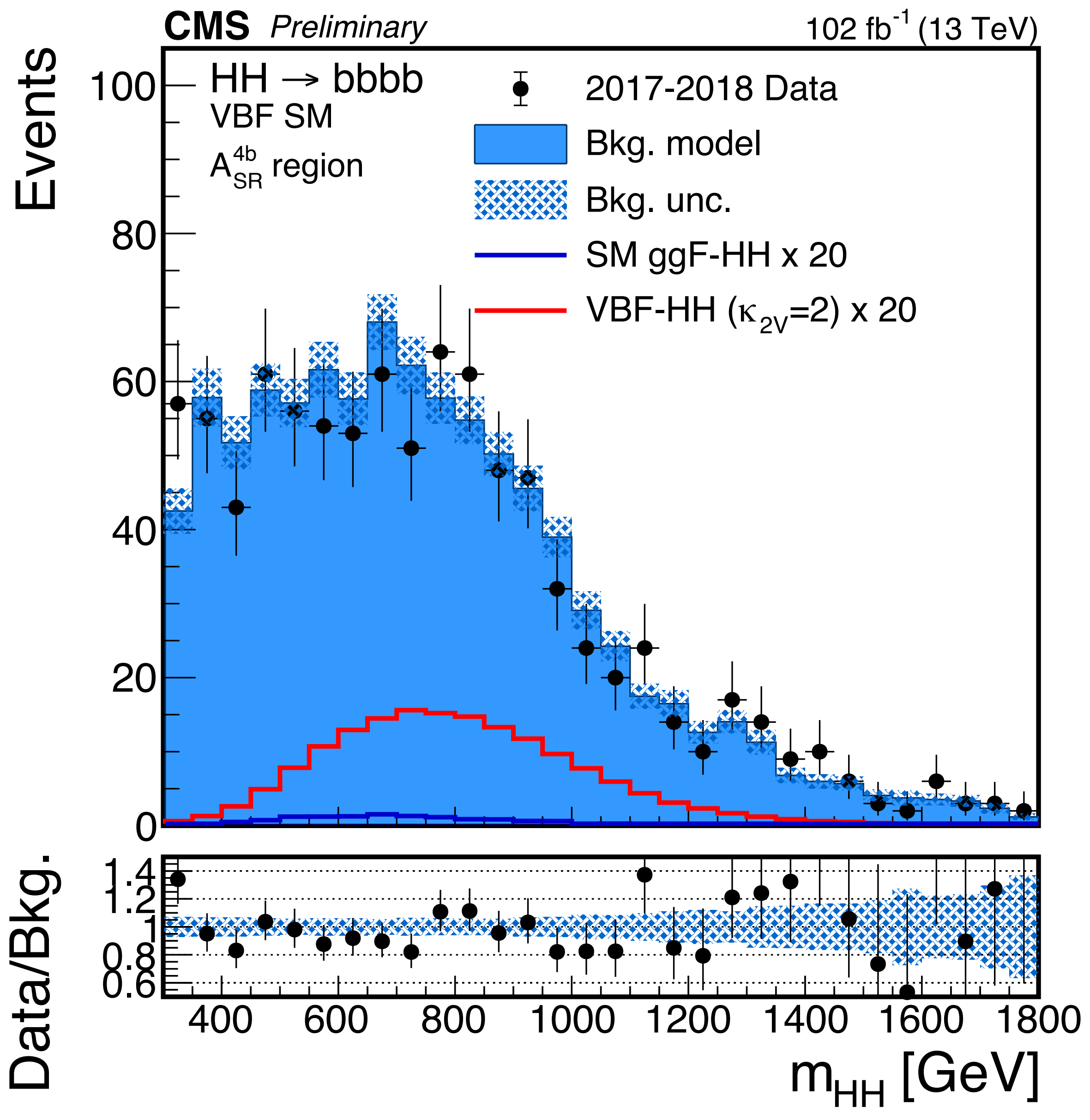


SM-like cat.

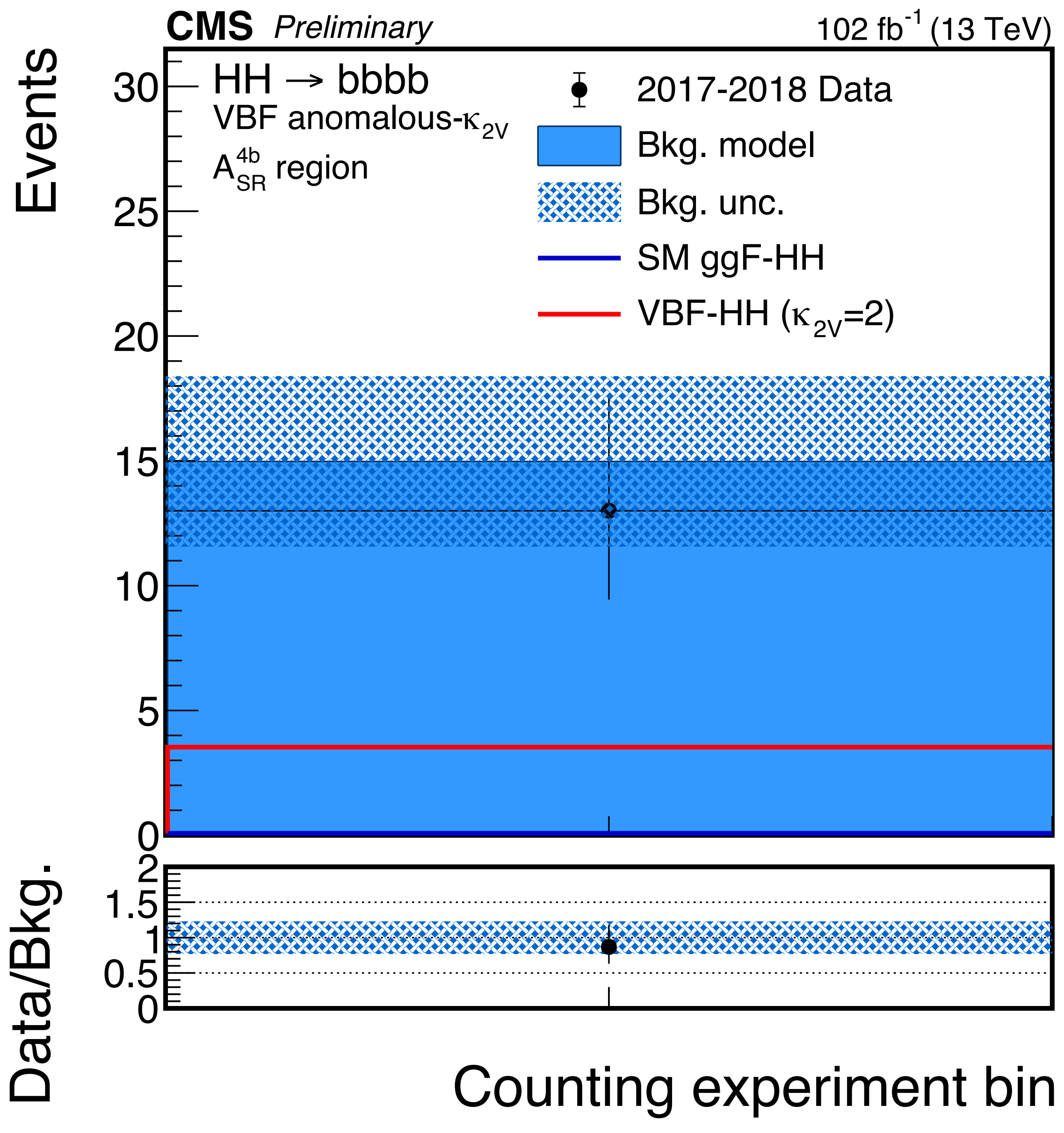


BSM-like cat.

2017 + 2018 data

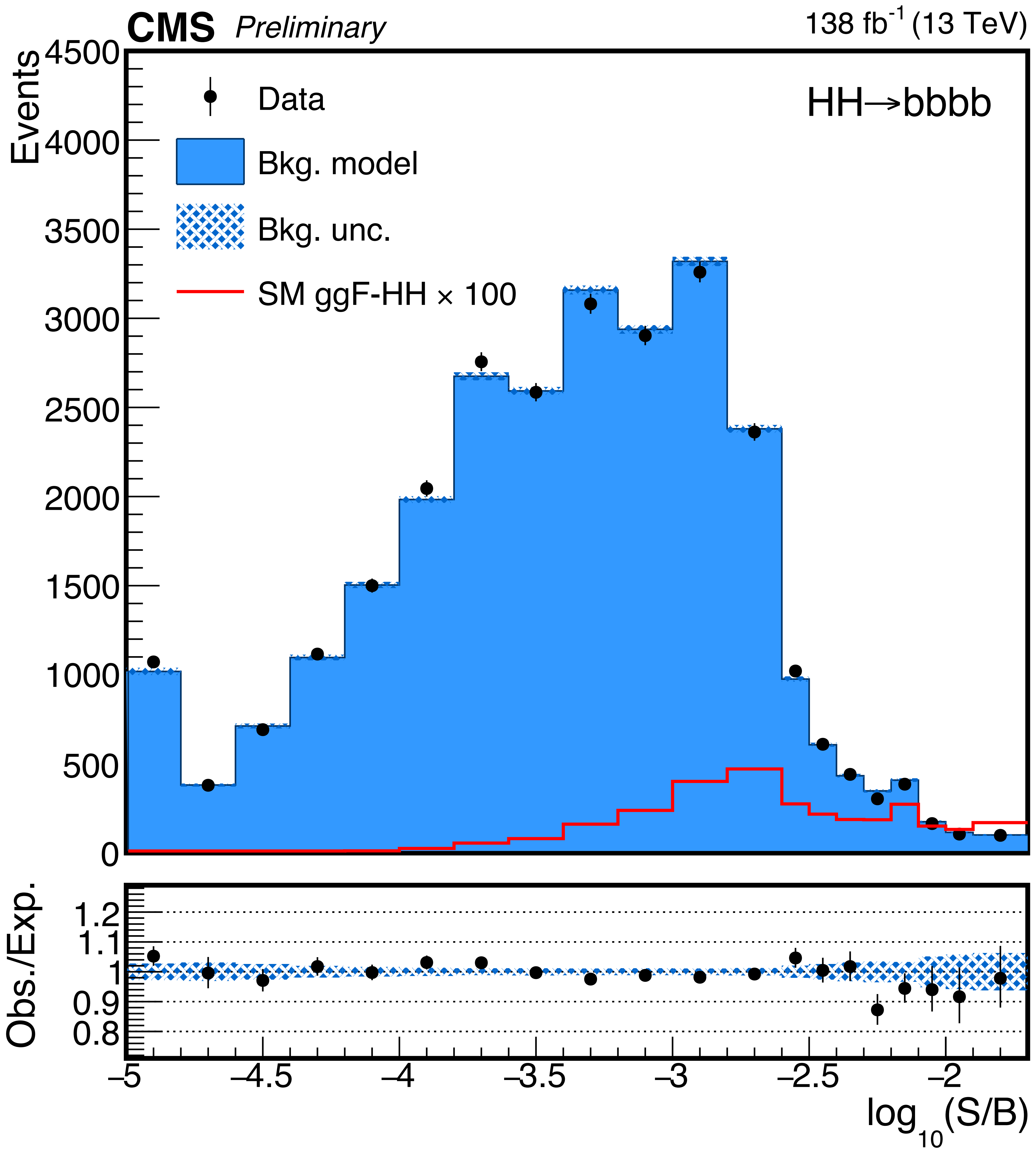


SM-like cat.



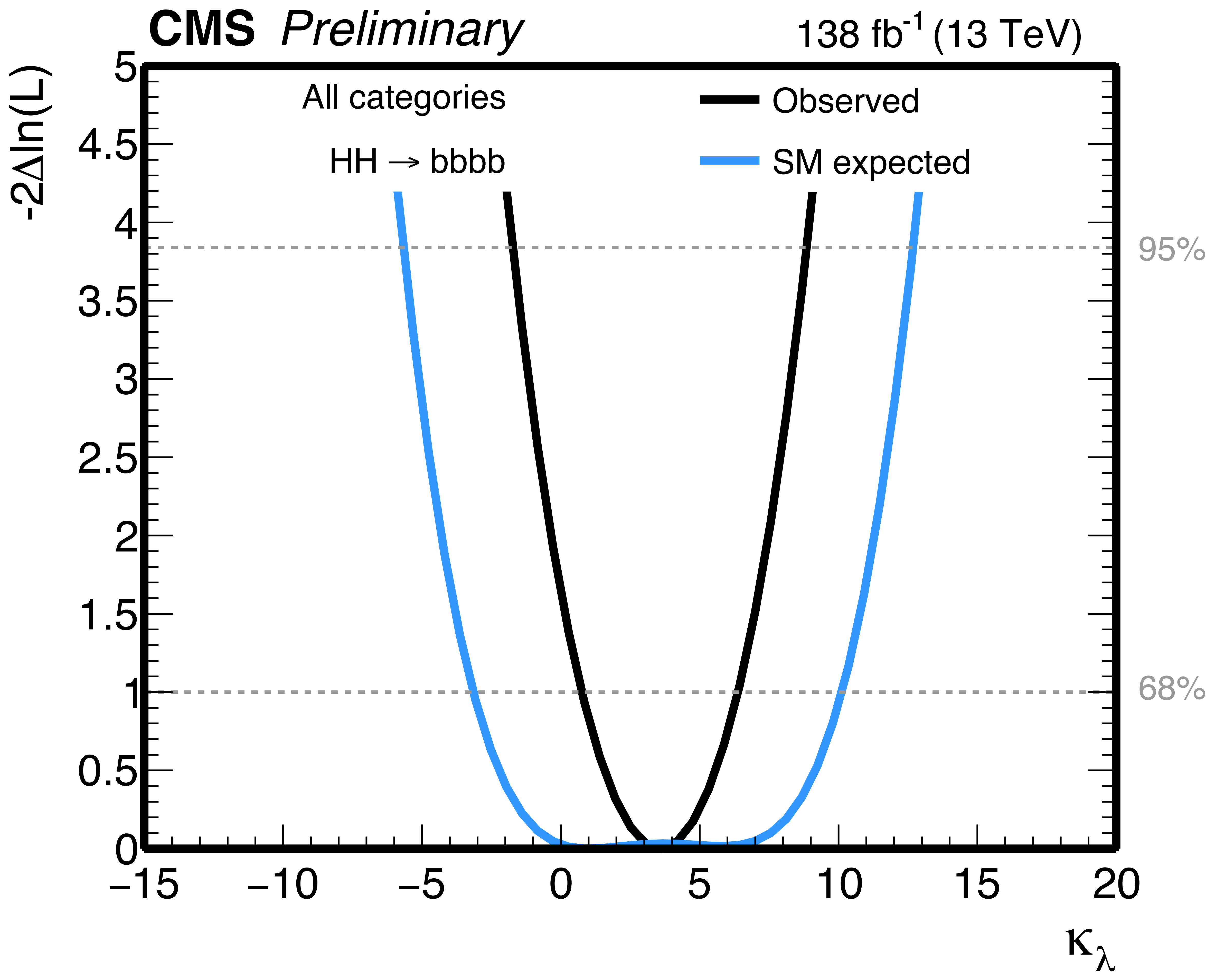
BSM-like cat.

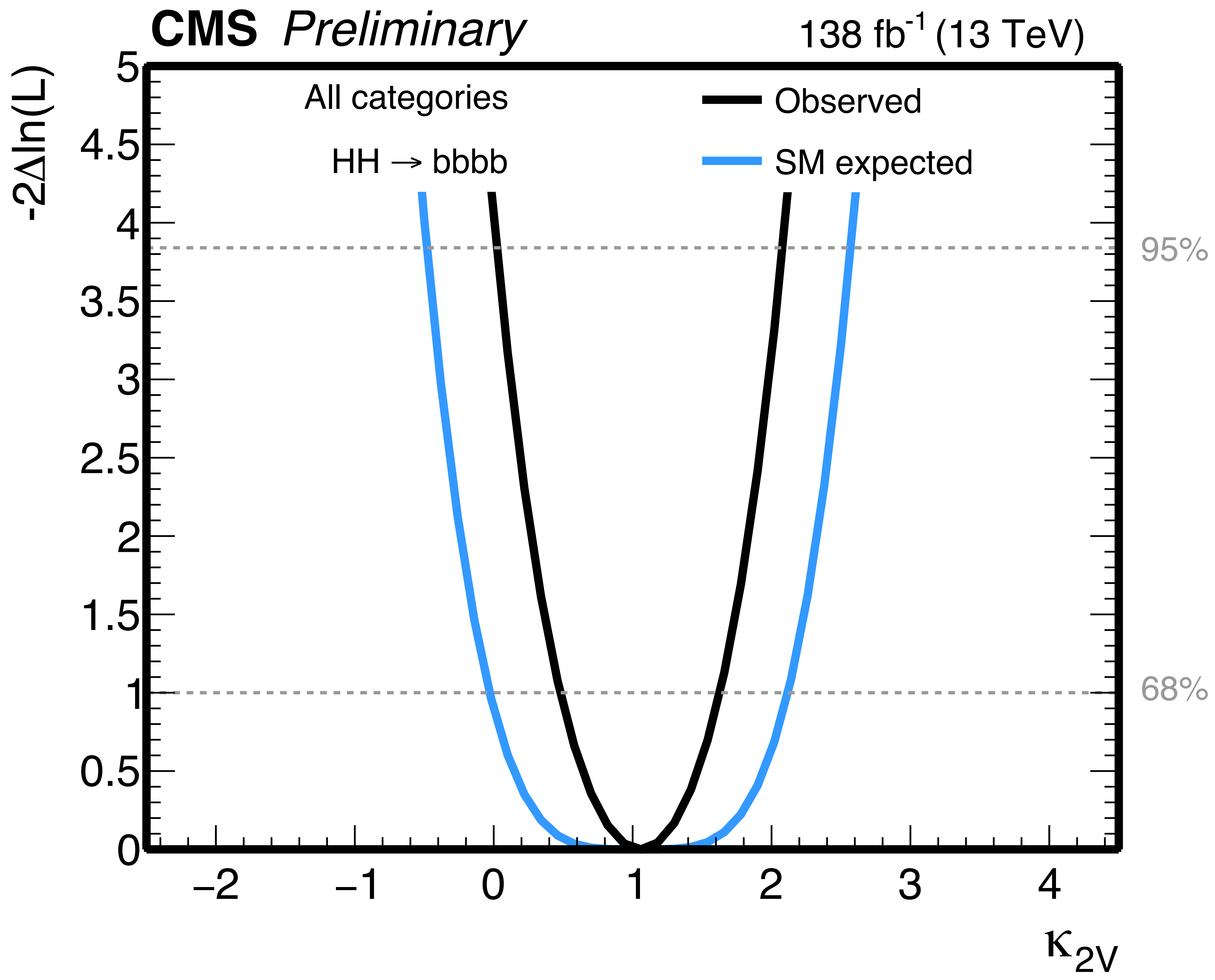
Grouping all the categories



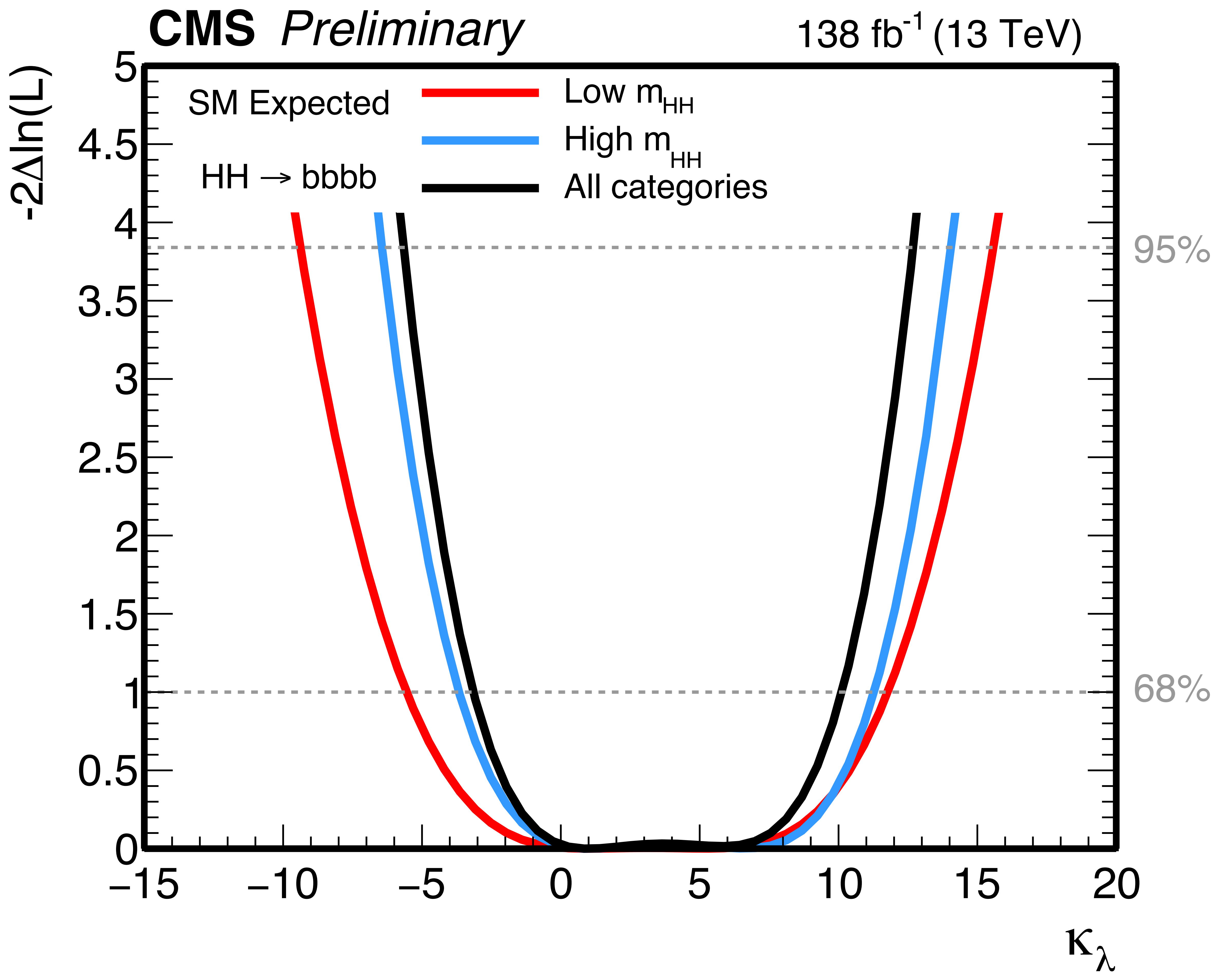
- Events yields aggregated from all categories, sorted by ascending $\log_{10}(S/B)$
  - considering a SM HH signal
- the underfluctuation at high S/B, directly stemming from the deficit in the ggF high m_{HH} category, is clearly visible

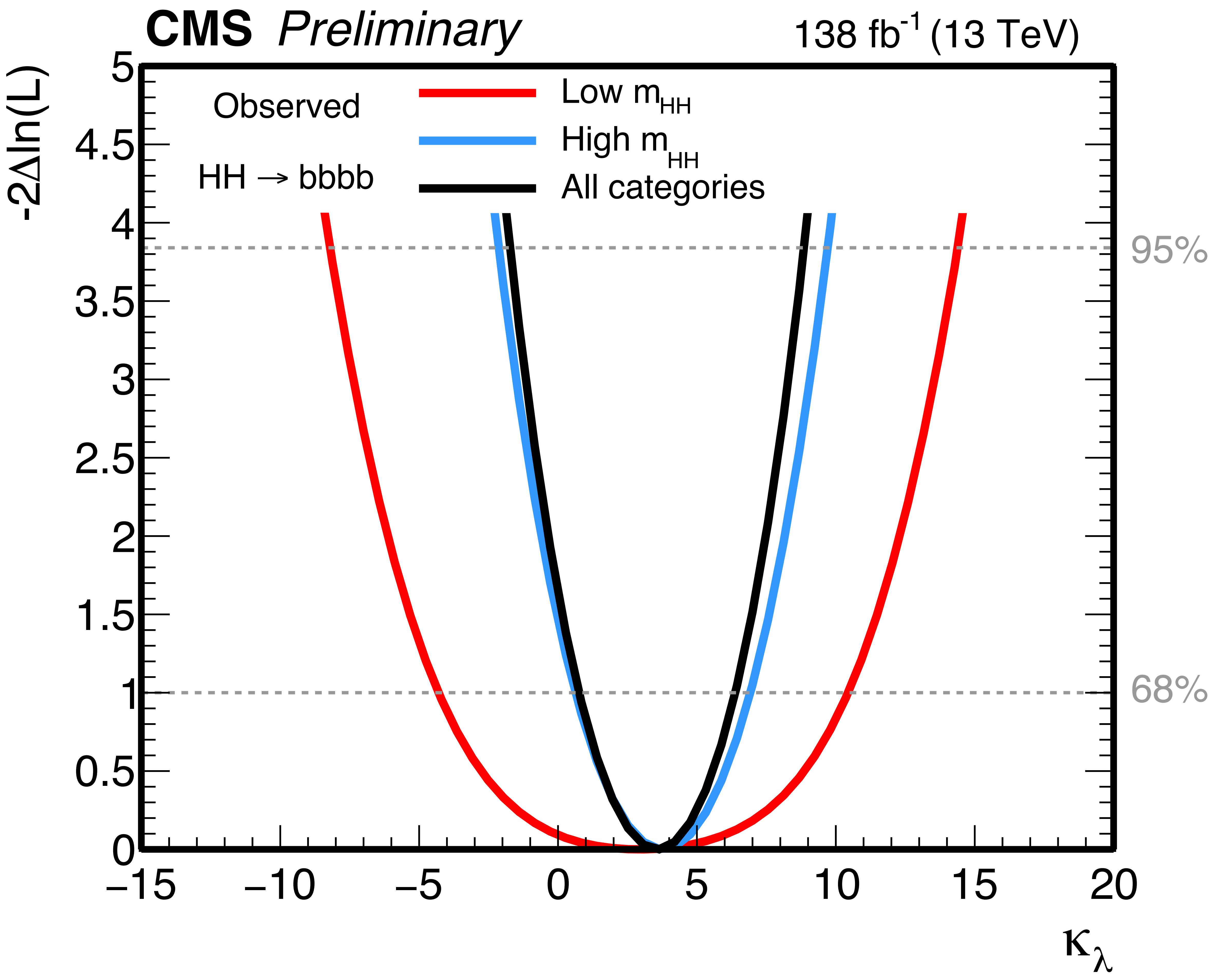
Likelihood scans

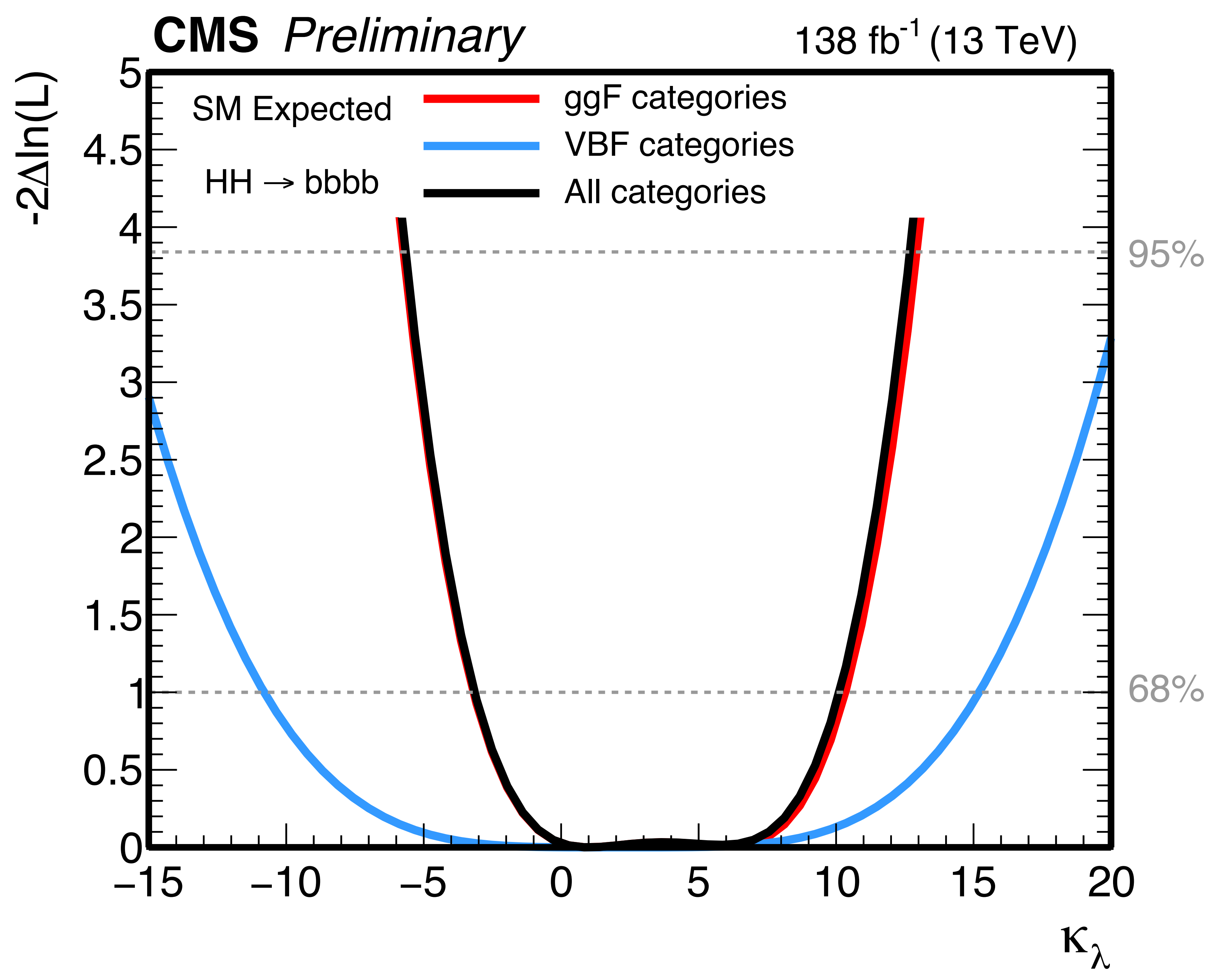


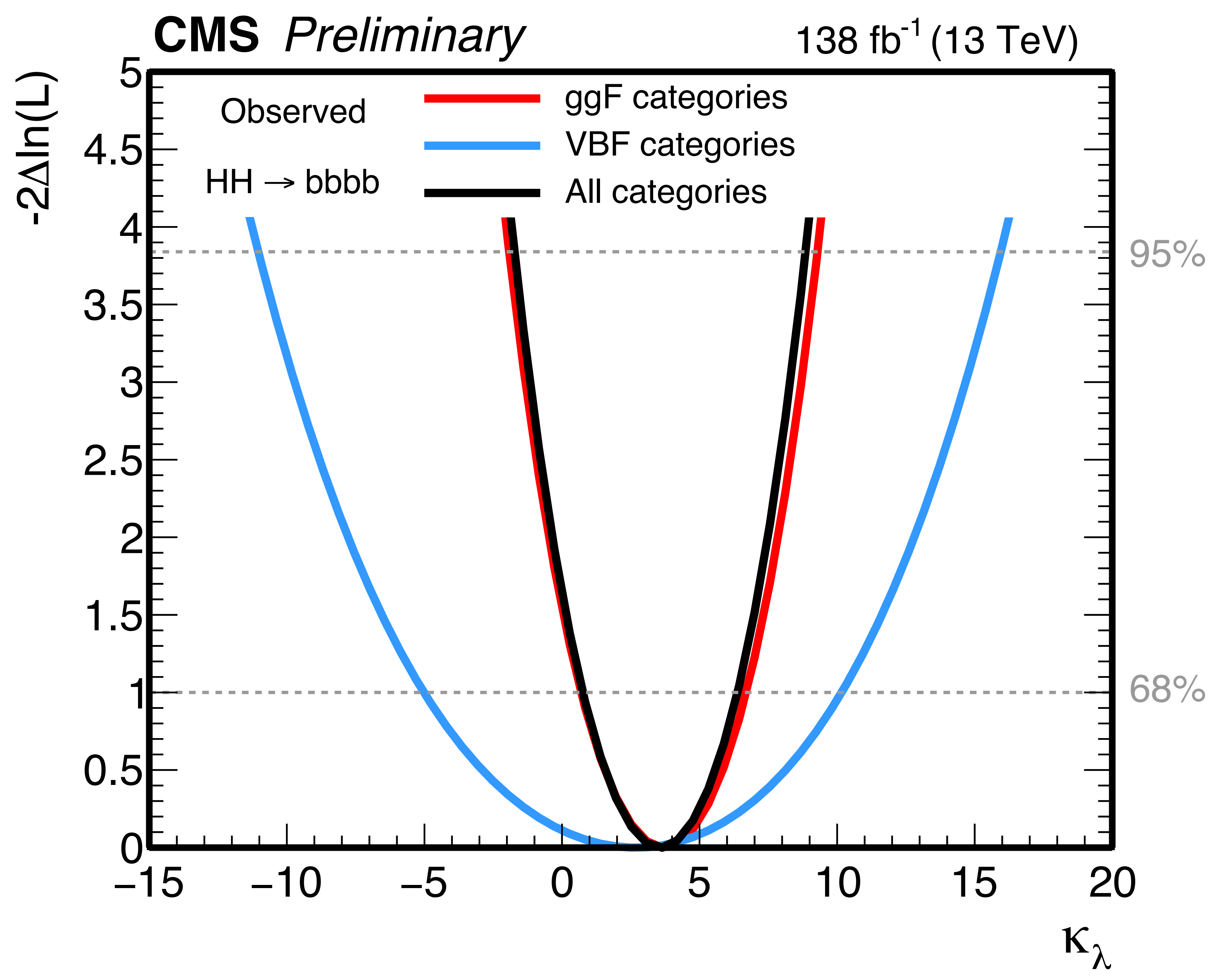


Mass categorization



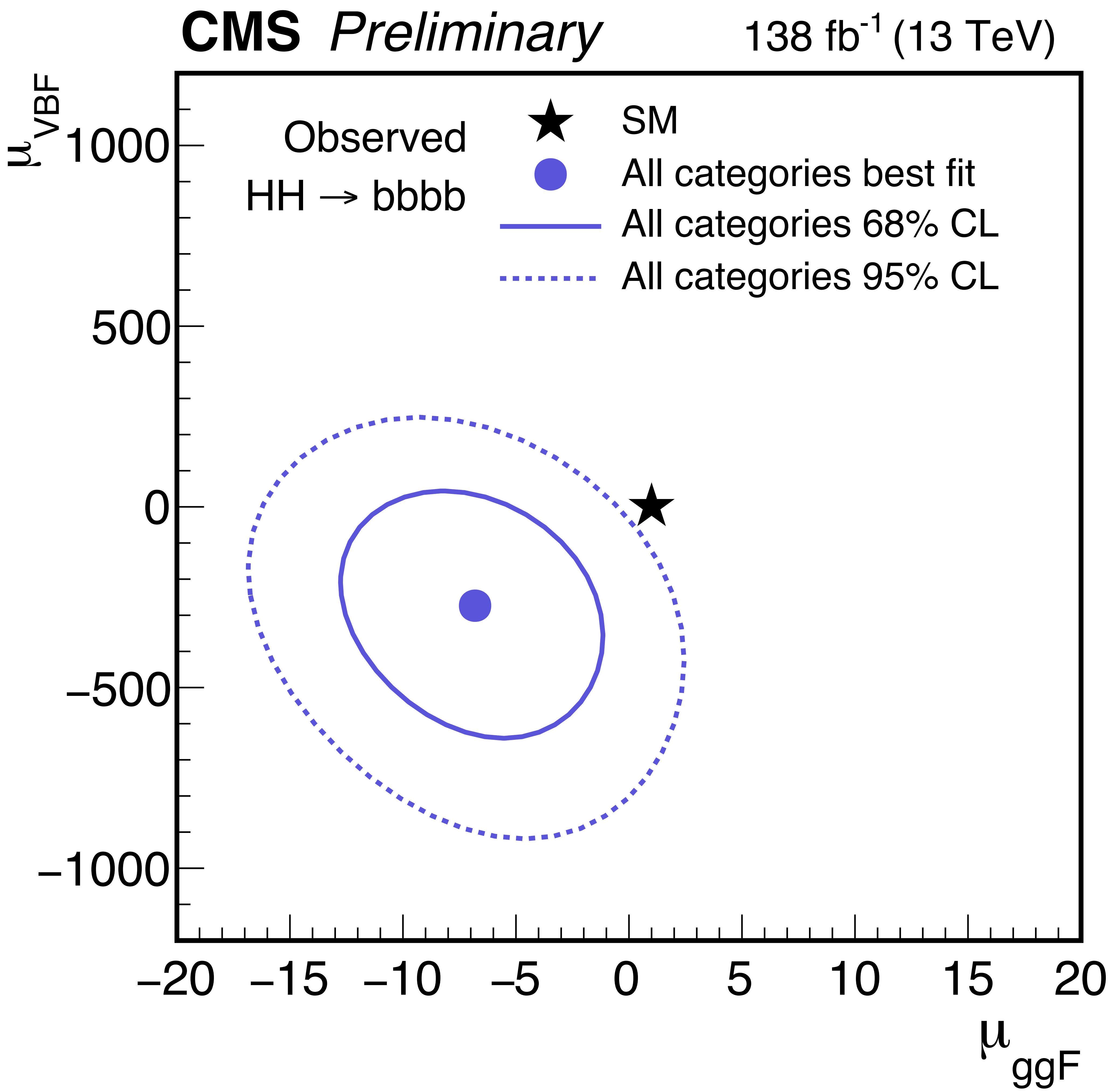






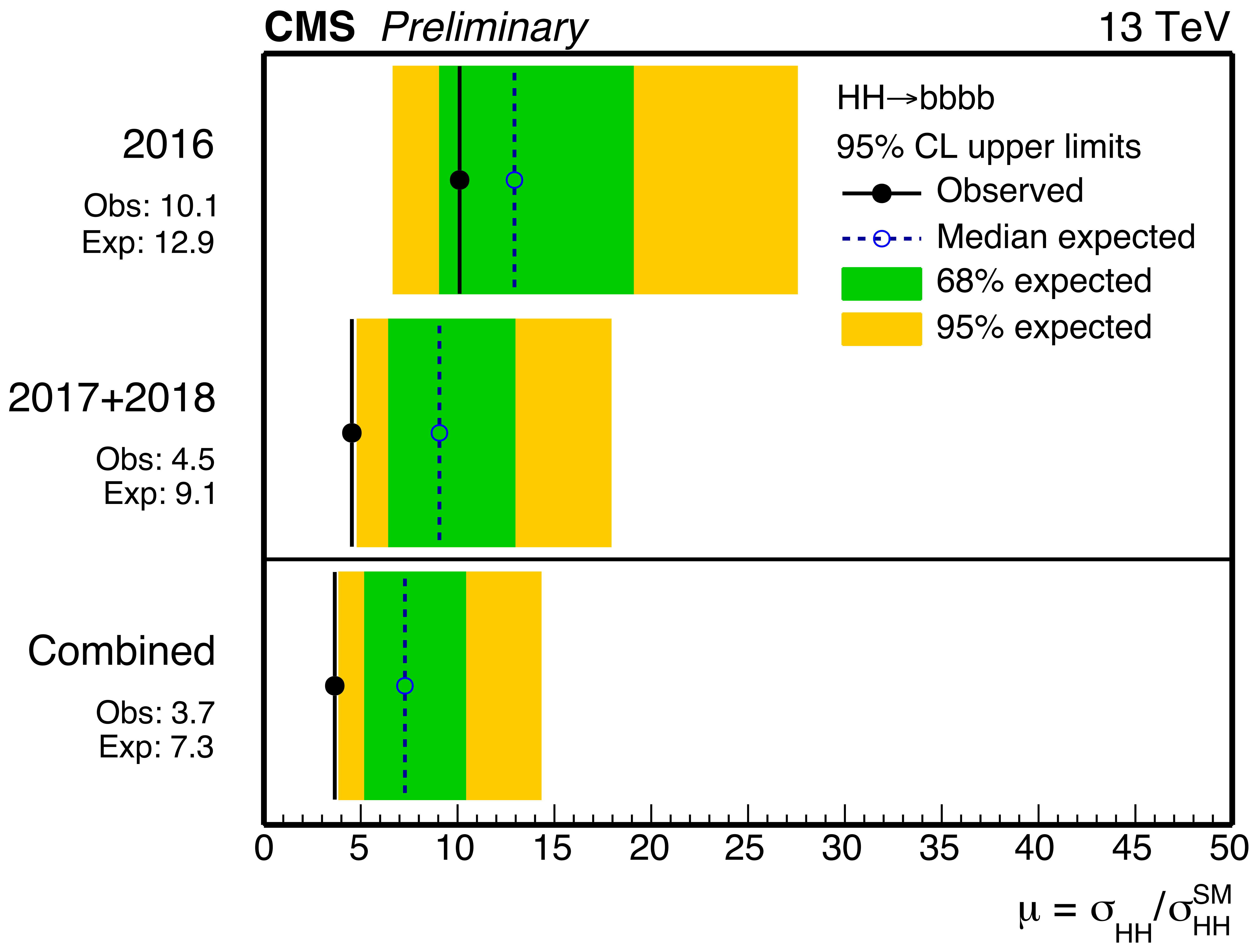
- The self-coupling determination is entirely driven by the ggF categories

Fitting the ggF and VBF strengths



- Separate fit of the ggF and VBF signal strengths
  - in both cases the SM couplings are assumed (for the expected signal shapes and acceptances)

Limits by data taking year



- Sensitivity lead by the 2017+2018 data (×2.8 more data)
- TIghter trigger thresholds partially reduce the sensitivity of the 2017+2018 analysis compared to the simple sqrt(L) scaling of the 2016 result