VBF H→ $c\bar{c}$ search at Run3

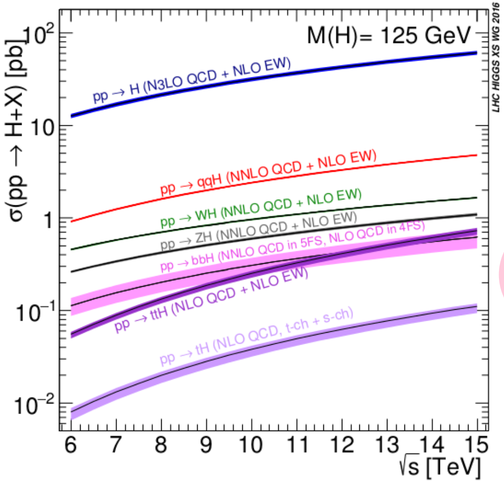
CMS Bari meeting

Angela Zaza

20/12/2024

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

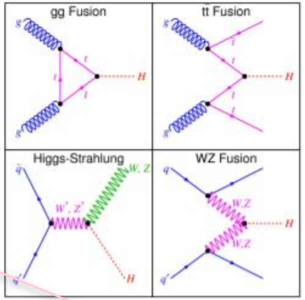
Higgs production at LHC



ggF

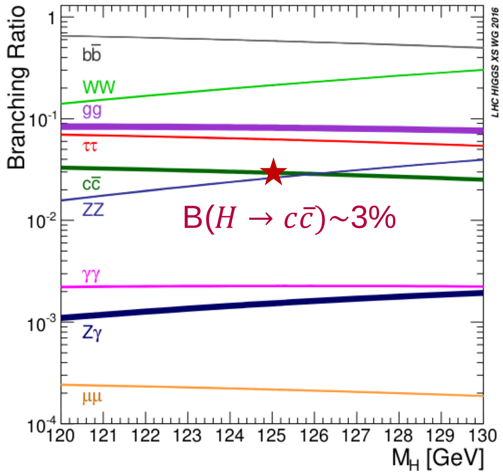
ttH (and bbH)

VH



VBF

Higgs decay modes



- Couplings to weak bosons and third generation fermions measured with 10-20% precision → compatible with the SM
- Couplings to second generation quarks (Hcc) **extremely challenging**: small rate and overwhelming QCD background

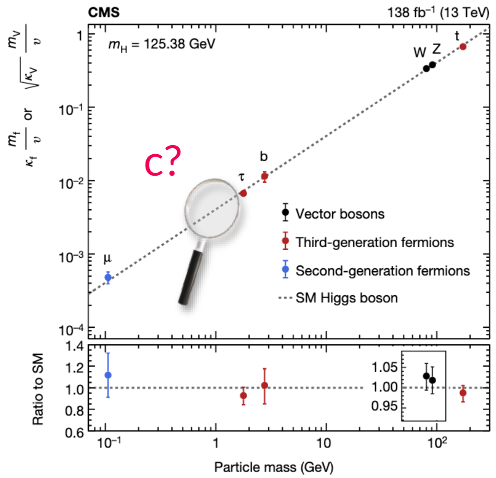
VH mechanism investigated at CMS with Run-2 data

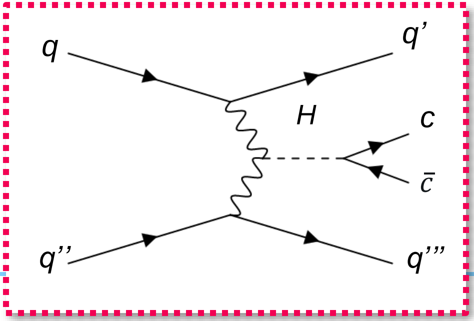
upper limit: $\frac{\sigma(VH)\cdot B(H\to c\bar{c})}{\sigma(VH)_{SM}\cdot B(H\to c\bar{c})_{SM}} < 14$

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

PhD Program:

- Investigate **VBF** production mode for the first time
  - Higher cross section (~7% of the total XS)
  - Overwhelming QCD background
  - q' and q''' jets produced forward and backward, approaching the beam-pipe
- Perform the search using Run-3 2023 data
  - Develop a trigger specific for VBF $H \to c\bar{c}$
  - Explore Neural Network approaches for signal to background discrimination
  - Estimate upper limit on the signal strength

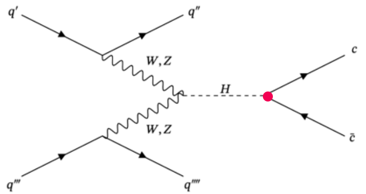


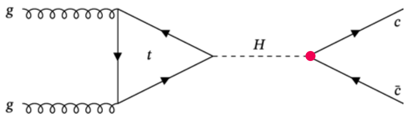


Analysis strategy

Signal: $\mathbf{VBF}\ H \to cc$
$gg\,H \to cc$ (considered as signal because difficult to disentangle)

Main Backgrounds: QCD (dominant) } continuum
$\mathbf{VBF}\ H \to bb,\ gg\,H \to bb$
$\mathbf{QCD}\ Z/W \to qq,\ \mathbf{EWK}\ Z/W \to qq$ } peaking





Trigger: **HLT_QuadPFJet100_88_70_30_PNetTag1CvsAll0p5_VBF3Tight**
(online from 2023)

Offline preselection (similar to trigger) → **BDT for signal vs QCD discrimination** → **Signal and background modelling** → **Final result: expected upper limit on the signal strength**

Signal and background modelling:
QCD → data-driven*
resonant bkg → MC

- Higgs boson reconstructed from the two most c-tagged jet (PNet)

*MC simulated QCD multijet events poorly model data

HLT path for VBF $H \to c\bar{c}$

HLT_QuadPFJet100_88_70_30_PNetTag1CvsAll0p5_VBF3Tight

Trigger acceptance: 1.9%

Kinematic sequence

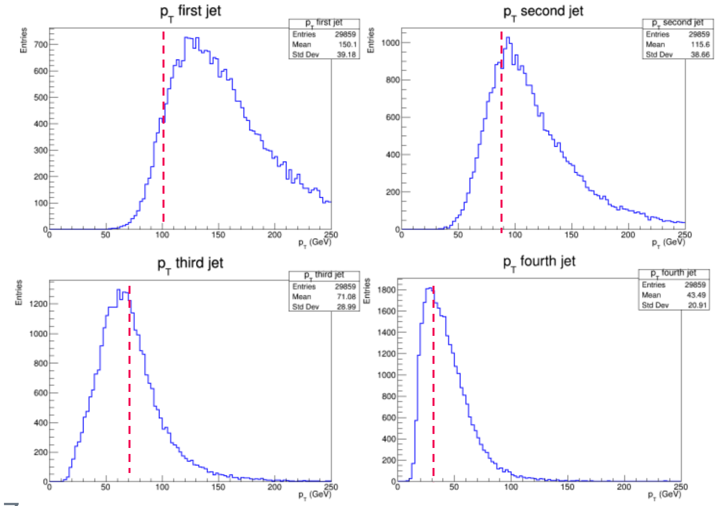
4 jets with p_T > 100, 88, 70, 30 GeV
2 jets with p_T > 30 GeV η < 2.5

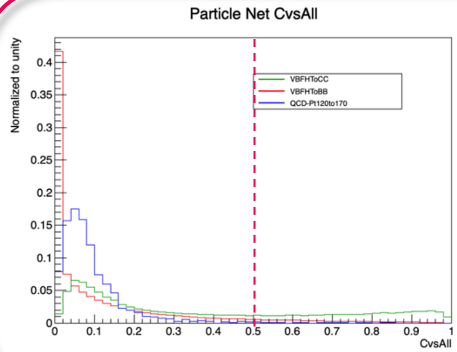
C-tagging sequence

1 jet with p_T > 30 GeV, η < 2.5 and CvsAll>0.5

VBF-tagging sequence

Among the 4 p_T-leading jets:
2 jets with highest CvsAll score: c-jets
other 2 jets: VBF jets (qq), with
$\Delta\eta(qq) > 3.5$
$M(qq) > 460\ GeV$





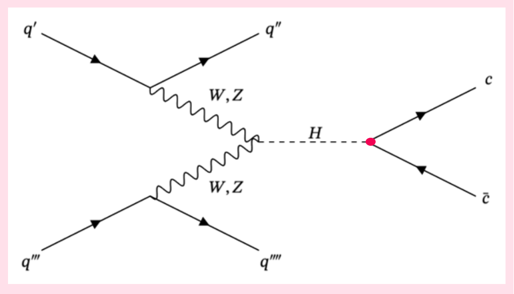
ParticleNet: powerful jet flavour tagging algorithm based on Dynamic Graph Convolutional Neural Network

$$\text{CvsAll} = \frac{\text{prob(c)}}{\text{prob(c)+prob(b)+prob(uds)+prob(g)+prob(tauh)}}$$

Offline pre-selection

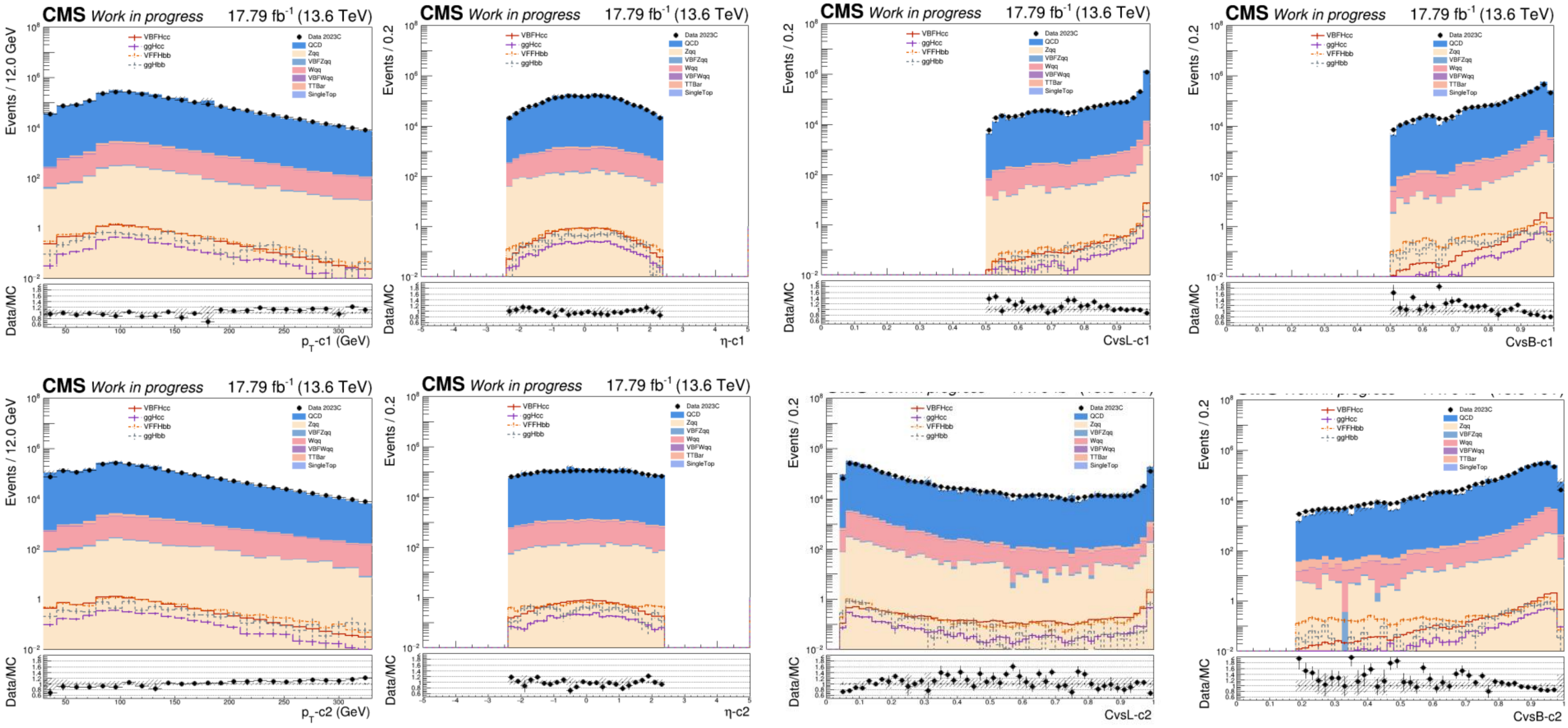
Gen weights, PU reweigthing, trigger SFs, JECs, jet veto maps applied

- Trigger HLT_QuadPFJet100_88_70_30_PNet1CvsAll0p5_VBF3Tight_v
- Electron/Muon veto
- MET pT < 170 GeV
- 4 leading p_T jets with pt>105, 90, 75, 35 GeV and η < 4.7 matching with HLT objects
- 2 jets with highest CvsL score and η < 2.4: c-jets
- Other 2 jets: VBF-jets
- Jet c1: CvsAll > 0.51, CvsL > 0.16, CvsB > 0.304
- Jet c2: CvsL > 0.054, CvsB > 0.182
- VBF jets: Mass > 500 GeV, $\Delta\ \eta$ > 3.8



After offline pre-selection

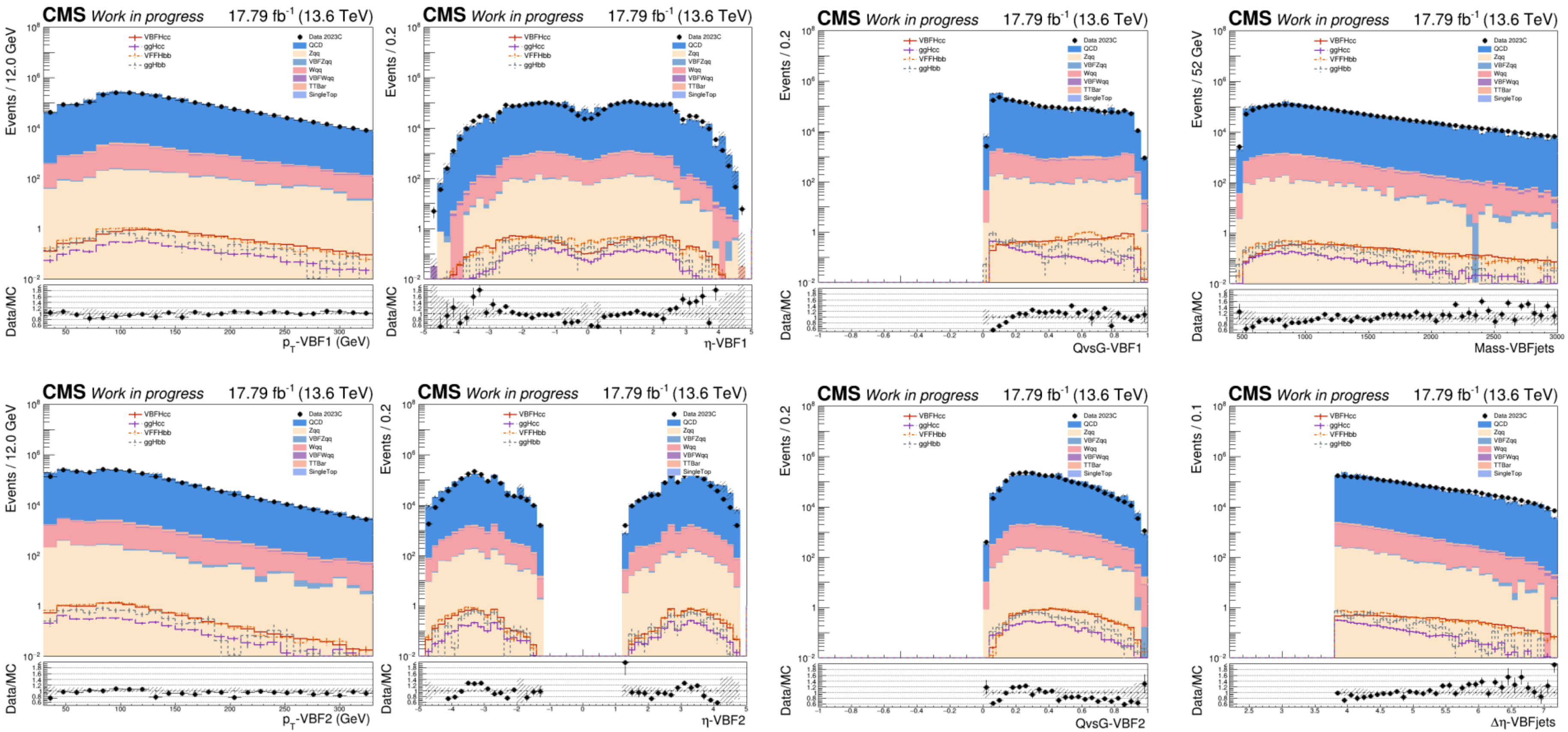
2023 C



K=1.8 applied to QCD

After offline pre-selection

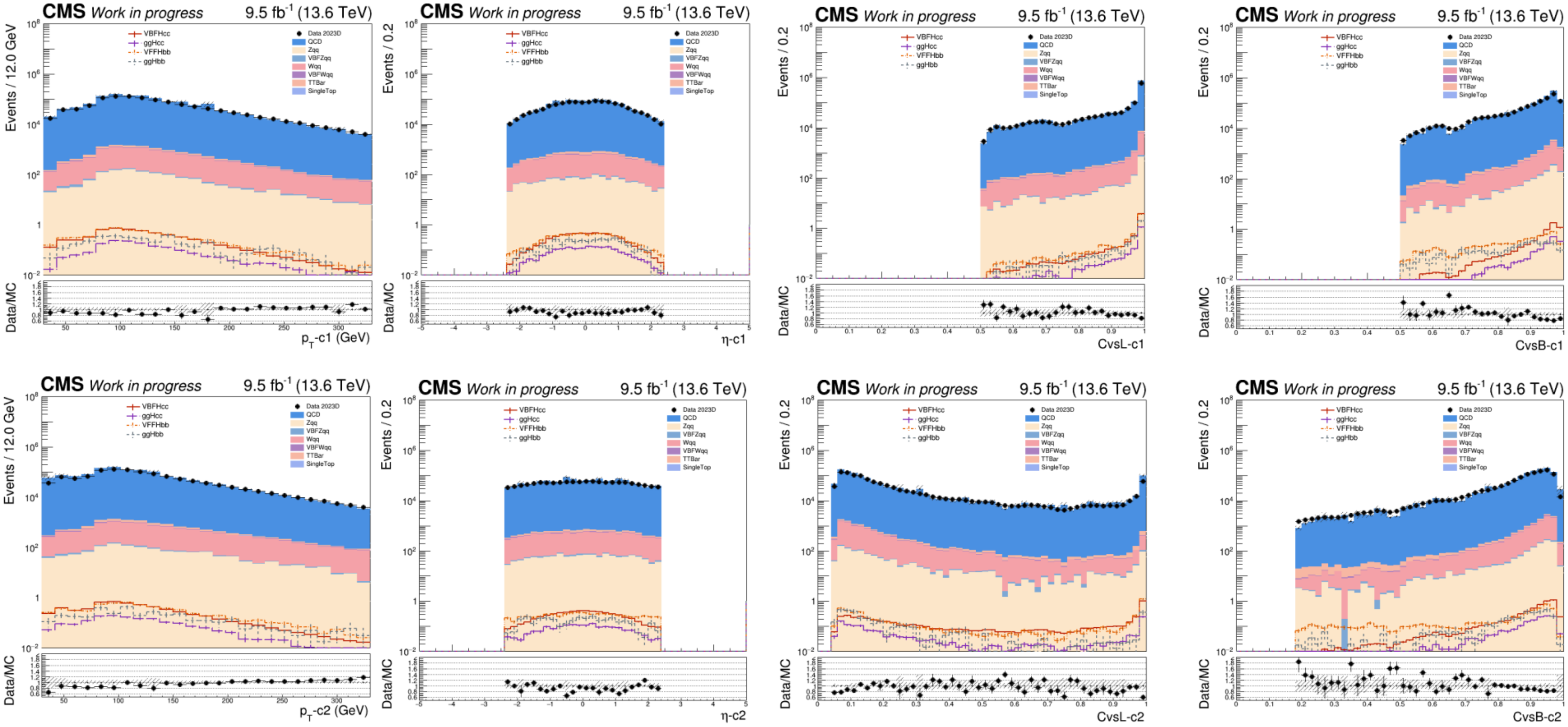
2023 C



K=1.8 applied to QCD

After offline pre-selection

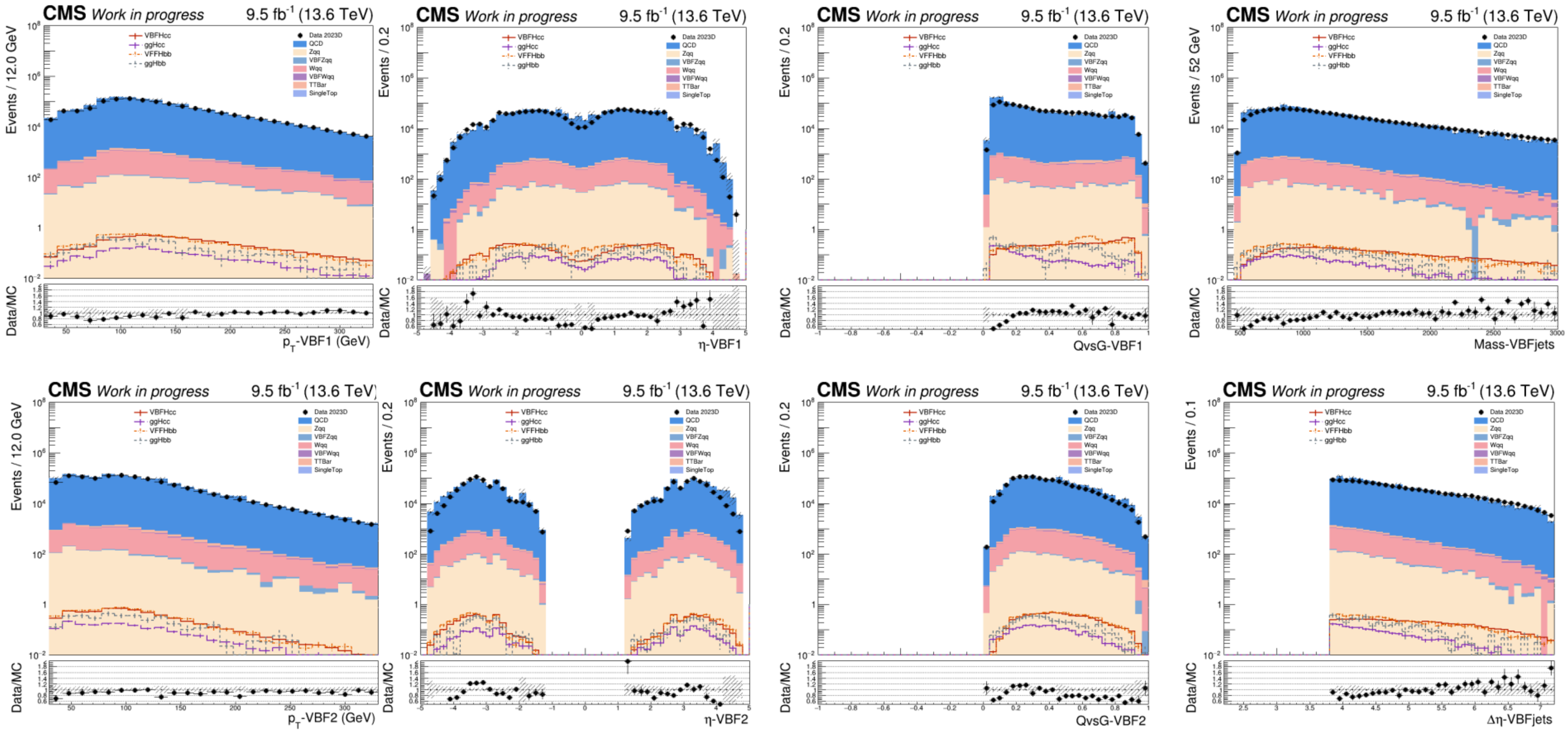
2023 D



K=1.8 applied to QCD

After offline pre-selection

2023 D



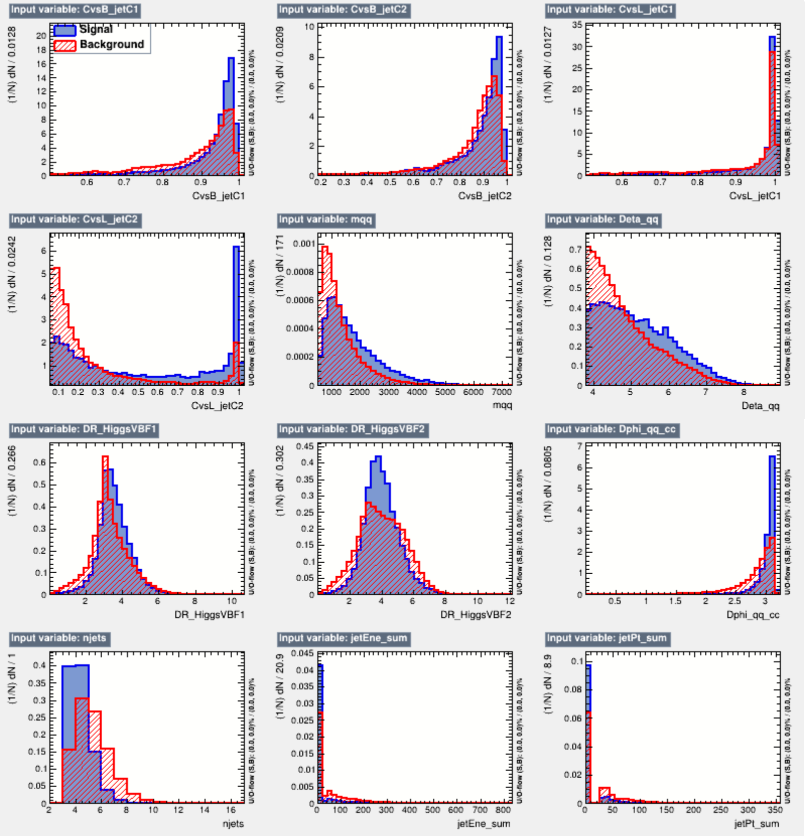
K=1.8 applied to QCD

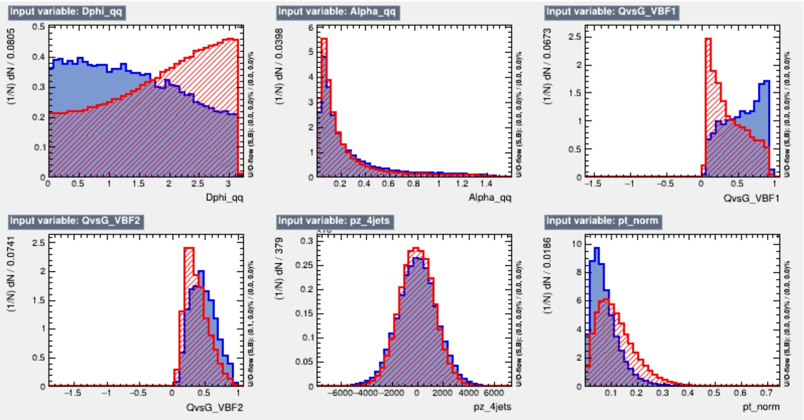
BDT model for H $\rightarrow c\bar{c}$ vs QCD discriminator

BDTG (Gradient Boosted Decision Tree) → Signal/QCD background discrimination

- Signal: VBF $H \rightarrow c\bar{c}$ ~60k events
- QCD background: data in sidebands
  (H_{mass}<80 GeV and H_{mass} >200 GeV)

BDT features

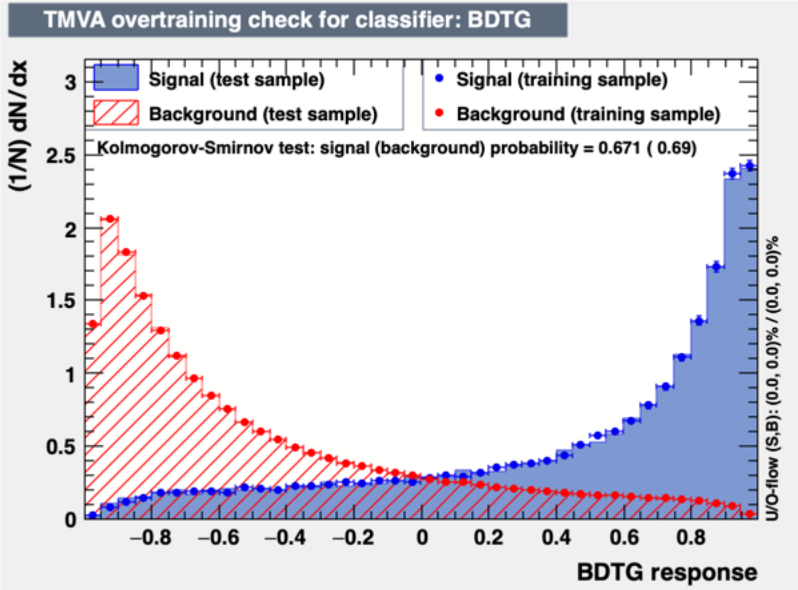


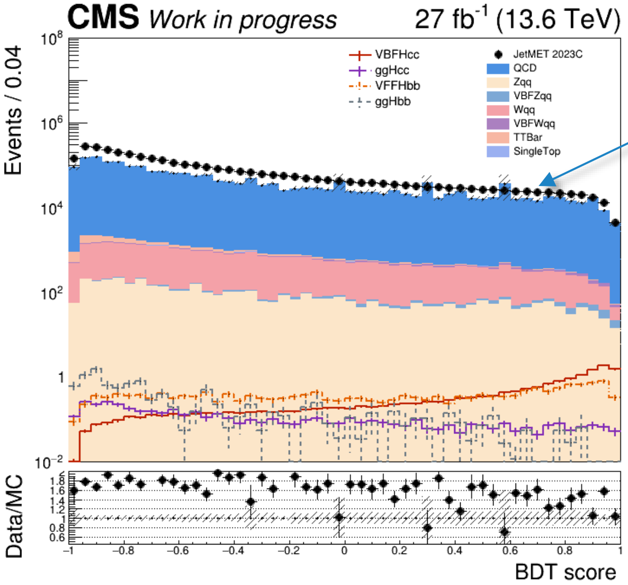


- α_{qq}: $\mathrm{Min}(\alpha_{q1}, \alpha_{q2})$, where $\alpha_{q1/q2}$ is the angle between lead/sublead VBF-jet and the boosted system of the VBF jet pair.
- QvsG ParticleNet score
- c-tagging: CvsL and CvsB Particle Net scores of the two c-jets.
- $p_z^{q1} + p_z^{q2} + p_z^{c1} + p_z^{c2}$: total longitudinal momentum of the selected four jets.
- $\sum \vec{p}_T^{\,i} / \sum p_T^i$, where i=q1, q2, c1, c2: Normalized sum of the transverse momentum.
- Angular distance: ΔR between the Higgs bososn candidate and the leading and subleading VBF jets.
- $\Delta(\phi_{qq} - \phi_{cc})$: difference of the azimuthal angle of the VBF jet pair system and the c-jet pair system (H candidate).
- The jet multiplicity in the region of $|\eta|$<2.4 above 20 GeV.
- Sum of the energy and the transverse momentum of all the jets above p_T > 30 GeV and $|\eta|$<2.4 excluding the selected four jets.

- m_{qq}: invaroant mass of the two VBF jets.
- $|\Delta\eta_{qq}|$: absolute pseudorapidity difference of the two VBF jets.
- $\Delta\phi_{qq}$: absolute azimuthal angle difference of the two VBF jets.

H $\rightarrow c\bar{c}$ vs QCD discriminator





- QCD from MC (mismodeling effects) better agreement in the signal region

preliminary

- Overall good agreement between training and test distributions
- Three categories based on the BDT score

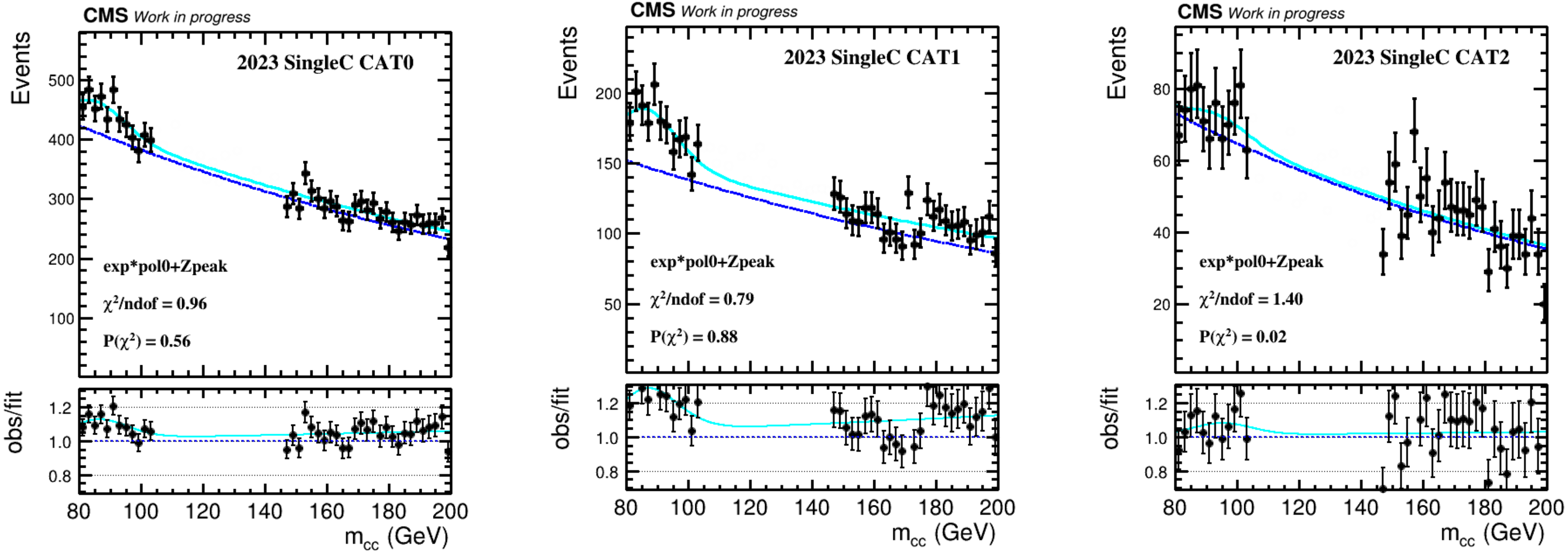
category	BDT score
CAT0	0.8 – 0.9
CAT1	0.9 – 0.95
CAT2	0.95 – 1

A. Zaza

Continuum background modelling

- ▷The continuum background shape is extracted from an exponential fit of the mass spectrum in the sideband: [80,104] – [146,200] GeV
- ▷Z boson peak visible

Era C+D



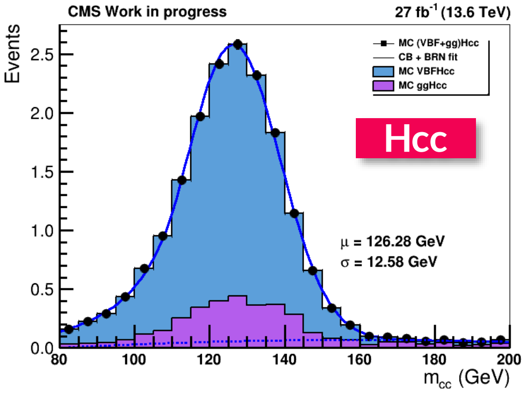
QCD yield: 19491.4

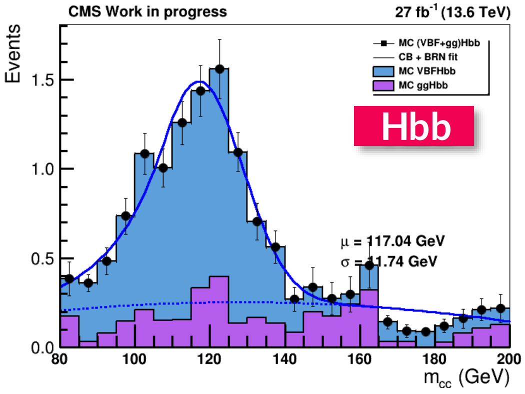
QCD yield: 8130.8

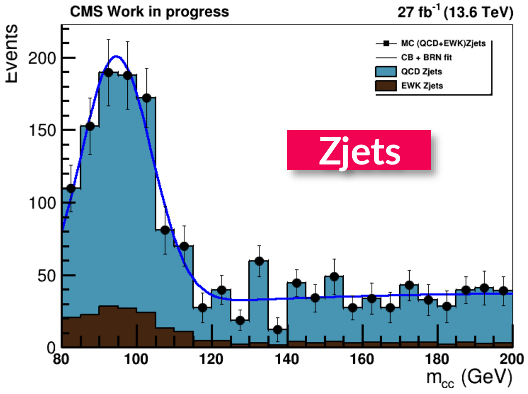
QCD yield: 3119.7

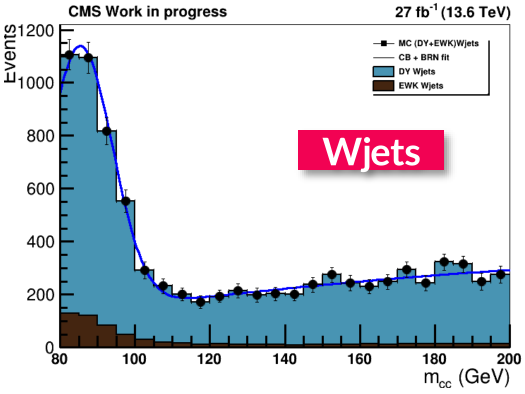
MC shapes (Hcc, Hbb, Zqq, Wqq)

preliminary









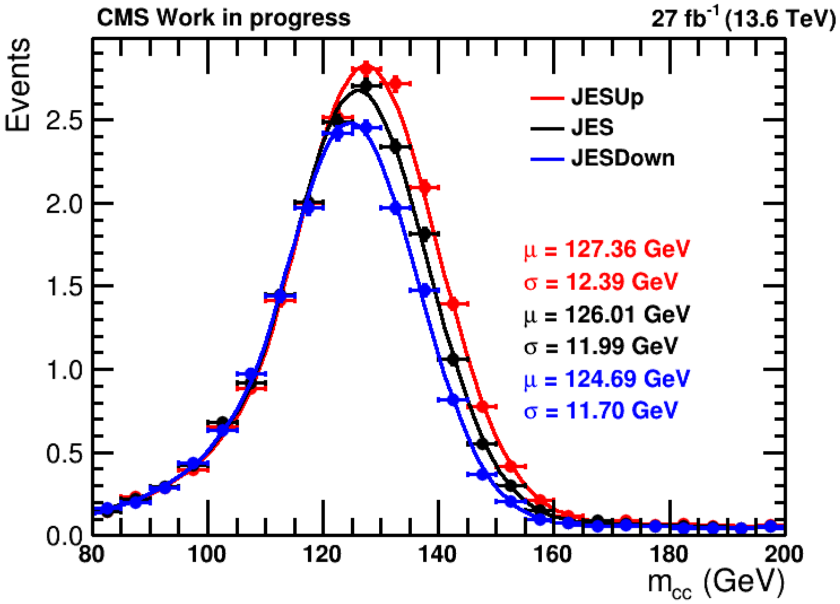
No thresholds applied on the BDT score

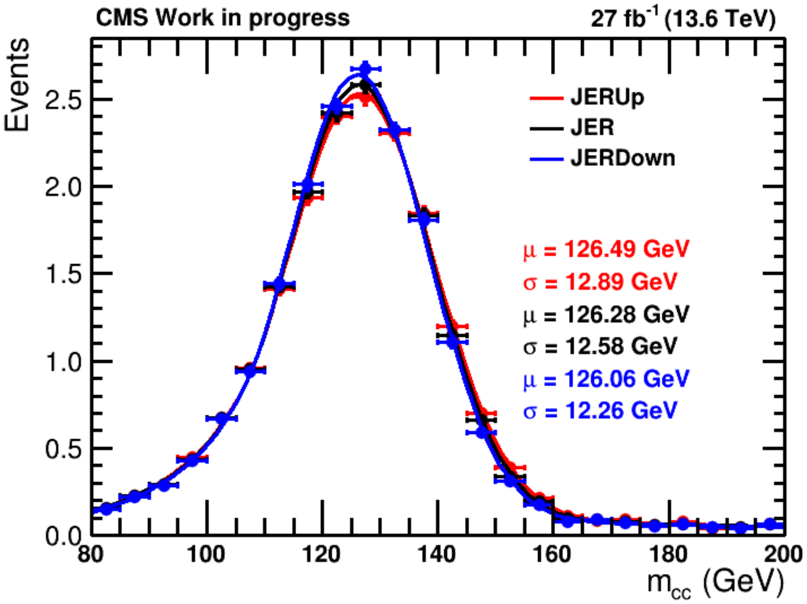
Yields in each category (histo integrals)

process	CAT0	CAT1	CAT2
VBFHcc	2.63	2.13	2.01
ggHcc	0.14	0.09	0.06
VBFHbb	1.39	0.98	0.51
ggHbb	0.06	0	0
Zjets	92.02	45.80	24.71
Wjets	241.06	52.89	24.21

MC distributions fitted with CB and Bernstein polynomial function

Effect of jet correction uncertainties on signal fit





JES and JER uncertainties impacts on the parameters of the fit → treated as shape uncertainties

Systematic uncertainties for combine

- BR H→cc
- Lumi
- Trigger SFs
- JES and JER
- QCD scale ggH
- QCD scale qqH
- Pdf ggH
- Pdf qqH
- VBF dipole recoil
- α_S ggH
- α_S qqH
- QCD Fit Bias
- c-tag SF (10%)

Simplistic approach implemented up to now:

- Fit the QCD background with an alternative function: exp*pol1 fit (instead of exp*pol0)
  → limit changes within 10%

- conservative 20% uncertainties assumed on the signal
  (spurious signal method)

```
bias_SingleC0_2023     param    0    0.2
bias_SingleC1_2023     param    0    0.2
bias_SingleC2_2023     param    0    0.2
bias_SingleC0_2023 rateParam SingleC0_2023 ggH_hcc_bias  0 [-1,1]
bias_SingleC1_2023 rateParam SingleC1_2023 ggH_hcc_bias  0 [-1,1]
bias_SingleC2_2023 rateParam SingleC2_2023 ggH_hcc_bias  0 [-1,1]
bias_SingleC0_2023 rateParam SingleC0_2023 qqH_hcc_bias  0 [-1,1]
bias_SingleC1_2023 rateParam SingleC1_2023 qqH_hcc_bias  0 [-1,1]
bias_SingleC2_2023 rateParam SingleC2_2023 qqH_hcc_bias  0 [-1,1]
```

- To be fine tuned by generating toys with nominal function and fit with alternative function

Expected upper limit

preliminary

- UL extracted using parametric shape of signal and background
- QCD yield extracted from data in sidebands with a simultaneous fit with W and Z peaks
- Shape tamplate of samples VBFHcc/bb, ggHcc/bb, Zjets and Wjets estimated before categorization, than the model is applied to each category

```
combine -M AsymptoticLimits --run blind Datacards_VBFHCC.root
```

$$\mu = \frac{XS(VBF) \cdot BR(H \to c\bar{c})}{XS(VBF)^{SM} \cdot BR(H \to c\bar{c})^{SM}} \leq 30.87 \quad @\,95\%\,CL$$

Extrapolation to Run-2 Luminosity: UL~14
→ comparable to Run-2 VH expected limit

Thank you

angela.zaza@cern.ch

Backup

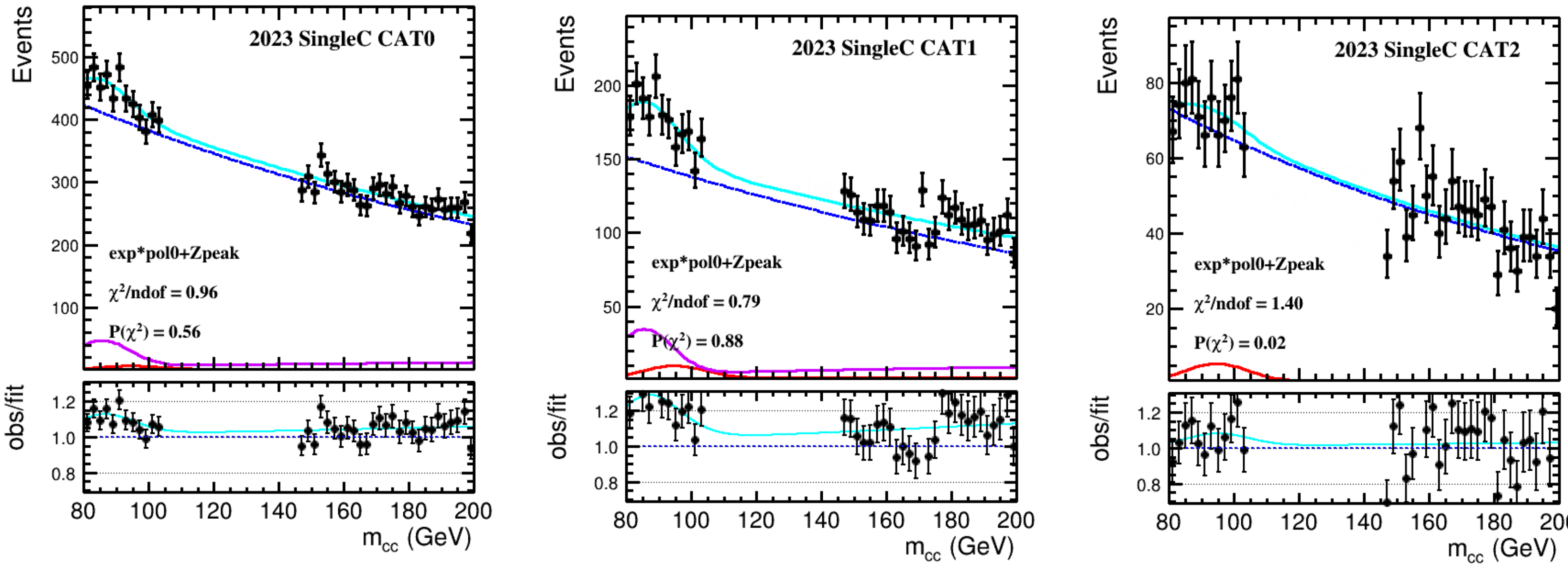
angela.zaza@cern.ch

Continuum background modelling

- ▷The continuum background shape is extracted from an exponential fit of the mass spectrum in the sideband: [80,104] – [146,200] GeV

Era C+D

Z peak: red, W peak: violet



Datasets

Era	Dataset	GT
2023 C 17.79 fb^{-1}	/JetMET0/Run2023C-PromptReco-v1/MINIAOD /JetMET0/Run2023C-PromptReco-v2/MINIAOD /JetMET0/Run2023C-PromptReco-v3/MINIAOD /JetMET0/Run2023C-PromptReco-v4/MINIAOD /JetMET1/Run2023C-PromptReco-v1/MINIAOD /JetMET1/Run2023C-PromptReco-v2/MINIAOD /JetMET1/Run2023C-PromptReco-v3/MINIAOD /JetMET1/Run2023C-PromptReco-v4/MINIAOD	130X_dataRun3_Prompt_v3
2023D 9.5 fb^{-1}	/ParkingVBF0/Run2023D-PromptReco-v1/MINIAOD /ParkingVBF0/Run2023D-PromptReco-v2/MINIAOD /ParkingVBF1/Run2023D-PromptReco-v1/MINIAOD /ParkingVBF1/Run2023D-PromptReco-v2/MINIAOD /ParkingVBF2/Run2023D-PromptReco-v1/MINIAOD /ParkingVBF2/Run2023D-PromptReco-v2/MINIAOD /ParkingVBF3/Run2023D-PromptReco-v1/MINIAOD /ParkingVBF3/Run2023D-PromptReco-v2/MINIAOD /ParkingVBF4/Run2023D-PromptReco-v1/MINIAOD /ParkingVBF4/Run2023D-PromptReco-v2/MINIAOD /ParkingVBF5/Run2023D-PromptReco-v1/MINIAOD /ParkingVBF5/Run2023D-PromptReco-v2/MINIAOD /ParkingVBF6/Run2023D-PromptReco-v1/MINIAOD /ParkingVBF6/Run2023D-PromptReco-v2/MINIAOD /ParkingVBF7/Run2023D-PromptReco-v1/MINIAOD /ParkingVBF7/Run2023D-PromptReco-v2/MINIAOD	130X_dataRun3_Prompt_v4

MC	GT= 130X_mcRun3_2023_realistic_v14

/VBFHto2C_M-125_TuneCP5_13p6TeV_powheg-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v3/MINIAODSIM
/GluGluHto2C_M-125_TuneCP5_13p6TeV_powheg-minlo-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v2/MINIAODSIM
/VBFHto2B_M-125_TuneCP5_13p6TeV_powheg-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v3/MINIAODSIM
/GluGluHto2B_M-125_TuneCP5_13p6TeV_powheg-minlo-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v2/MINIAODSIM
/QCD-4Jets_HT-*_TuneCP5_13p6TeV_madgraphMLM-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v2/MINIAODSIM
/TTto2L2Nu_TuneCP5_13p6TeV_powheg-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v2/MINIAODSIM
/TTto4Q_TuneCP5_13p6TeV_powheg-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v2/MINIAODSIM
/TTtoLNu2Q_TuneCP5_13p6TeV_powheg-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v2/MINIAODSIM
/TbarWplusto4Q_TuneCP5_13p6TeV_powheg-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v15-v2/MINIAODSIM
/TWminusto4Q_TuneCP5_13p6TeV_powheg-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v2/MINIAODSIM
/TbarWplustoLNu2Q_TuneCP5_13p6TeV_powheg-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v15-v2/MINIAODSIM
/TWminustoLNu2Q_TuneCP5_13p6TeV_powheg-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v2/MINIAODSIM
/Zto2Q-2Jets_PTQQ-*_2J_TuneCP5_13p6TeV_amcatnloFXFX-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v3/MINIAODSIM
/VBFZto2Q_TuneCP5_13p6TeV_madgraph-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v3/MINIAODSIM
/Wto2Q-2Jets_PTQQ-100to200_2J_TuneCP5_13p6TeV_amcatnloFXFX-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v3/MINIAODSIM
/VBFWto2Q_TuneCP5_13p6TeV_madgraph-pythia8/Run3Summer23MiniAODv4-130X_mcRun3_2023_realistic_v14-v3/MINIAODSIM

Summary and next steps

- Started writing the AN (PhD thesis as basis)
- Plan to get approval for Summer25 conferences (e.g. EPS in July)

To do:

- Bias study with combine
- Apply PNet jet energy regression
- Apply c-tagging correction (not available yet)
- BDT discriminant trained with TMVA → plan to move to XGBoost


Available on the CMS information server **CMS AN-24-243**

CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

2024/12/13
Archive Hash: untracked
Archive Date: 2024/12/13

Search for VBF $H \to c\bar{c}$ with Run3 2023 data

A. Zaza, R. Venditti, S. Mukherjee, A. Raspereza, A. Tumasyan,
A. Colaleo, L. Longo, A. Pellecchia, F. Simone, D. Troiano
CERN

Abstract

In this note, a search for the Higgs boson decay in a charm quark-antiquark pair in the Vector Boson Fusion production channel is presented. The study is conducted on proton-proton collision data at $\sqrt{s}$ = 13.6 TeV collected during the 2023 running period, which correspond to an integrated luminosity of 27 fb^{-1}. An upper limit on the signal strength is set at 95% CL.

DRAFT

This box is only visible in draft mode. Please make sure the values below make sense.

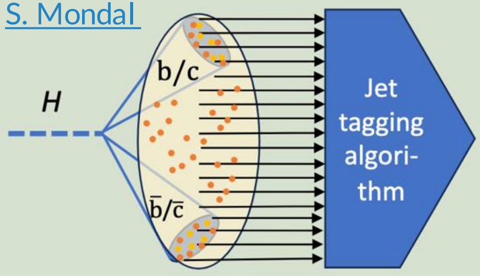
PDFAuthor:	A. Zaza, R. Venditti, S. Mukherjee, A. Raspereza, A. Tumasyan
PDFTitle:	Search for VBF H rightarrow c barc with Run3 2023 data
PDFSubject:	CMS
PDFKeywords:	CMS, your topics

Please also verify that the abstract does not use any user defined symbols

Plans for 2024 analysis

- Use the VBF parking triggers
  no selection applied on the jet flavour → possibility to study simultaneously Hcc/bb
- Explore new approaches:
  - GNN instead of the BDT for event cathegorization
  - new PAIReD jet tagger

S. Mondal



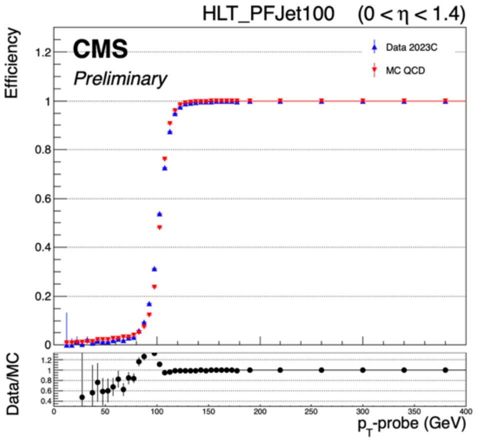
New PhD student:
Lisa Generoso

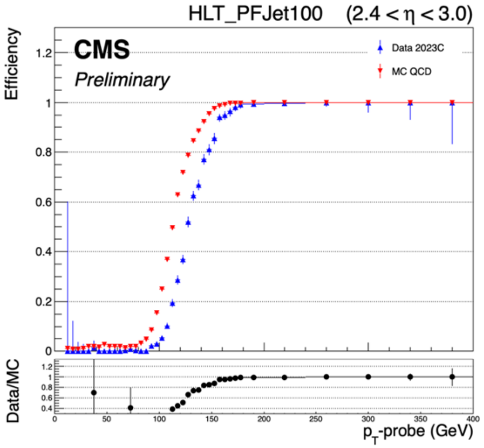
Trigger p_T scale factors

The efficiency of each of the first three highest p_T thresholds is evaluated with a **tag and probe method**:

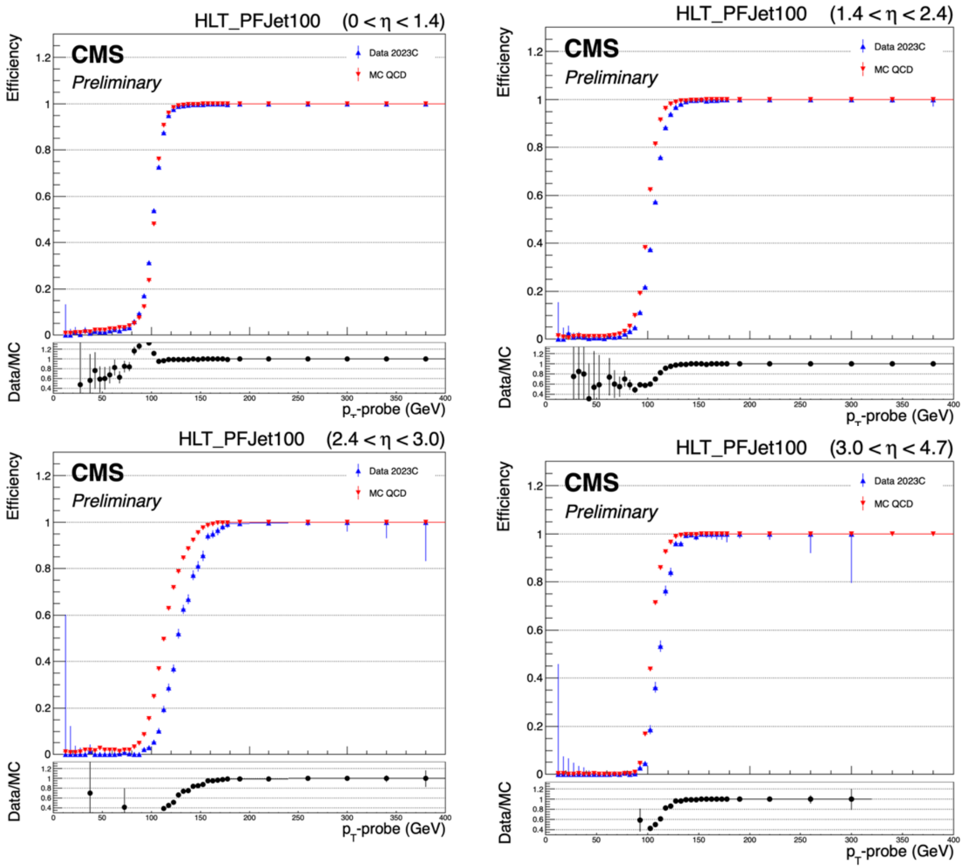
- ▷ Trigger: single AK4 jet with p_T > 60 (80) GeV
- ▷ Events with back-to-back di-jet topology selected
- ▷ Tag: leading offline AK4 jet with p_T > 110 (130) GeV matched to an HLT object
- ▷ Probe: subleading offline AK4 jet
- ▷ For each p_T threshold (*thr*), efficiency is evaluated as

$$\text{Efficiency} = \frac{p_T\, probe, matched\ to\ HLT\ jet\ with\ p_T > thr}{p_T\, probe}$$





Trigger p_T scale factors



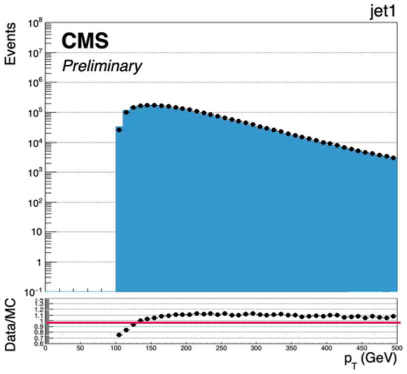
- <u>Top panel</u>: Efficiency as a function of the p_T of the probe jet evaluated on Data (blue) and MC QCD (red) in four different intervals of η for the first p_T threshold: 100 GeV

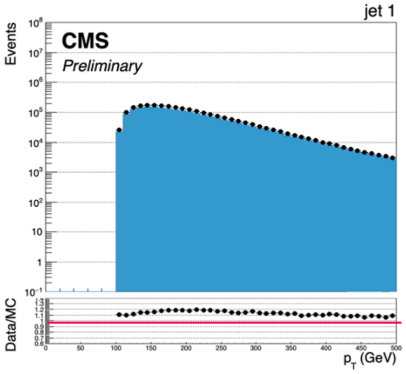
- <u>Bottom panel</u>: ratio between data and MC efficiencies

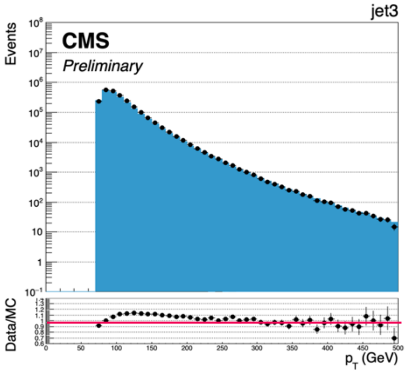
Analogous plots for the other p_T thresholds can be found in the backup slides

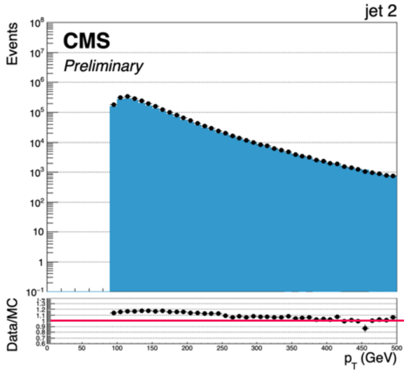
Validation plots for HLT p_T SFs

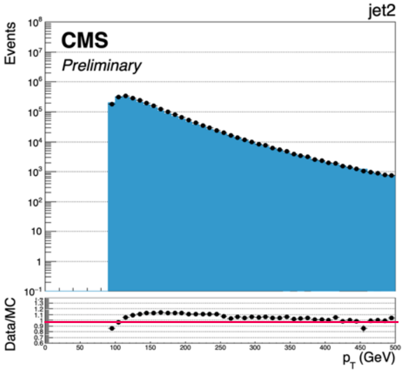
Before SF | After SF | Before SF | After SF

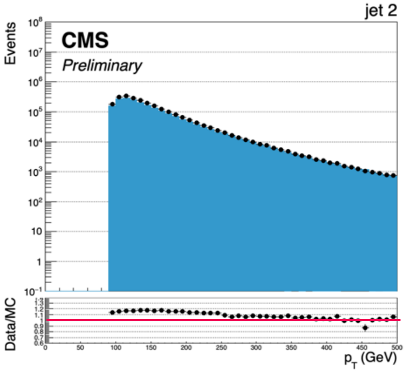


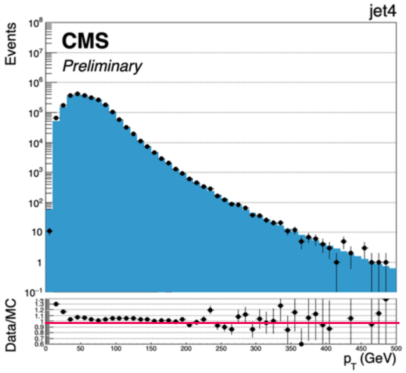


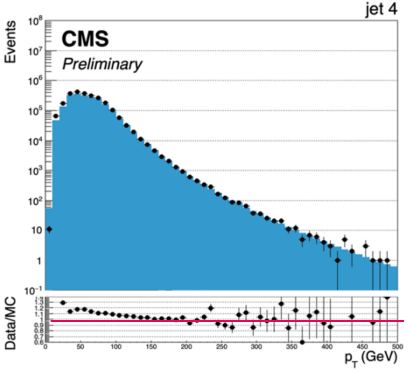










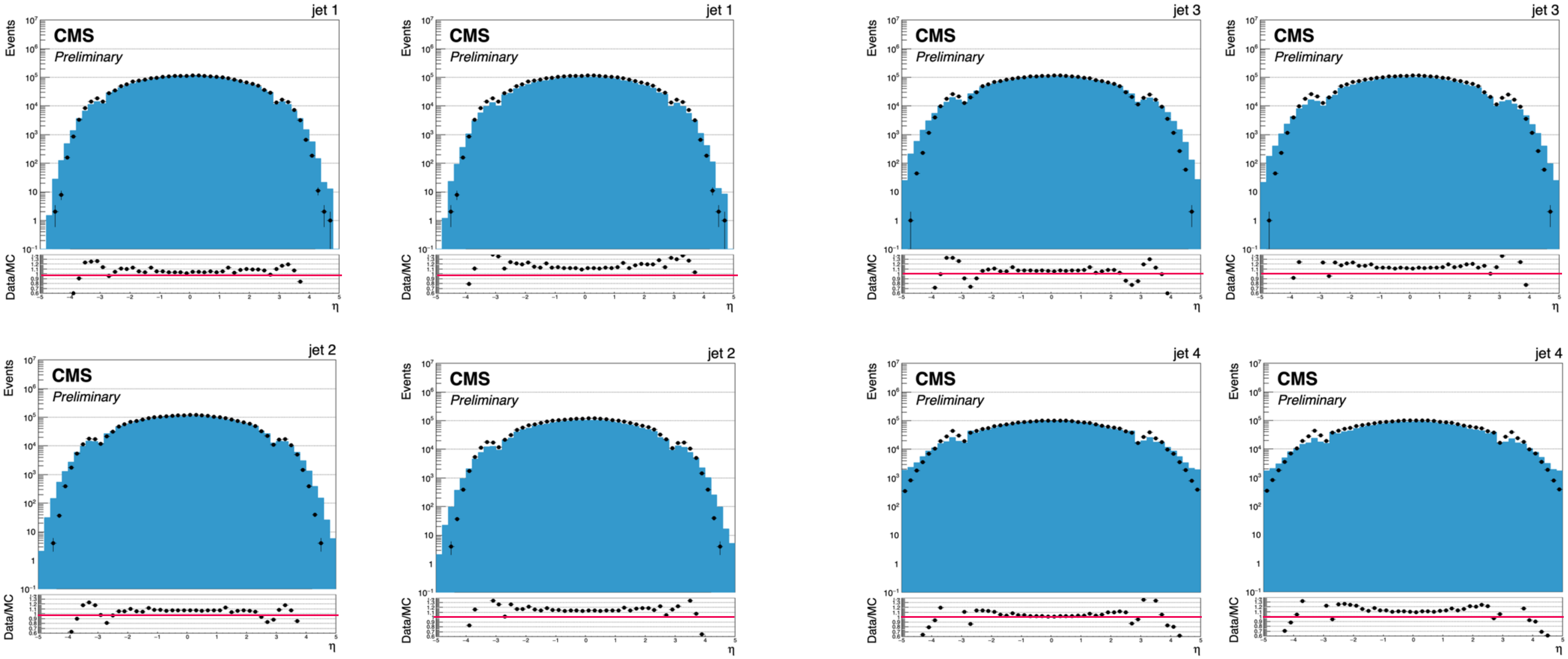


Events selected with
Control HLT path: equal to the signal path, but does not include VBF and c-tagging sequences

JEC applied

Validation plots for HLT p_T SFs

Before SF | After SF | Before SF | After SF



Events selected with
Control HLT path: equal to the signal path, but does not include VBF and c-tagging sequences

JEC applied

Trigger c-tagging scale factors

The efficiency on the Pnet CvsAll trigger selection is evaluated by using the control path (no c-tag and VBF sequences)

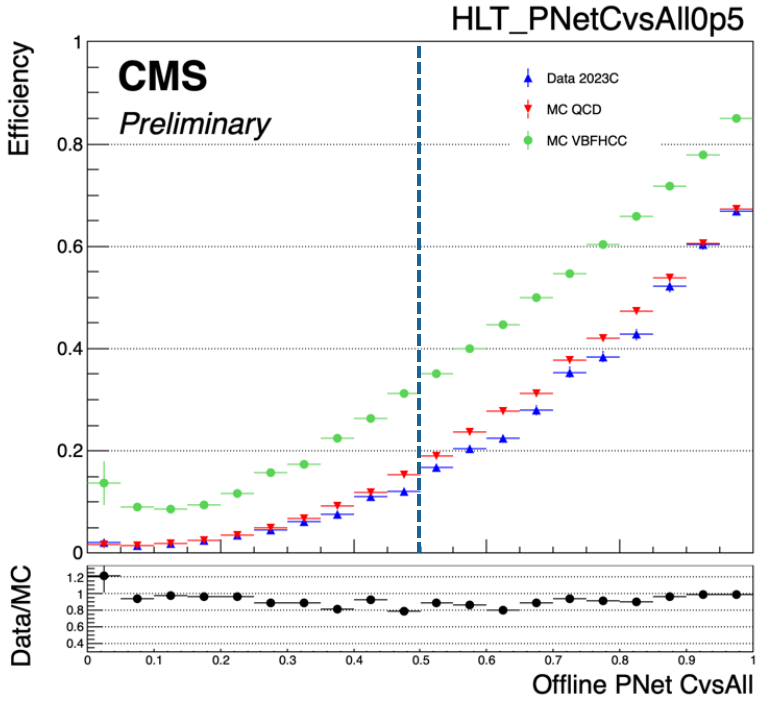
The following event selection is applied:

- at least 4 offline jets with p_T > 105, 90, 75, 35 GeV
- first three p_T leading offline jets matching with a HLT object with p_T > 100, 88, 70 GeV
- Among the 4 p_T-leading jets:
  - 2 jets with highest CvsAll score identified as c-jets
  - other 2 jets identified as VBF jets (qq), with
    $\Delta\eta$(qq) > 3.8
    Mass(qq) > 500 GeV

$$\text{Efficiency} = \frac{\text{PNet CvsAll offline score of the most c-tagged jet, } \mathbf{HLT{:}signal{+}control}}{\text{PNet CvsAll offline score of the most c-tagged jet, } \mathbf{HLT{:}control}}$$

Final SF = SF_{pT} x SF_{VBF} x SF_{ctag}

missing



Suboptimal Data/MC agreement