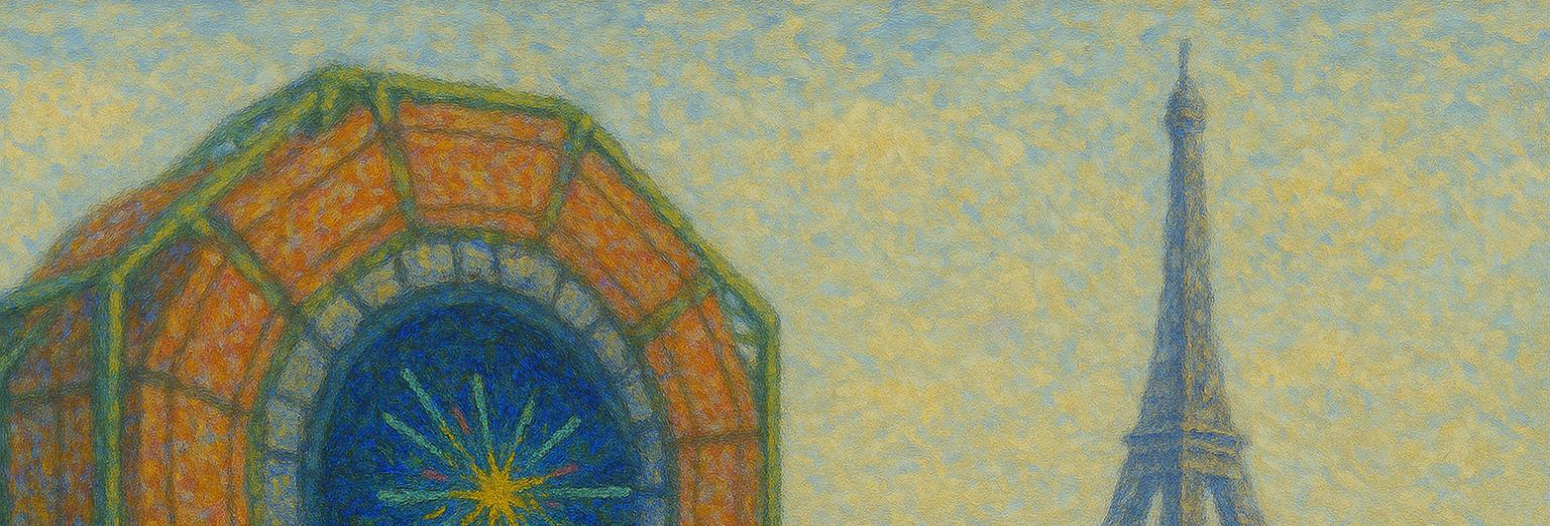
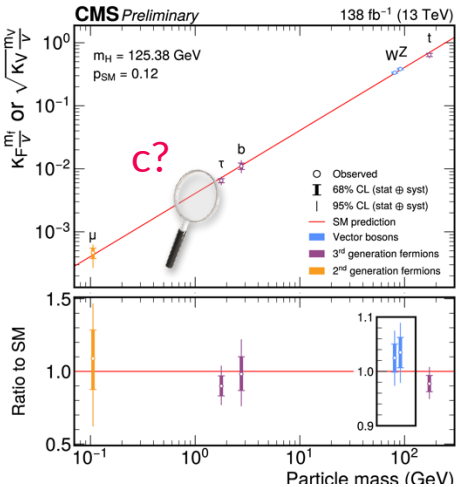


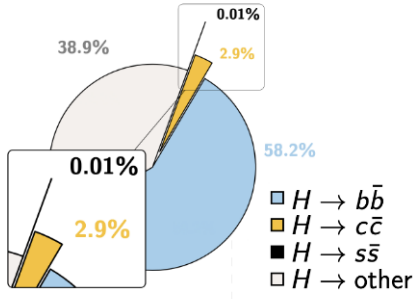
Search for $H \to c\bar{c}$ at the CMS experiment

Higgs Hunting 2025
15-17 July 2025, Orsay Paris

Angela Zaza on behalf of the CMS collaboration

Motivation of the search





SM $m_H = 125$ GeV $H \to q\bar{q}$ branching fractions, $H \to u\bar{u}/d\bar{d}$ too small to show!

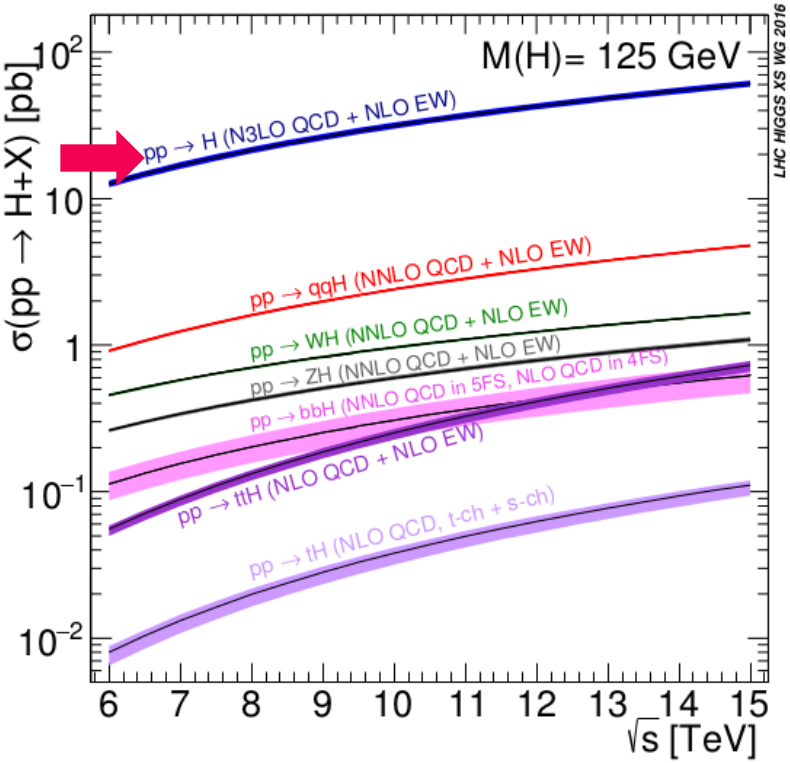
- Higgs boson coupling with gauge bosons and third generation fermions measured with precision of ~10% (Walaa's talk)
- First evidence for $H \to \mu\mu$ at 3σ with Run-2 data
- Higgs coupling to **second generation quarks** still out of reach
  - one of the highest priority goals of CMS and ATLAS physics program
  - Discrepancies from SM prediction could provide valuable insights into new physics

Challenges of the search

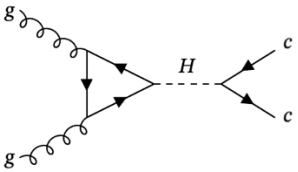
- Small branching ratio (2.9%)
- Large QCD multijet background
- Difficulty to identify jets originated by charm quarks

Great enhancement in sensitivity thanks to new developments in **heavy flavour tagging**

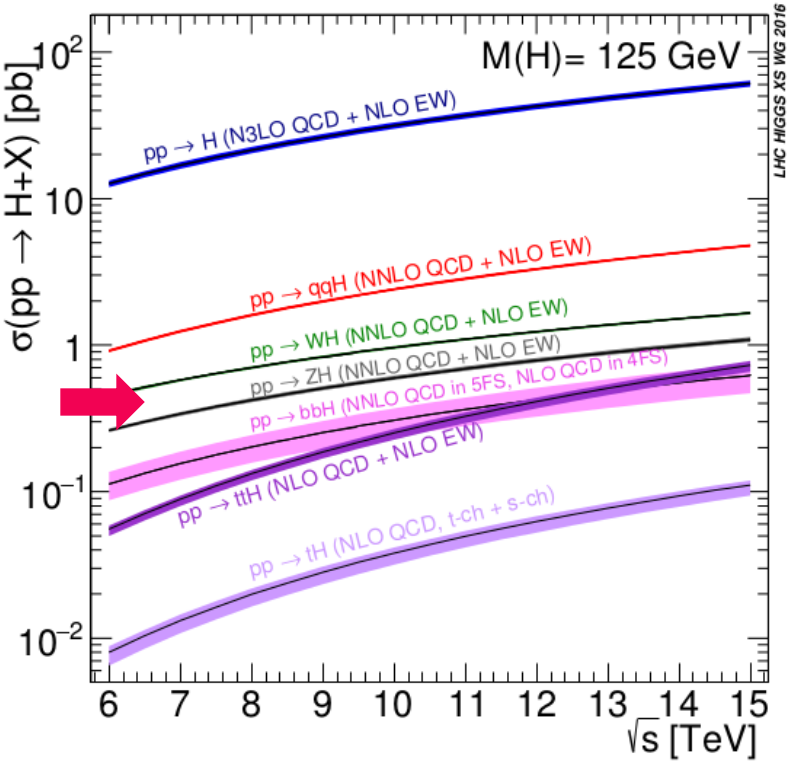
Searches performed by CMS on Run2 data (138 fb^{-1})



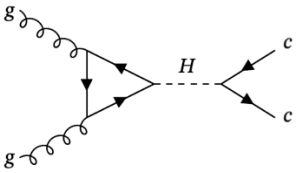
→ gg H →cc search (Phys. Rev. Lett. **131**, 041801)
- 87% of the total cross section
- Higgs p_T > 450 GeV (boosted analysis)



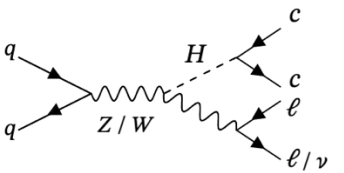
Searches performed by CMS on Run2 data (138 fb^{-1})



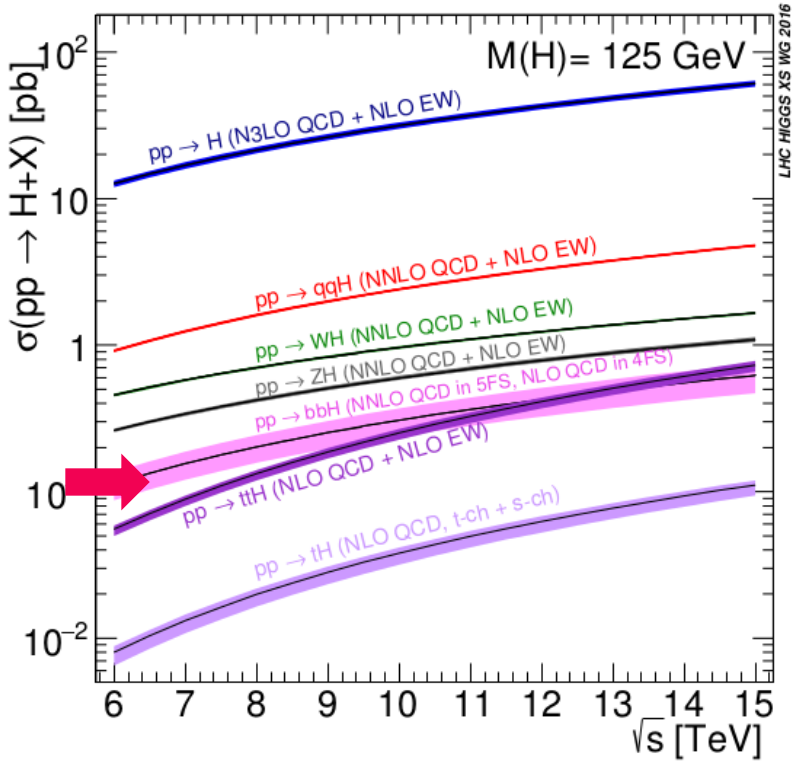
→ gg H →cc search (Phys. Rev. Lett. **131**, 041801)
- 87% of the total cross section
- Higgs p_T > 450 GeV (boosted analysis)



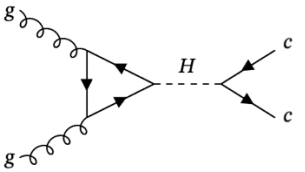
→ VH H→cc search (Phys. Rev. Lett. **131**, 061801)
- 4% of the total cross section
- resolved and boosted analyses



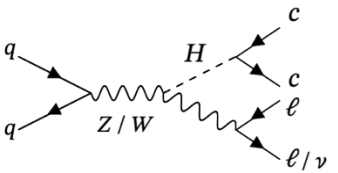
Searches performed by CMS on Run2 data (138 fb^{-1})



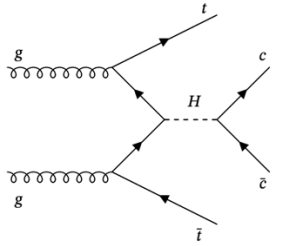
→ gg H →cc search (Phys. Rev. Lett. **131**, 041801)
- 87% of the total cross section
- Higgs p_T > 450 GeV (boosted analysis)



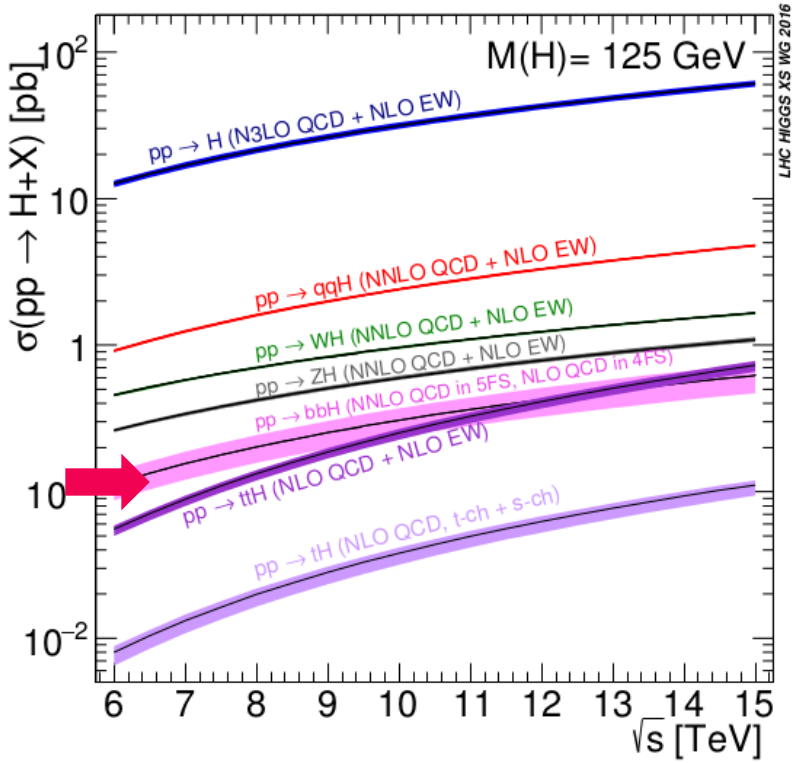
→ VH H→cc search (Phys. Rev. Lett. **131**, 061801)
- 4% of the total cross section
- resolved and boosted analyses



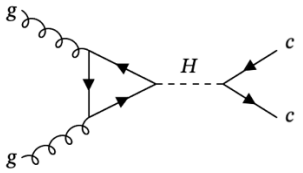
→ ttH (HIG-PAS-24-018)
- 1% of the total cross section
- simultaneous extraction of Hcc, Hbb, Zcc and Zbb



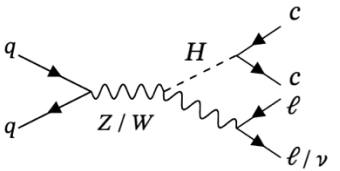
Searches performed by CMS on Run2 data (138 fb^{-1})



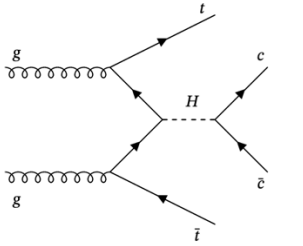
→ gg H →cc search (Phys. Rev. Lett. **131**, 041801)
- 87% of the total cross section
- Higgs p_T > 450 GeV (boosted analysis)



→ VH H→cc search (Phys. Rev. Lett. **131**, 061801)
- 4% of the total cross section
- resolved and boosted analyses



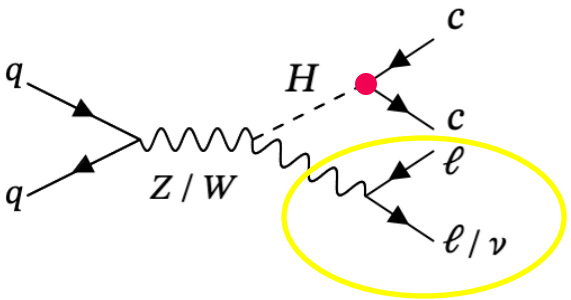
→ ttH (HIG-PAS-24-018)
- 1% of the total cross section
- simultaneous extraction of Hcc, Hbb, Zcc and Zbb



Most sensitive channels
Discussed in this talk

Search for VH→cc

- **Most sensitive** channel to the H→cc decay:
  QCD background suppressed by targeting the leptonic decays of the Z/W boson

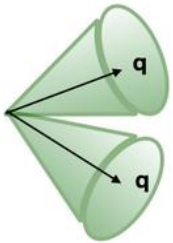


- Three categories based on the Z/W decay modes:
  - **0L**: $Z \rightarrow \nu\nu$
  - **1L**: $W \rightarrow l\nu$
  - **2L**: $Z \rightarrow l^{+}l^{-}$

- Backgrounds: V+jets, $t\bar{t}$ (extracted from dedicated CRs)
  QCD (suppressed), single-top, VH→bb (estimated from MC)

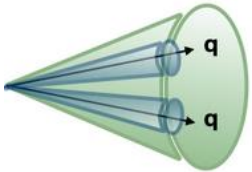
Resolved analysis (p_T(H) < 300 GeV)
Higgs boson reconstructed from
2 c-tagged small radius jets (AK4)

~95% of events



Boosted analysis (p_T(H) > 300 GeV)
Higgs boson reconstructed as
a single large radius jet (AK15)

~5% of events



https://cds.cern.ch/record/2809446

Search for VH→cc

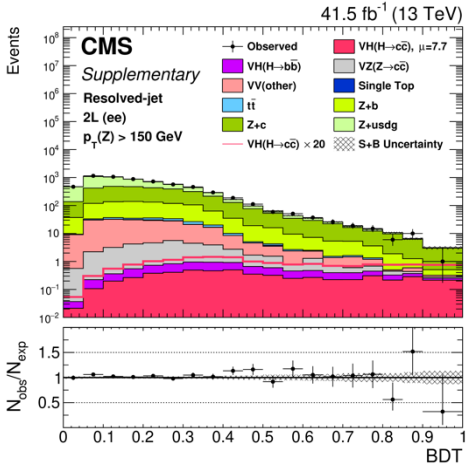
Resolved analysis

- Higgs candidate reconstructed from two AK4 c-tagged jets (**DeepJet**)
- BDT for signal vs background discrimination
- Signal strength modifier μ extracted from a maximum likelihood fit to data of the **BDT score**

$$\mu = \frac{(\sigma B)_{obs}}{(\sigma B)_{SM}}$$

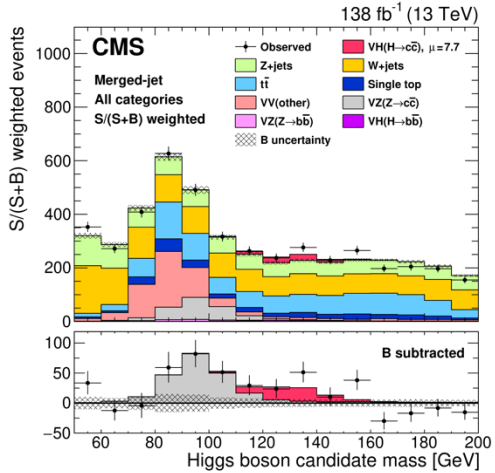
Post-fit distribution of the BDT score in the 2L(ee) category in 2017 data

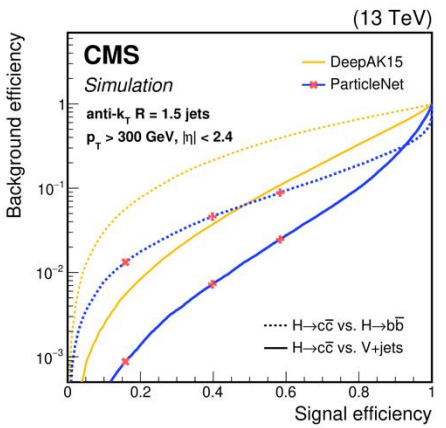
CMS-results HIG-21-008



Merged analysis

- Higgs candidate reconstructed from the p_T-dominant AK15 jets
- **ParticleNet** X→cc tagger
- BDT for signal vs background discrimination
- μ extracted from a fit to **AK15 jet regressed mass**





Combined m(H_{cand}) distribution in all channels of the merged-jet analysis

Search for VH→cc - Results

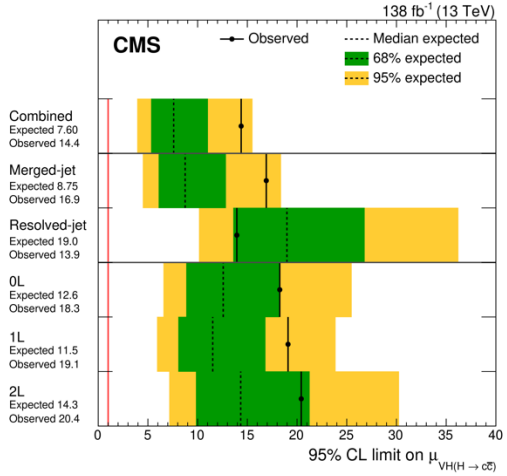
- Simultaneous fit of the two analyses
- Upper limit on the signal strength $\mu_{VH(cc)}$:

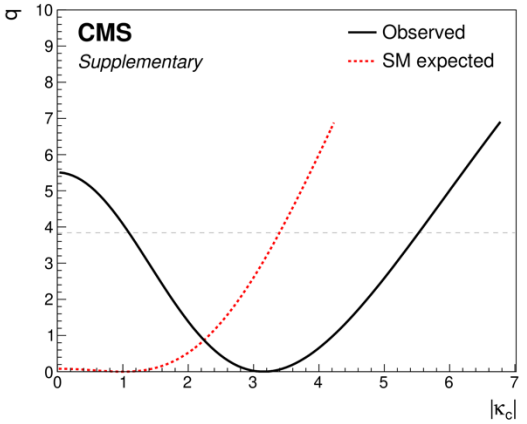
$$\frac{\sigma(VH)\cdot B(H\to c\bar{c})}{\sigma(VH)_{SM}\cdot B(H\to c\bar{c})_{SM}} < 14\ (7.6)\ \ @95\%\ CL$$

- Constraints on the Higgs-charm Yukawa coupling modifier k_c:

$$1.1 < |k_c| < 5.5\ (|k_c| < 3.4)\ \ @95\%CL$$

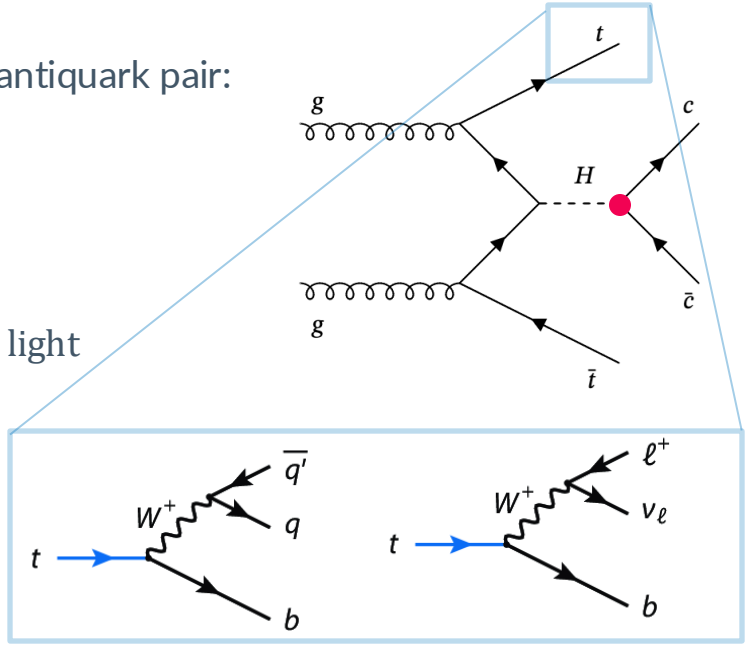
- Validation on **Z→cc**:
  first time observed at a hadron collider with a significance of $\mathbf{5.7\sigma}$





Search for ttH→cc - Overview

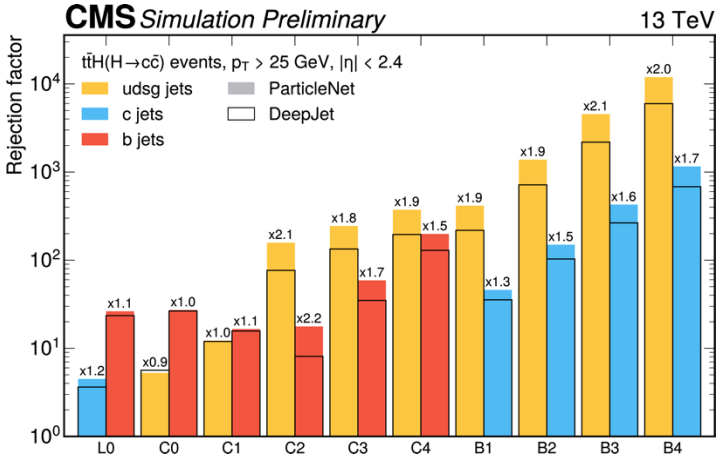
- Three categories based on the decay mode of the top quark-antiquark pair:
  - **0L**: fully hadronic
  - **1L**: semileptonic
  - **2L**: dileptonic
- Backgrounds:
  - **$t\bar{t}$+jets** categorized in $t\bar{t}+\geq 2b$, $t\bar{t}+b$, $t\bar{t}+\geq 2c$, $t\bar{t}+c$, $t\bar{t}+$ light
    data-driven corrections estimated in CRs
  - **$t\bar{t}H(bb)$, $t\bar{t}Z(bb/cc)$** measured simultaneously to signal
  - **minor backgrounds:** single top, $t\bar{t}W$, tWZ, QCD
- Heavy flavour tagging → **ParticleNet**
  defined 11 mutually exclusive tagging categories
- ttH signal vs tt+jets background discrimination with a **multiclass event classifier** based on **ParT**
  4 SRs and 5 CRs defined based on the classifier output

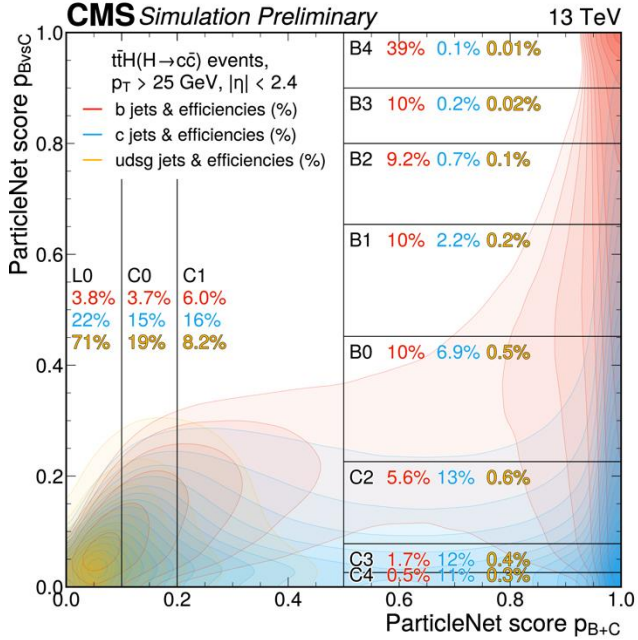


- Signals extracted via a binned likelihood fit to data of the ParT event classifier discriminant

ttH→cc – Heavy flavour tagging

- ParticleNet tagger for AK4 jets used in Run2 analysis for the first time
- 2 scores used to categorize events:
  p_{B+C} → discriminate b and c jets from light jets
  p_{BvsC} → discriminate between b and c jets
- 11 exclusive categories defined
  → 5 b-jet (B0-B4), 5 c-jet (C0-C4) and 1 light-jet (L0) enriched categories

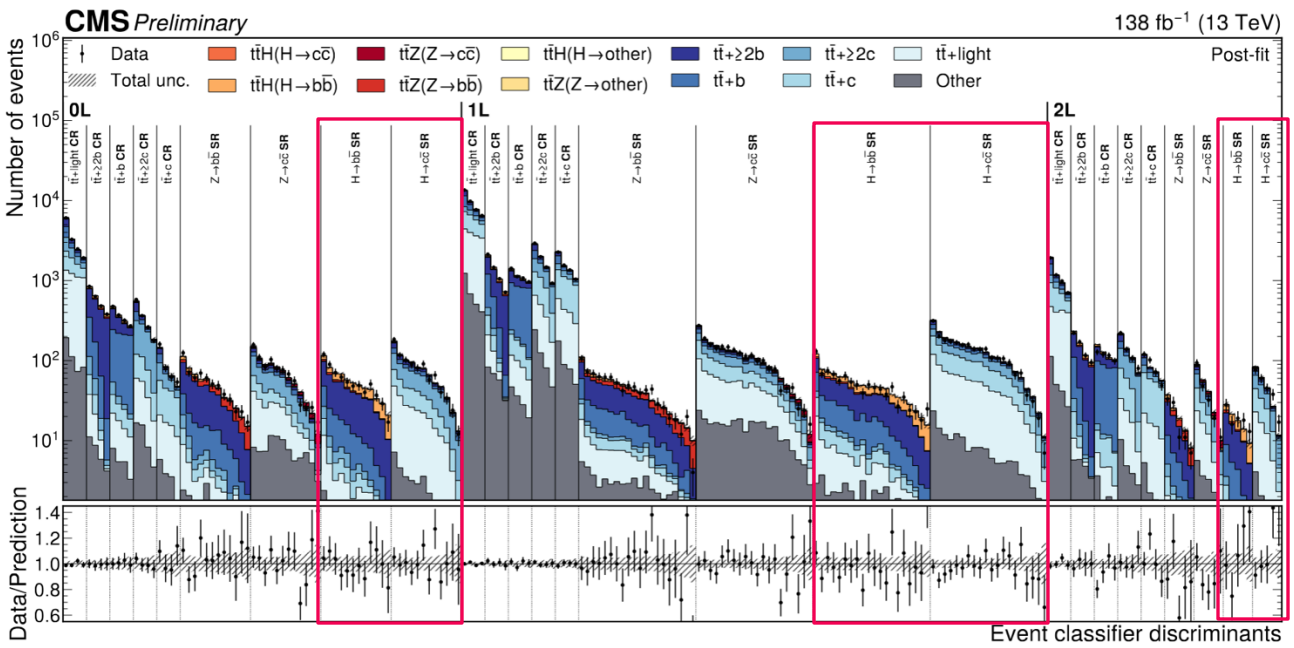




ParticleNet → **2x improvement** in background rejection wrt DeepJet

ttH→cc – Multiclass discriminator

- Multiclass event classifier based on ParT
- Final-state object (jets, leptons, missing energy) properties as input
- 10 (9) output classes in the 0L (1L, 2L) categories defined
  - **Signal:** $t\bar{t}H(bb)$, $t\bar{t}H(cc)$, $t\bar{t}Z(bb)$, $t\bar{t}Z(cc)$
  - **$t\bar{t}$+jets :** $t\bar{t}+\geq 2b$, $t\bar{t}+b$, $t\bar{t}+\geq 2c$, $t\bar{t}+c$, $t\bar{t}+\text{light}$
  - **QCD** (only in 0L)
- Signals extracted via a binned profile likelihood fit to data of the **ParT discriminator scores**



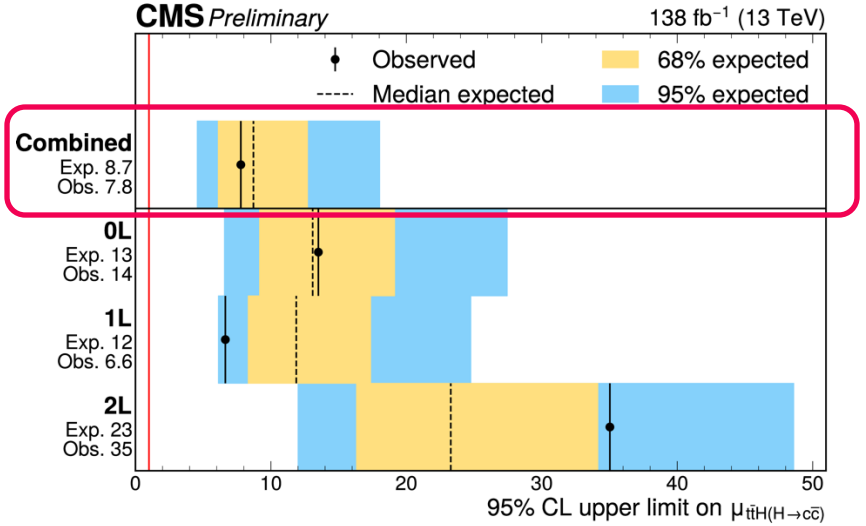
ttH→cc - Results

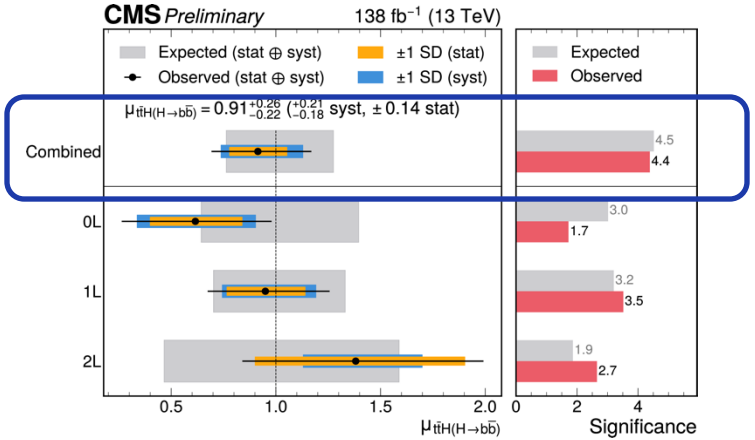
- First upper limit on ttH→cc

$$\mu = \frac{\sigma(ttH)\cdot B(H\to c\bar{c})}{\sigma(ttH)_{SM}\cdot B(H\to c\bar{c})_{SM}} < 7.8\ (8.7)\ \ @95\%\ CL$$

- Constraints on the Higgs-charm Yukawa coupling modifier k_c:

$$|k_c| < 3.0\ \ (|k_c| < 3.3)\ \ @95\%CL$$





ttH→ bb measurement

- Compatible with SM
- Highest sensitivity to ttH→bb

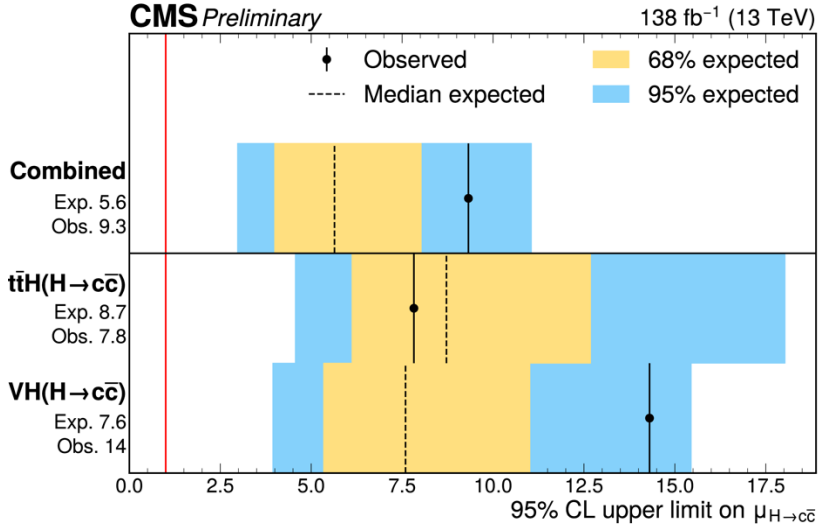
Search for H→cc - Combination

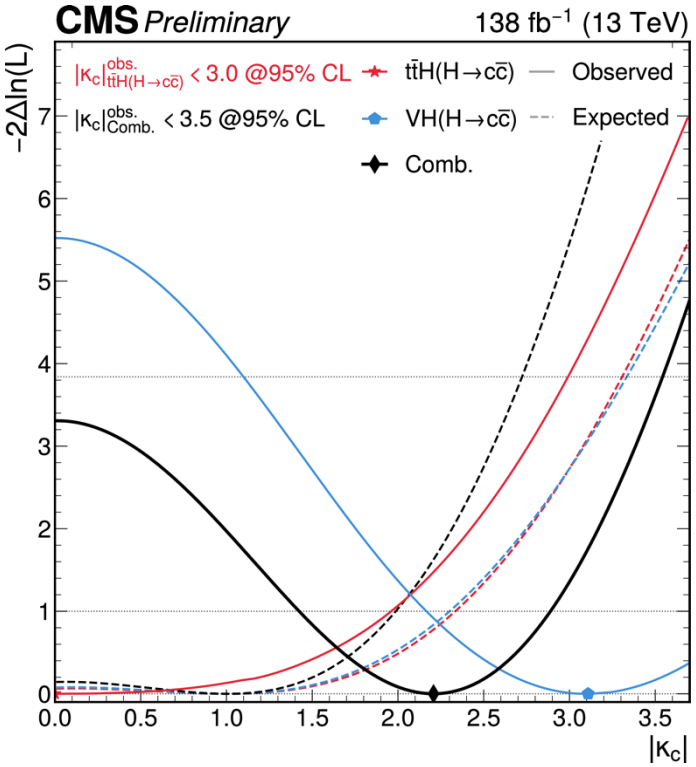
VH and ttH combination:

$\mu_{Hcc} < 9.3\ (5.6) \quad @95\%CL$

$|k_c| < 3.5\ (|k_c| < 2.7) \quad @95\%CL$

→ **Best result to date!**





Search for H→cc - Conclusions

- Search for H→cc greatly benefits from development of **new taggers** and sophisticated **ML techniques** for background reduction
- Run-2 analyses (138 fb^{-1}) investigated ggH, **VH** and **ttH** production modes
- First measurement of ttH→cc competitive to VH ATLAS and CMS results
- VH and ttH combination → **most stringent constraint to date on k_c**

$$\mu_{Hcc} < 9.3\ (5.6) \quad @95\%CL$$

$$|k_c| < 3.5\ \ (|k_c| < 2.7) \quad @95\%CL$$

→ **Best result to date!**

HL-LHC Projection

Projection on **VH→cc** combination by **ATLAS** and **CMS**

$$|k_c| < 1.5\ \ @95\%CL$$

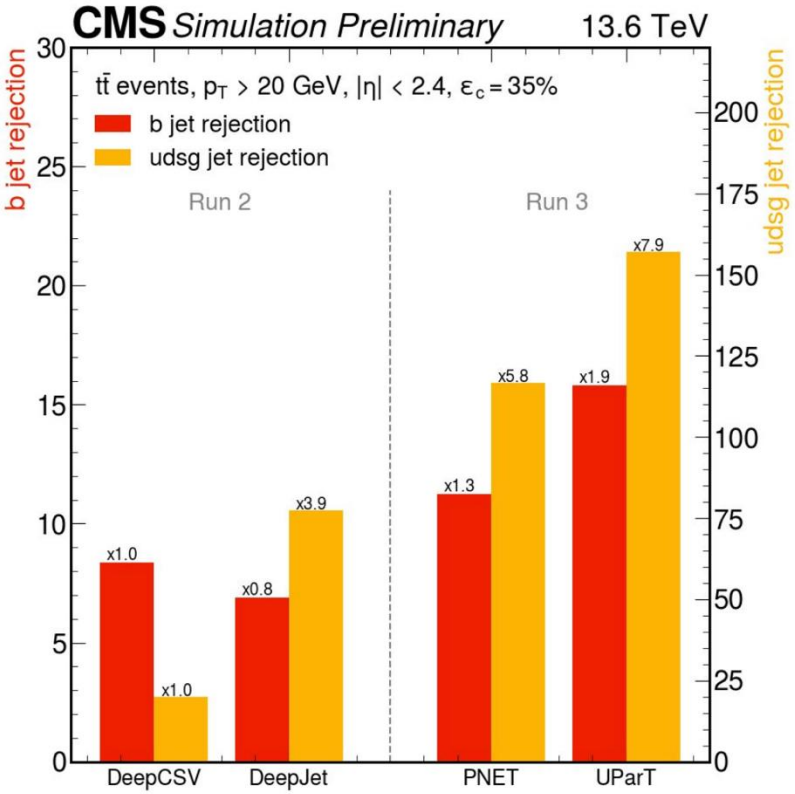
ATL-PHYS-PUB-2025-018
CMS-HIG-25-002

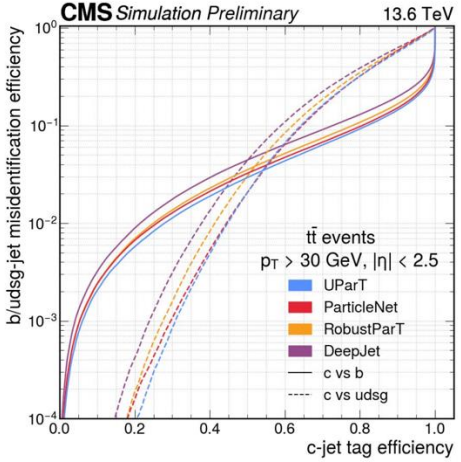
Further developments foreseen for Run3 and HL-LHC on trigger, flavour tagging and analysis strategies

Backup

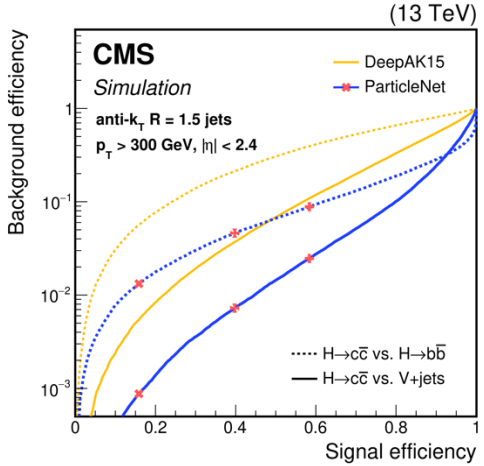
angela.zaza@cern.ch

C-tagging in CMS





AK4 jets

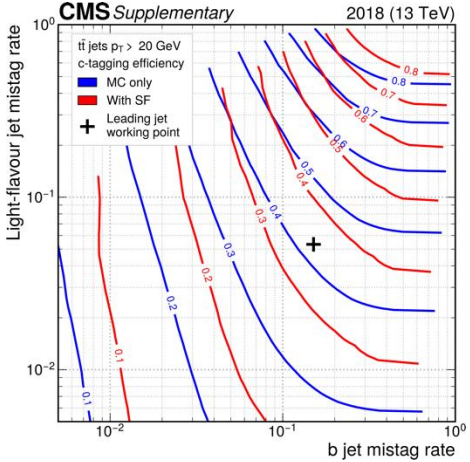


AK15 jets

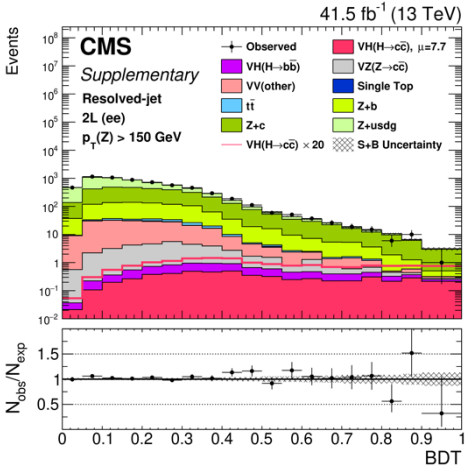
Resolved analysis VH→cc

- Higgs candidate reconstructed from two AK4 c-tagged jets
- Vetoed events with AK15 jets with p_T > 300 GeV
- BDT for signal vs background discrimination
  - Separate BDTs trained for each category
  - Kinematic and particle flavour variables
- **V+jets** and **tt̄** dominant background yields extracted from data in dedicated CRs obtained by inverting the c-tagging selections
- Signal strength modifier μ extracted from a maximum likelihood fit to data of the BDT output score

$$\mu = \frac{(\sigma B)_{obs}}{(\sigma B)_{SM}}$$



DeepJet c-tagging

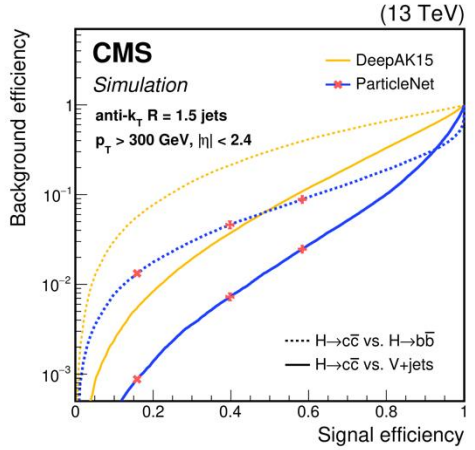


Post-fit distribution of the BDT score in the 2L(ee) category in 2017 data

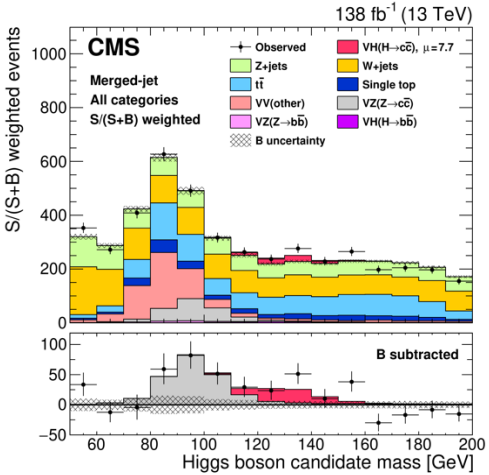
CMS-results HIG-21-008

Merged analysis VH→cc

- Higgs candidate reconstructed from the p_T-dominant AK15 jets
- **ParticleNet** X→cc tagger
  - **5x** better **H→bb rejection** and **2x** better **sensitivity** than DeepAK15
- BDT for signal vs background discrimination
  - Input variables not correlated with Higgs candidate mass
  - Signal and CR regions defined accordingly to the BDT score
- **3 categories** defined based on cc-score of the Higgs candidate
- Signal strength modifier μ extracted from a maximum likelihood fit to AK15 jet regressed mass



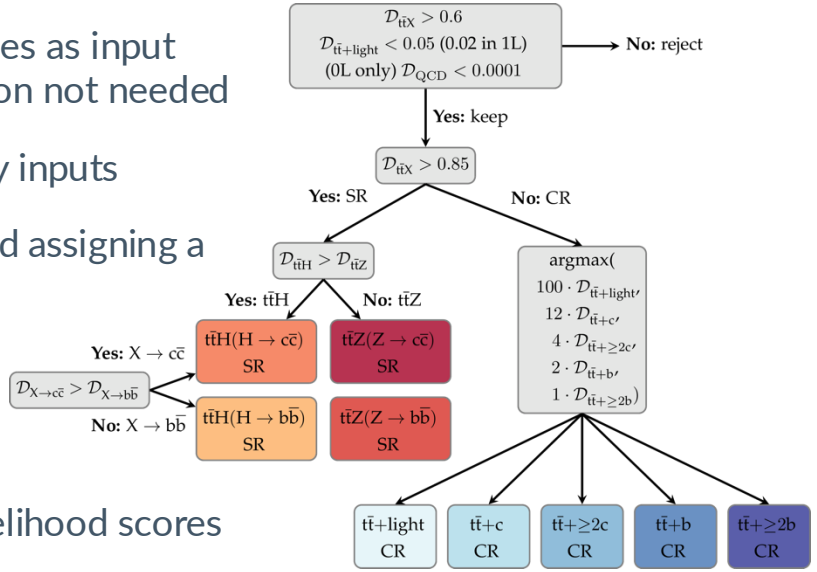
ParticleNet Hcc-tagging



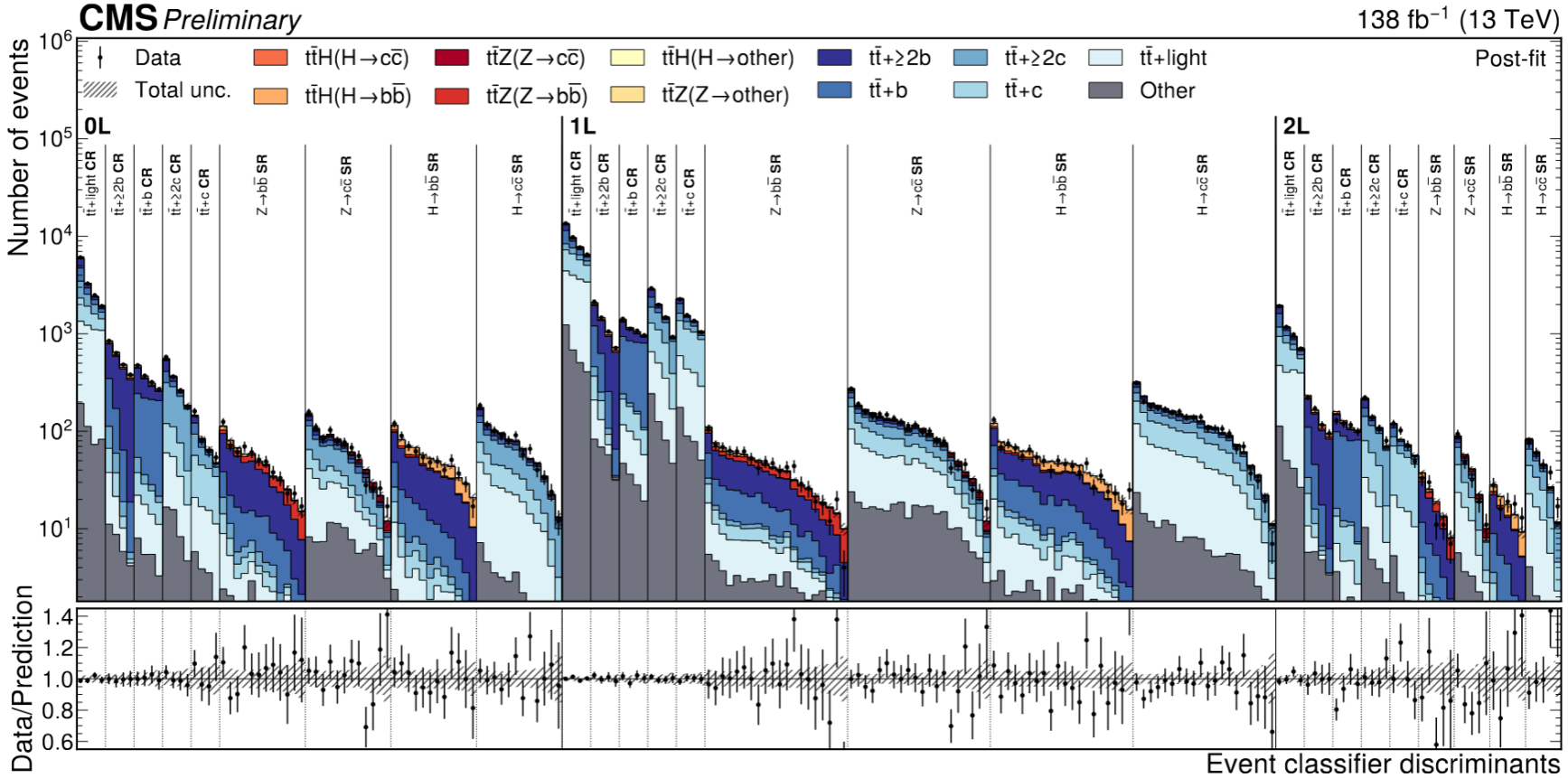
Combined m(H_{cand}) distribution in all channels of the merged-jet analysis

ttH→cc – Multiclass discriminator

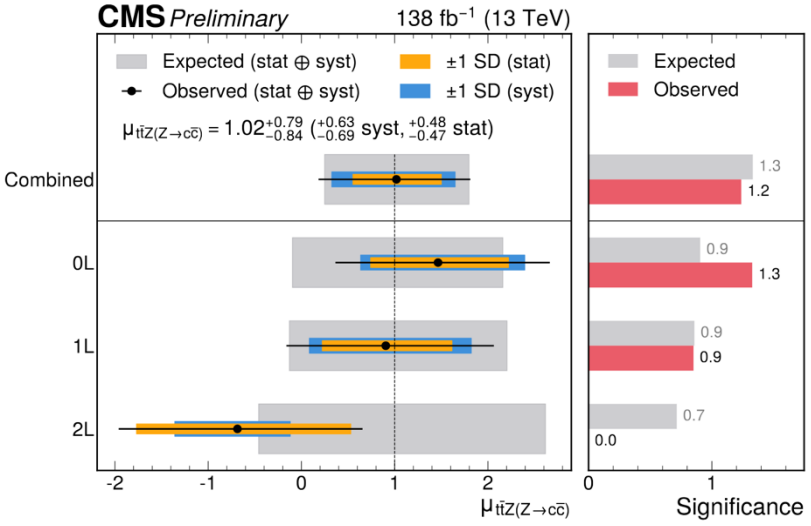
- Multiclass event classifier based on ParT
- Final-state object (jets, leptons, missing energy) properties as input
  → prior reconstruction of the top quarks and Higgs boson not needed
- Tagging categories (B0-B4, C0-C4) encoded in 10 binary inputs
- 10 (9) output classes in the 0L (1L, 2L) categories defined assigning a likelihood score to each class
  - **Signal:** $t\bar{t}H(bb)$, $t\bar{t}H(cc)$, $t\bar{t}Z(bb)$, $t\bar{t}Z(cc)$
  - **$t\bar{t}$+jets** : $t\bar{t}+\geq 2b$, $t\bar{t}+b$, $t\bar{t}+\geq 2c$, $t\bar{t}+c$, $t\bar{t}+light$
  - **QCD** (only in 0L)
- 4 SRs and 5 CRs defined based on thresholds on the likelihood scores
  $t\bar{t}$+jets normalization factors estimated in the CRs
- Signals extracted via a binned profile likelihood fit to data of the ParT discriminator scores

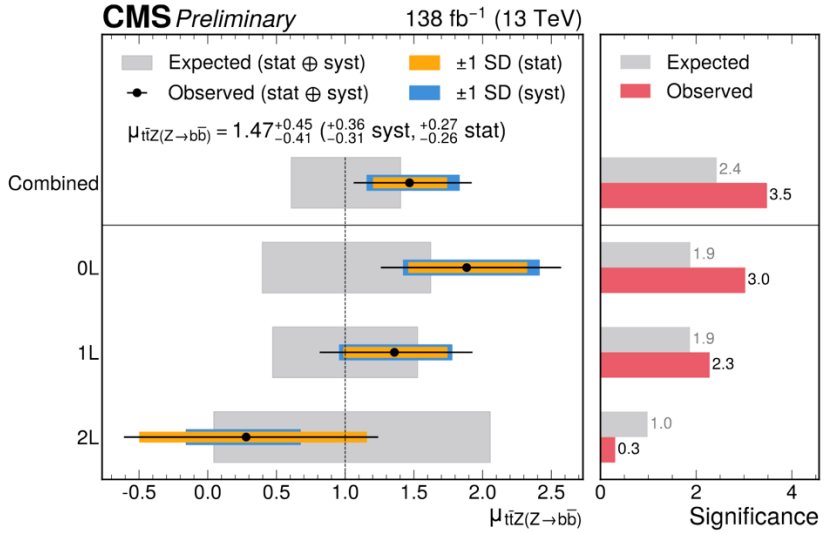


ttHcc – ParT Classifier



ttZcc/bb


CMS Preliminary
138 fb^{-1} (13 TeV)
Expected (stat ⊕ syst)
Observed (stat ⊕ syst)
±1 SD (stat)
±1 SD (syst)
$\mu_{t\bar{t}Z(Z\to c\bar{c})} = 1.02^{+0.79}_{-0.84}$ ($^{+0.63}_{-0.69}$ syst, $^{+0.48}_{-0.47}$ stat)
Combined
0L
1L
2L
-2
-1
0
1
2
$\mu_{t\bar{t}Z(Z\to c\bar{c})}$
Expected
Observed
1.3
1.2
0.9
1.3
0.9
0.9
0.7
0.0
0
1
Significance


CMS Preliminary
138 fb^{-1} (13 TeV)
Expected (stat ⊕ syst)
Observed (stat ⊕ syst)
±1 SD (stat)
±1 SD (syst)
$\mu_{t\bar{t}Z(Z\to b\bar{b})} = 1.47^{+0.45}_{-0.41}$ ($^{+0.36}_{-0.31}$ syst, $^{+0.27}_{-0.26}$ stat)
Combined
0L
1L
2L
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
$\mu_{t\bar{t}Z(Z\to b\bar{b})}$
Expected
Observed
2.4
3.5
1.9
3.0
1.9
2.3
1.0
0.3
0
2
4
Significance