Status of H→bb and H→cc search with full Run 2

27-07-2023

Laura Buonincontri
Analysis group: X. Cid Vidal, L. Sestini, C. Vazquez Sierra, A. Gianelle, D. Zuliani, D. Lucchesi

Introduction

Goal of my thesis: “Inclusive H⟶bb and H⟶ cc searches at LHCb with the full Run 2 (5.4 fb^{-1})”

- DNN-based jet flavour tagging to separate b vs c vs (u,d,s,g)
  - Inputs are not only SV variables. Global quantities of the jet and features of charged and neutral particles in the jet are also considered
  - Three outputs: probabilities to be a b, c, or light jet
  - $P_b+P_c+P_q=1$
- Significance calculated in the invariant mass range [66,146] GeV (5.4 fb^{-1}) is 0.15
  - expected number of data: 31 M
  - expected number of Higgs in bb events: 840

$$S_b = \frac{N_{H\to bb}}{\sqrt{N_{data}}}$$

- Significance not enough to measure H⟶bb and H⟶cc cross section
  - limit on the production cross section will be set on the inclusive H⟶bb and H⟶ cc production cross sections and prospects for Yukhawa couplings at HL-LHC

Analysis Strategy

- Model of the bb (cc) QCD background for the H⟶bb (H⟶ cc) search:
    - Define a signal region (SR) and a control region (CR) on **data** by applying requirements on the DNN output:
        - Signal region, signal is expected: both jets in the final state are identified as b quarks;
        - Control region, mixed flavour in order to not have resonances: events that have the same kinematic characteristics as SR, but one quark is identified as b (c) quark, the other as light quark.
    - Transfer Function calculated from the Monte Carlo QCD bb or cc samples + correction function
- Fit to the di-jet invariant mass distribution
- Set of upper limits on inclusive production cross sections with CLs method
- Evaluation of the systematic uncertainties

Data and Monte Carlo samples

- Run II data-taking in the year 2016, center of mass energy 13 TeV
- Events used are those that pass the HLT2 lines: *HLTQEEJetsDiJetSVSV* and *HLTQEEJetsDiJetSV* (prescaled)
  - *HLTQEEJetsDiJetSVSV*: for both jets p_T>17 GeV, SV in jet cone
  - *HLTQEEJetsDiJetSV*: events with two reconstructed jets, at least one with SV
- Stripping lines required are *StrippingHLTQEEJetsDiJetSVSV* and *StrippingHLTQEEJetsDiJetSV* that store events where HLT2 lines are true. Tot integrated luminosity 1.6 fb^{-1} for both samples.
- Monte Carlo: H⟶bb, H⟶cc, Z⟶bb, Z⟶cc, QCD (bb,cc,qq)
- **Selection requirements**: P_T>20 GeV, 2.2 < η< 4.2, |Δφ| > 1.5, one of the two jets L0Chain TOS and HLT1Chain TOS

Monte Carlo requirements	H⟶bb	H⟶cc	Z⟶bb	Z⟶cc
Selection requirements+ **both jets SV-tagged**	11%	1.4%	14%	1.1%
Selection requirements+ **at least one jet SV-tagged**	30%	13%	34%	10%

- L0Chain TOS:
  L0HadronDecision_TOS || L0MuonDecision_TOS || L0PhotonDecision_TOS || L0DiMuonDecision_TOS || L0ElectronDecision_TOS || L0MuonEWDecision_TOS || L0JetPhotonDecision || L0JetPhotonDecision
- HLT1Chain TOS:
  Hlt1TrackMVADecision_TOS || Hlt1TwoTrackMVADecision_TOS || Hlt1TrackMuonDecision_TOS || Hlt1TrackMVATightDecision_TOS || Hlt1TwoTrackMVATightDecision_TOS ||Hlt1DiMuonHighMassDecision_TOS || Hlt1DiMuonLowMassDecision_TOS ||Hlt1SingleMuonHighPTDecision_TOS || Hlt1DiMuonNoL0Decision_TOS;

$H \rightarrow bb$: signal region

- Region where the signal is expected
- DNN P_b and P_q requirements tuned to maximize the significance in the SVSV data sample $S_b = \frac{N_{H\rightarrow bb}}{\sqrt{N_{data}}}$

$$N_{sign} = \mathcal{L} \cdot \sigma_{th} \cdot A \cdot \epsilon$$

$$\mathcal{L} = 1644.16 \text{ pb}^{-1}$$

ϵ includes the selection efficiencies, the cut on the DNN probability

- Maximum significance reached when both jets have $P_b>0.25$

Requirements	$P_b>0.25$ on both jets
$N_{H\rightarrow bb}$	257
N_{Data}	2.1e+07
Purity ($N_{H\rightarrow bb}/N_{Data}$)	1.3e-05
Significance	0.057

Purity and significance calculated in the entire mass spectrum

$H \rightarrow bb$: control region

- The goal is to have a data sample with enough data statistics while minimizing the number of signal events
- This search is done both in SVSV and in SVjet samples
- Different criteria have been studied to choose the best working point:
  - Significance minimization
  - Purity minimization $N_{H\rightarrow bb}/N_{Data}$
  - **$F=N_{H\rightarrow bb}/(N_{Data})^{3/2}$ minimization** (the idea was to increase the data statistics while keeping purity close to the minimum)
- Statistics higher in the SVSV sample, but SV+jet sample offers higher purity (less signal contamination in the CR)

Working point SR	P_b>0.25 on both jets
N_{Data}	2.1e+07
Purity	1.3e-05
Significance	0.057

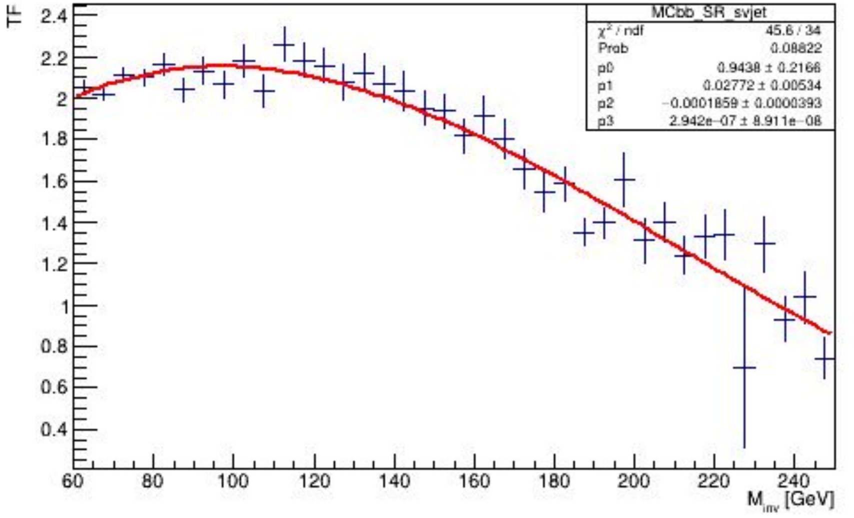
Working Point **SVSV**, CR
P_b>0.25 on one jet P_b< 0.25 on the other jet
N data: 3.23e+06
Significance: 0.009
Purity: 5e-06
F: 2.7e-09

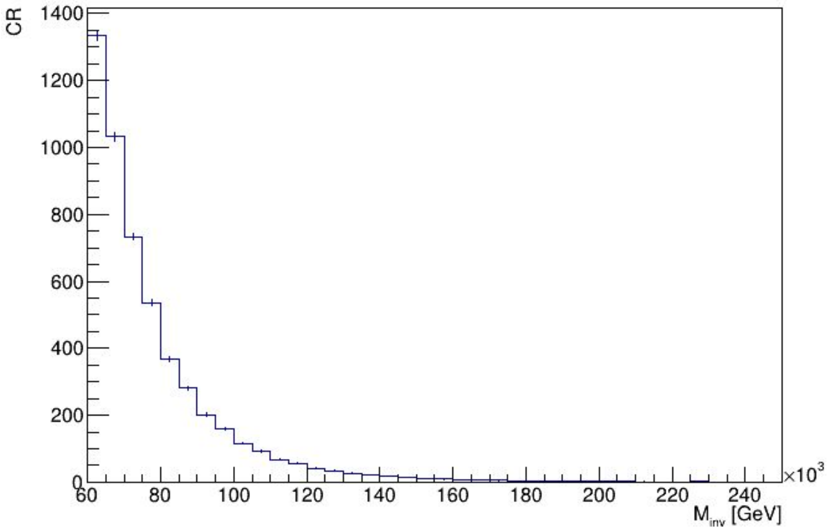
Working point **SV+jet**, CR
P_b>0.25 on the jet with SV, P_b< 0.22 on the other jet
N data: 1.1e+06 (with pre-scaling)
Significance: 0.02
Purity: 2.2e-06
F: 2.1e-10

Transfer Function fit Range [60,250] GeV

- QCD bb background sample used to calculate the Transfer Function as a function of the dijet invariant mass

$$\mathrm{TF} = \frac{n_{events,SR}}{n_{events,CR}}$$

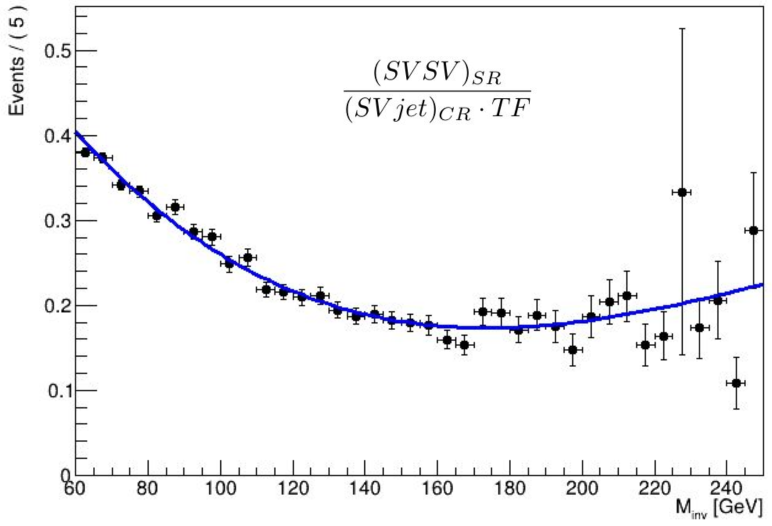




H →bb: full fit model

- Model to fit data in SVSV Signal Region, range [60,250] GeV:
  - H→bb, Z→bb, H→cc, Z→cc distributions from Monte Carlo
  - QCD background modeled as data in CR x TF (with CR and TF built from with SVjet samples)
  - Correction function (shape: Bernstein function)
- Determination of the proper number of coefficients of the Bernstein function: F-test
  - decide whether or not to reject the lower polynomial degree in favor of the higher polynomial degree
  - Bernstein polynomial degree: 4

- Correction function initial parameters: calculated by fitting a small portion of data
- Parameters of the correction function left free to vary in the fit to data
- N_{data} and N_{Hbb} left free in the fit to data

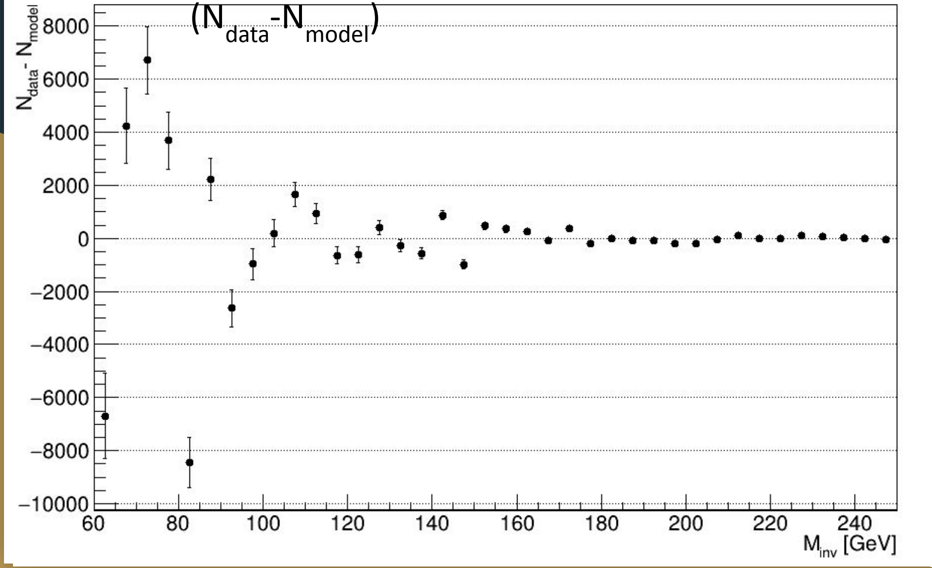


Fit to data results: H to bb

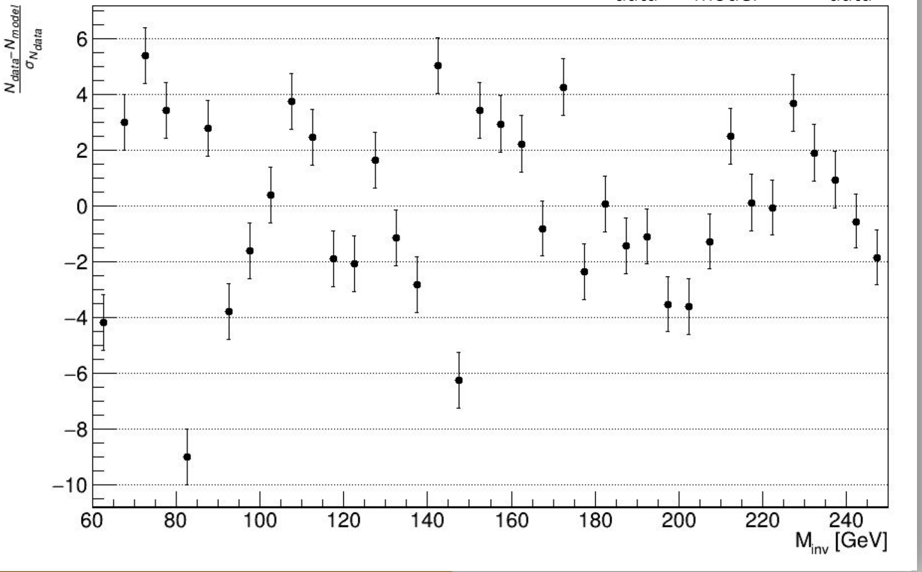
- Results on signal and background left blind

```
Error on Higgs: 3257.03
x0: 6.59769 +/- 0.573768
x1: 1.49541 +/- 0.132015
x2: 2.84324 +/- 0.250394
x3: 3.14267 +/- 0.279535
```

Residual: $(N_{data}-N_{model})$

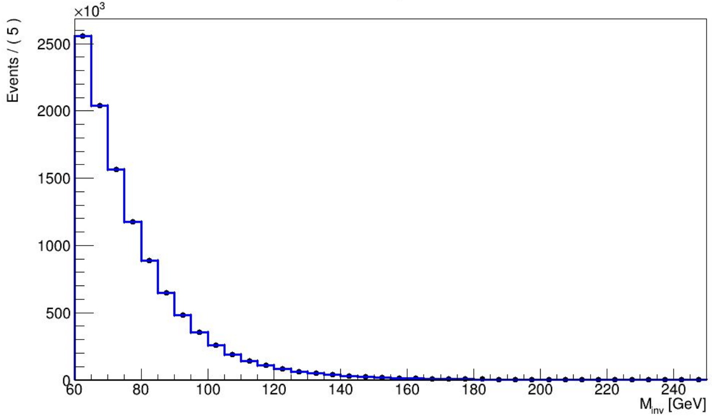


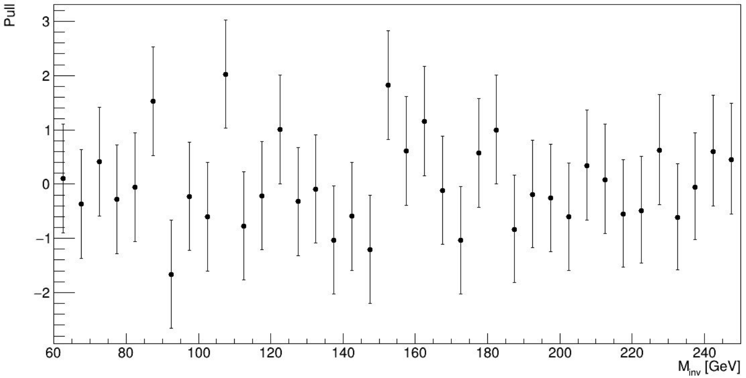
Pull: $(N_{data}-N_{model})/\sqrt{N_{data}}$

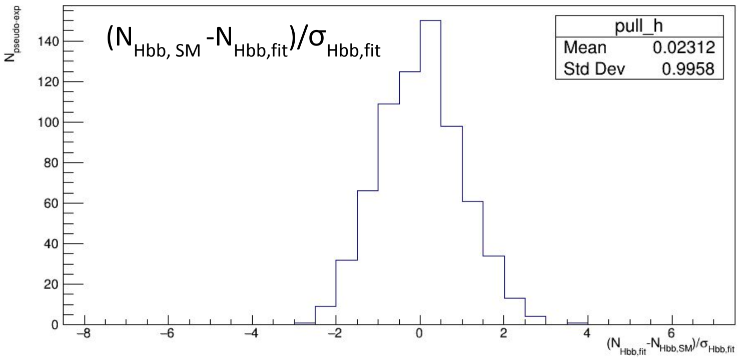


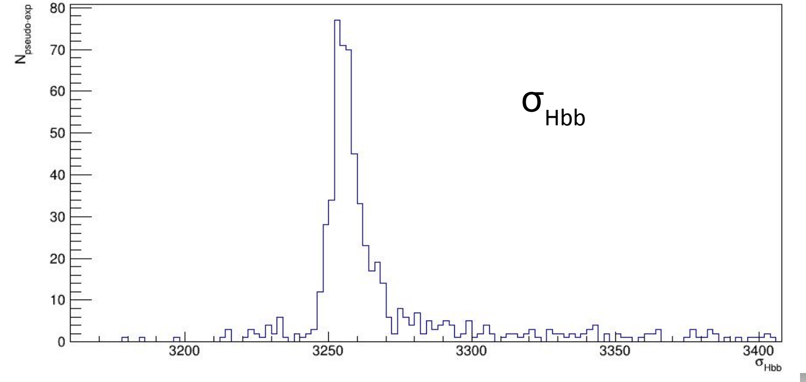
Monte Carlo fit validation

- Monte Carlo pseudo-experiment done with the Bern-4 correction function
- Initial values taken from the fit to data and then left free to vary in the fit

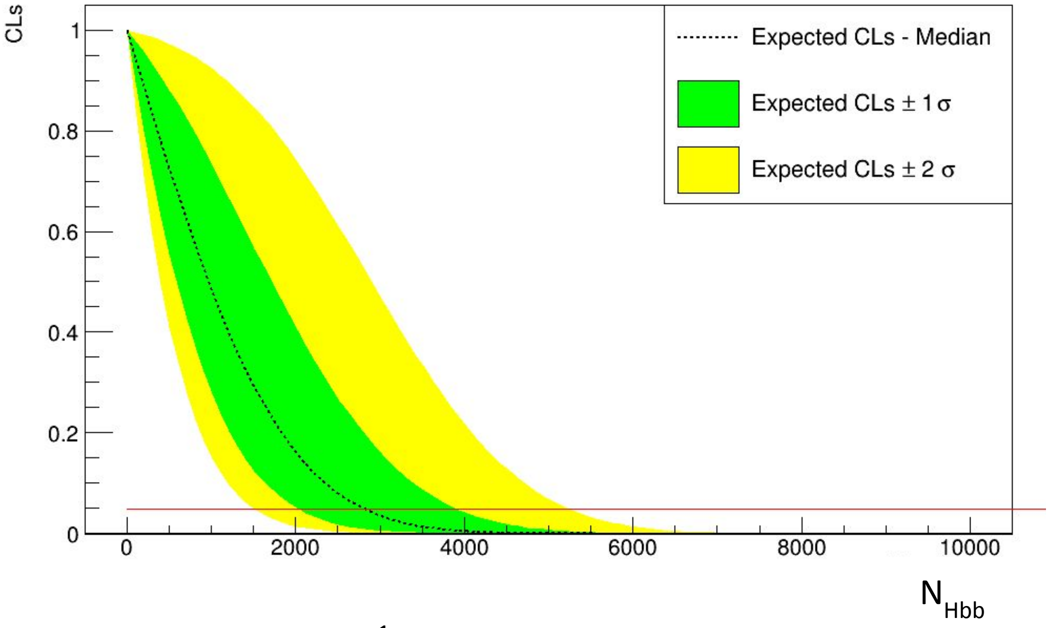








Expected Upper Limit on H to bb



Expected limit at 95% C.L.: 2807 in 1.6 fb^{-1} -> 12 x SM

Expected limit at 95% C.L. in 5.4 fb^{-1} (full Run 2) -> 6 x SM

Expected limit at 95% C.L. in 50 fb^{-1} (full Run 3+4) -> 2 x SM

Expected limit at 95% C.L. in 300 fb^{-1} (HL-LHC) -> 0.9 x SM

Signal and Control region H to cc

- Run II data-taking in the year 2016, center of mass energy 13 TeV, tot integrated luminosity 1.6 fb^{-1}
- Trigger: *HLTQEEJetsDiJetSVSV* and *HLTQEEJetsDiJetSV* (prescaled), *StrippingHLTQEEJetsDiJetSVSV* and *StrippingHLTQEEJetsDiJetSV*
- Monte Carlo: H→bb, H→cc, Z→bb, Z→cc, QCD (bb,cc,qq)
- **Selection requirements**: P_T>20 GeV, 2.2 < η< 4.2, |Δφ| > 1.5, one of the two jets L0Chain TOS and HLT1Chain TOS

- **SIGNAL REGION**: DNN P_c and P_q requirements tuned to maximize the significance in the SVSV data sample
- Maximum significance: both jets have P_c>0.15

Signal region	**P_c>0.15** on both jets
$N_{H\to cc}$	1.8 (entire mass spect)
N_{Data}	4 M (entire mass spect)
Purity ($N_{H\to cc}/N_{Data}$)	5.0e-07 (entire mass spect)
Significance	0.0008 (entire mass spect)

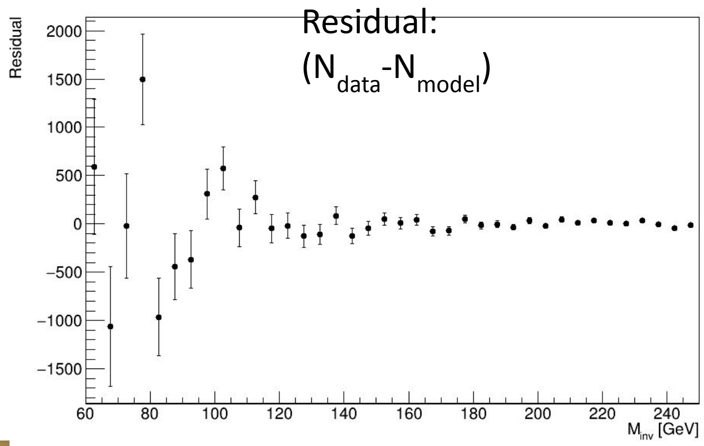
- **CONTROL REGION:** the goal is to have a data sample with enough data statistics while minimizing the number of signal events
- This search is done both in SVSV and in SVjet samples
- Low data statistics in the SV+jet sample, the SVSV is used

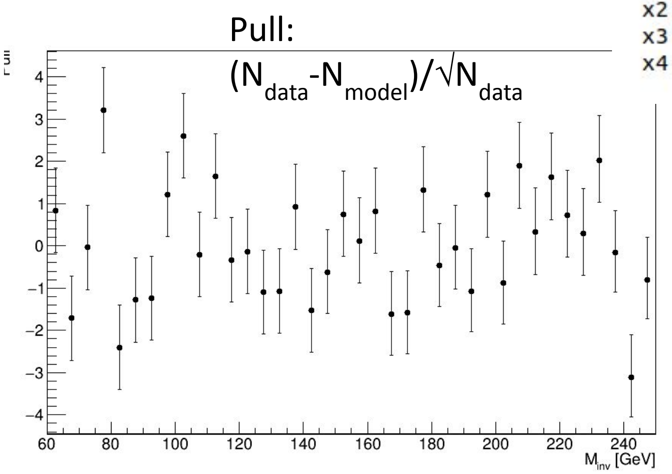
Control region	Minimum Signif. and purity
Cuts on P_c	**P_c>0.15 on one jet** **P_c< 0.02 on the other jet**
N data	4.5e+06
Significance	2.4e-05
Purity	1.1e-08

Fit to the SVSV signal region

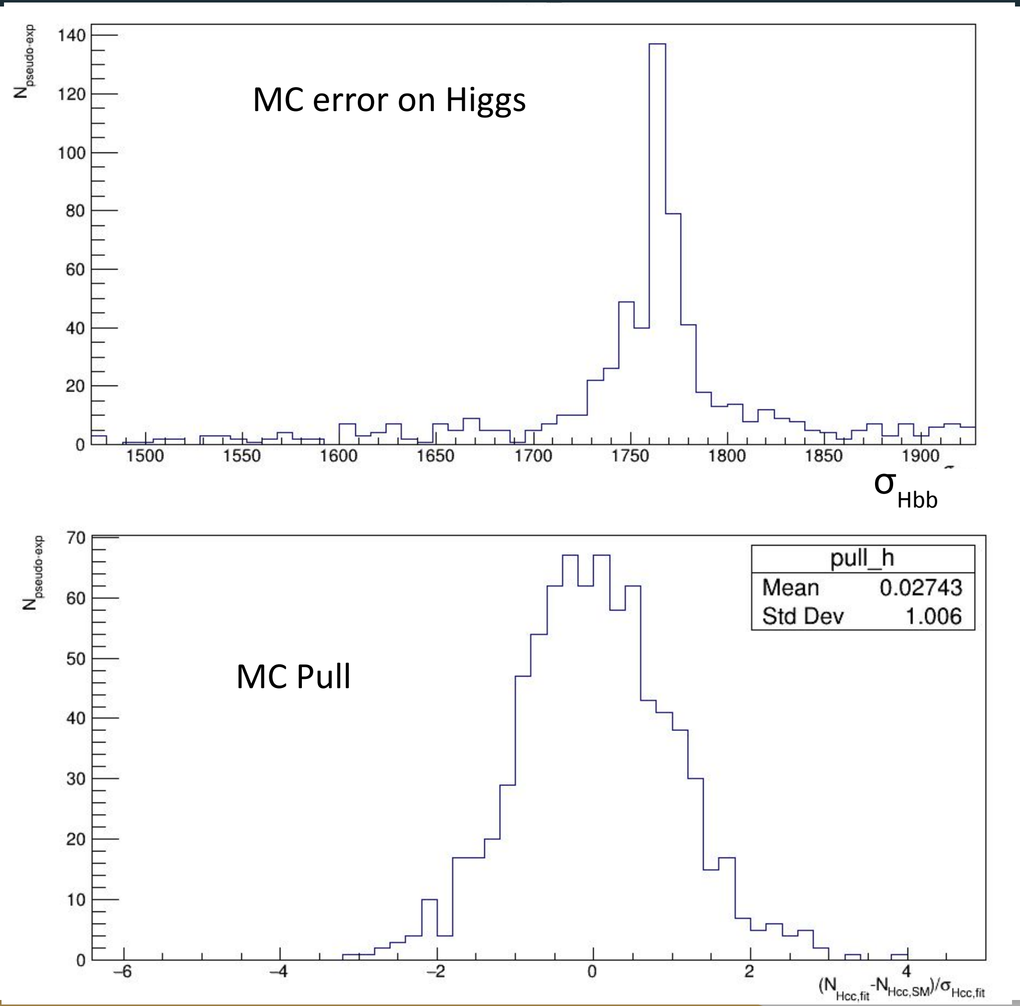
- Model to fit invariant mass distribution of data selected *HLTQEEJetsDiJetSVSV* Signal Region:
  - $H \to bb$, $Z \to bb$, $H \to cc$, $Z \to cc$ distributions from Monte Carlo
  - QCD background modeled as:
    - data SVSV in CR x Transfer Function (TF) determined using invariant mass distribution of Monte Carlo QCD bb and cc samples
    - Correction function (shape: Bernstein polynomial 5 coefficients)
- Correction function initial parameters: calculated by fitting a small portion of data and are left free in the fit to the full dataset
- N_{data} and N_{Hcc} left free in the fit to data

```
Error on Higgs to cc: 1798.69
x0: 4.98278 +/- 0.574746
x1: 4.72565 +/- 0.554297
x2: 8.11861 +/- 0.970454
x3: 7.00158 +/- 0.924254
x4: 6.85781 +/- 0.817415
```

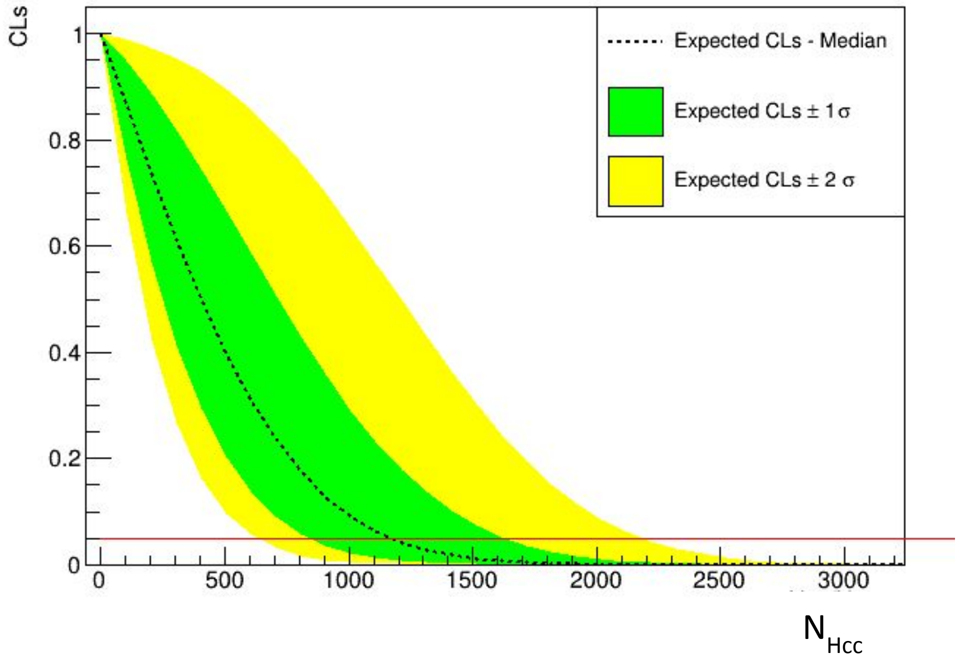




Monte Carlo fit validation H to cc



Expected Upper Limit on H to cc



L=1.6fb^{-1}: 1174 -> 665x SM

Scaling with luminosity:

L=5.6 fb^{-1} -> 356 x SM

L=50 fb^{-1} -> 116 x SM

L=300 fb^{-1} -> 47 x SM

Improvements to the H to cc significance

Upper limit roughly scales with the inverse of the Higgs significance S

- How much the significance increases without requiring SV, only DNN
  - Trigger line and stripping line with two jets required (Hlt2: Hlt2JetsDiJet and Stripping: FullDiJetsLine, total prescale: 1.3e-05)
  - Same selection requirements of the analysis: P_T>20 GeV, 2.2 < η< 4.2, |Δφ| > 1.5, one of the two jets L0Chain TOS and HLT1Chain TOS
  - DNN P_c and P_q optimized to maximize the significance

S (dijets): 0.0011

S (SVSV): 0.0008

Factor: 1.4

- C.o.m. will pass from 13 TeV to 14 TeV: $f = \frac{S(14TeV)}{S(13TeV)} = \frac{N_H(14TeV)}{N_H(13TeV)}\sqrt{\frac{N_{bkg}(13TeV)}{N_{bkg}(14TeV)}}$

$$\frac{N_{bb}(14TeV)}{N_{bb}(13TeV)} = \frac{(\sigma_{bb}\cdot A)(14TeV)}{(\sigma_{bb}\cdot A)(13TeV)} = 1.17+/-0.06$$

$$\frac{N_{qq}(14TeV)}{N_{qq}(13TeV)} = \frac{(\sigma_{qq}\cdot A)(14TeV)}{(\sigma_{qq}\cdot A)(13TeV)} = 1.21+/-0.13$$

$$\frac{N_{H}(14TeV)}{N_{H}(13TeV)} = \frac{(\sigma_{H}\cdot A)(14TeV)}{(\sigma_{H}\cdot A)(13TeV)} = 1.13+/-0.08$$

Factor~ 1

- Scaling factors for mixed flavour background are compatible with the bb and qq samples

Prospects on H to cc

*LHCb-CONF-2016-006
**https://agenda.infn.it/getFile.py/access?contribId=36&sessionId=4&resId=0&materialId=slides&confId=12253

- Final prospect for the upper limit on the inclusive Higgs cross section at HL-LHC, taking into account DNN improvement:

$$\frac{BR(H \to c\bar{c})}{BR_{SM}(H \to c\bar{c})} < 33$$

- V + H_0 (⟶ cc) limit: 6400xSM at 8 TeV in (*)
- From (**) limit recast for the H to cc cross section with the Higgs produced in association with a W/Z
  - V + H_0 (⟶ cc) limit: **50xSM** without any improvement in the analysis and the detector at HL-LHC
  - V + H_0 (⟶ cc) limit: **13xSM** with improvement in the c-tagging (best IP resolution) at HL-LHC
  - V + H_0 (⟶ cc) limit: **5-10xSM** with improvement in the c tagging and analysis at HL-LHC

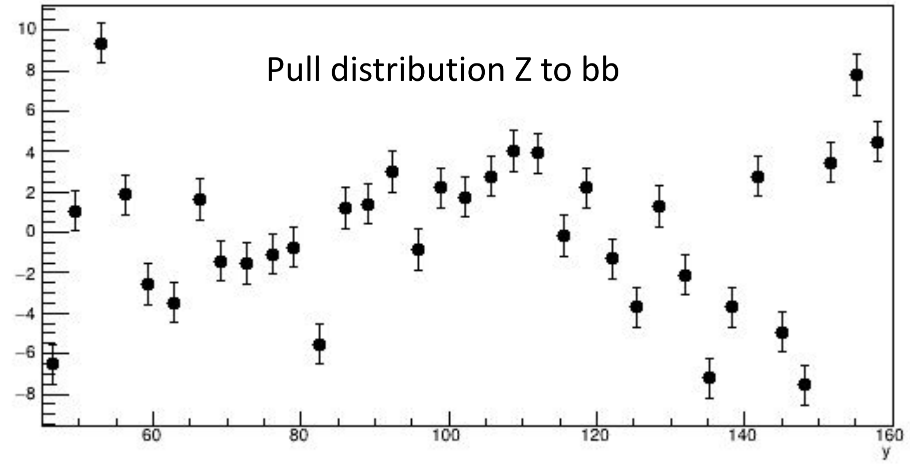
- With the improvement on c-tagging applied to the inclusive production, the 33xSM limit could become **~8.6xSM**

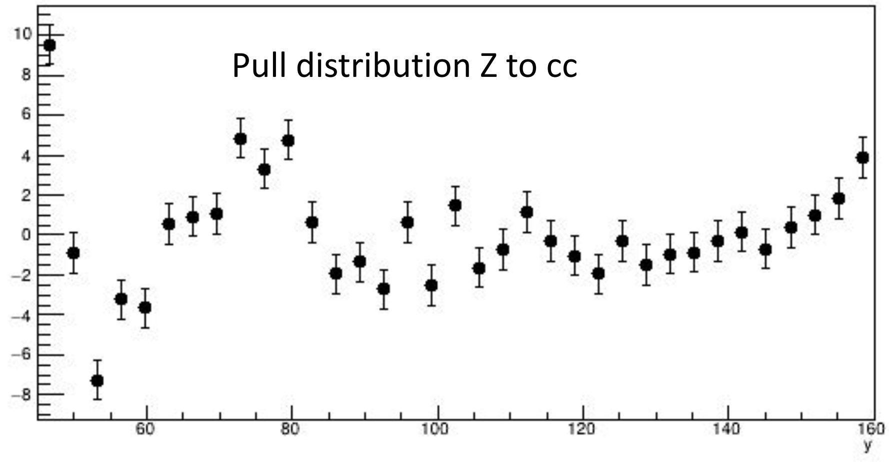
Z to bb and Z to cc fit

- Simultaneous fit of the Z to bb and Z to cc in the two signal regions, mass range [45,160] GeV:
  - One common signal strength parameter between the two fits for the Z to bb and one for the Z to cc
  - Number of data left free in the fit
  - The correction functions : Bernstein function with 4 coefficients for the cc and Bernstein function with 6 coefficients for bb

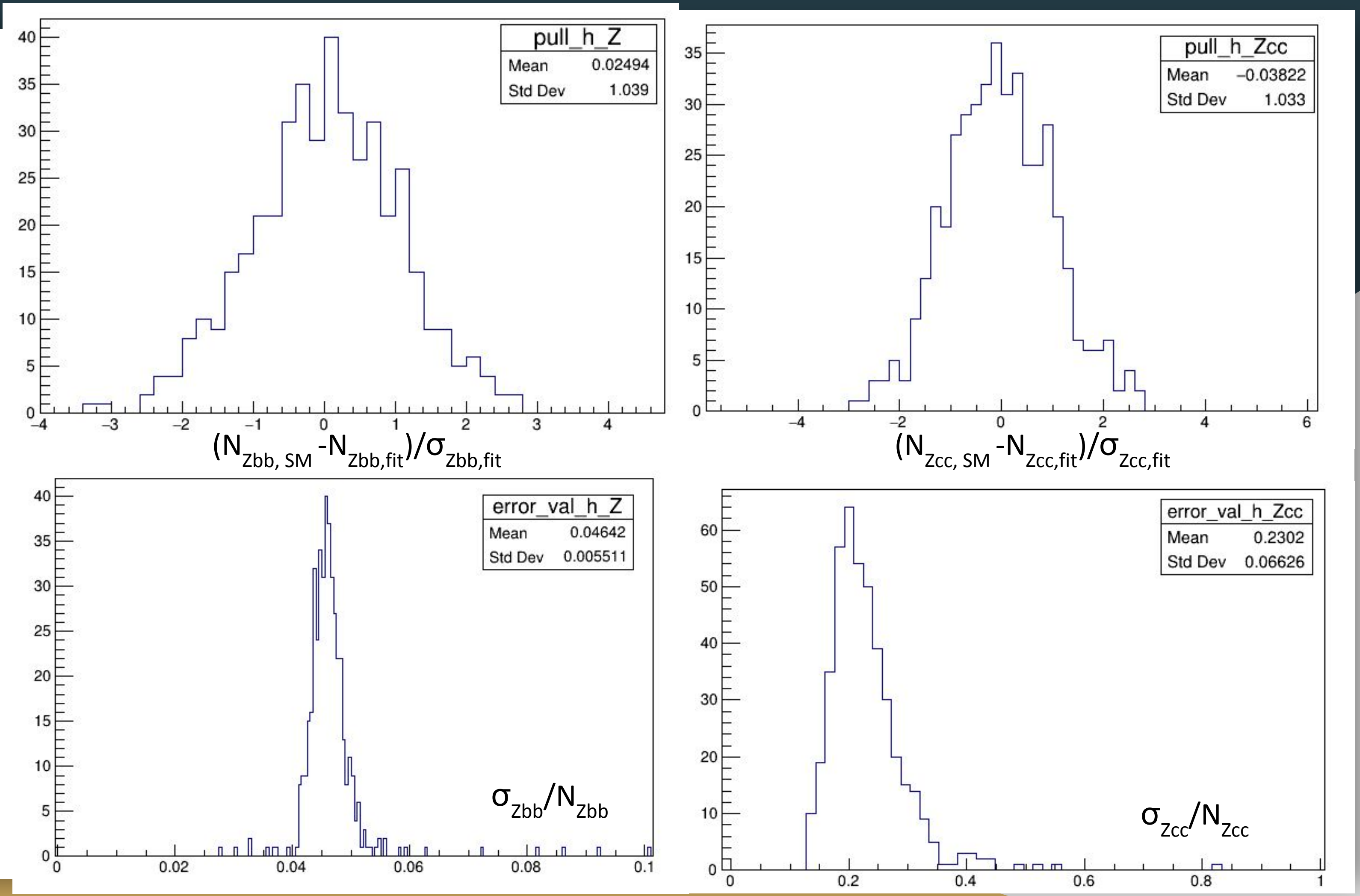
Z to bb signal strength:
6.5e-01 +/- 4.74e-02

Z to cc signal strength:
1.6e+00 +/- 2.26e-01





Monte Carlo



pull_h_Z
Mean 0.02494
Std Dev 1.039
$(N_{Zbb, SM} - N_{Zbb,fit})/\sigma_{Zbb,fit}$
pull_h_Zcc
Mean −0.03822
Std Dev 1.033
$(N_{Zcc, SM} - N_{Zcc,fit})/\sigma_{Zcc,fit}$
error_val_h_Z
Mean 0.04642
Std Dev 0.005511
σ_{Zbb}/N_{Zbb}
error_val_h_Zcc
Mean 0.2302
Std Dev 0.06626
σ_{Zcc}/N_{Zcc}

Current focus

- Calculation of systematic uncertainties started:
  - Systematic on correction function
    - to be added in limit calculation
  - Systematic on SV tagging:
    - Effects on limit to be determined

H to cc:

Modelling the correction with Bernstein poly, 5 coeff:
$L=1.6fb^{-1}$: 1174 -> 667x SM

Modelling the correction with Bernstein poly, 6 coeff
$L=1.6fb^{-1}$: 1171.9 -> 666x SM

	N Z in cc	N Z in bb	N H in cc	N H in bb
No Corr	15338	191391	2,1	265
Corr+1 sigma	18404	236841	2,2	353
Corr	15263	204406	1,7	302
Corr-1 sigma	12416	174431	1,3	258

jet p_T (GeV)	jet η	b jets	c jets
10–20	2.2–4.2	0.89 ± 0.04	0.81 ± 0.09
20–30	2.2–4.2	0.92 ± 0.07	0.97 ± 0.09
30–50	2.2–4.2	1.06 ± 0.08	1.04 ± 0.09
50–100	2.2–4.2	1.10 ± 0.09	0.81 ± 0.15

Next steps

- Complete the systematics and efficiency correction calculation:
  - Strategy defined for the calculation of some systematics: Control Region statistics, signal contamination in CR, DNN efficiency
  - For the others I will follow the same strategy that are explained in the analysis note “Measurement of differential $b\bar{b}$ and $c\bar{c}$ cross sections in the forward region of pp collisions at $\sqrt{s}$ = 13 TeV”: Jet Energy Correction, Jet Energy Scale, Jet Identification, Trigger…

TIMELINE

- by the end of August:
  - complete the systematics (upload results on the H to cc paper from time to time)
  - write the introduction chapters (detector, Higgs, event reconstruction)
- since September: apply to Z to bb and Z to cc fit, write the analysis

Contents

1 Introduction

2 Theoretical introduction: Higgs boson physics
- 2.1 EW theory/Higgs mechanism
  - 2.1.1 Yukawa couplings
- 2.2 Higgs boson physics at LHC and Future colliders
  - 2.2.1 Production mechanism at hadron collider and lepton colliders
  - 2.2.2 Higgs decay modes
  - 2.2.3 Double and triple Higgs production at hadron and muon colliders
- 2.3 Higgs coupling measurements
  - 2.3.1 Current experimental results on Higgs couplings (LHC)
  - 2.3.2 Prospects on Higgs couplings measurements at HL-LHC
  - 2.3.3 Prospects on Higgs couplings measurements at muon collider

3 LHCb detector description
- 3.1 LHC overview
- 3.2 Detector overview
  - 3.2.1 Vertex Locator
  - 3.2.2 Magnet
  - 3.2.3 Tracking system
  - 3.2.4 RICH
  - 3.2.5 Calorimeters
  - 3.2.6 Muon systems
- 3.3 Trigger
  - 3.3.1 L0 trigger
  - 3.3.2 HLT
- 3.4 Tracking performances

4 Jets reconstruction, flavour tagging and event selection at LHCb
- 4.1 Simulations
- 4.2 Jets reconstruction algorithms
  - 4.2.1 Particle flow
  - 4.2.2 Jet clustering with anti kt algorithm
  - 4.2.3 Jet Energy correction
- 4.3 Flavour tagging algorithm
  - 4.3.1 SV tagging algorithm
  - 4.3.2 Tagging Performance
- 4.4 DNN for flavour tagging
  - 4.4.1 DNN scheme
  - 4.4.2 Application of DNN to jets

5 Search for H to bb and H to cc at LHCb
- 5.1 Introduction
- 5.2 Monte Carlo and data samples
  - 5.2.1 Event selection and yield prediction
- 5.3 Analysis strategy
  - 5.3.1 Background determination
  - 5.3.2 Definition of signal and control region
  - 5.3.3 Definition of the transfer function
  - 5.3.4 Application of SR and CR to signal resonances
  - 5.3.5 CLs method

6 Upper limits on H to bb and H to cc cross sections at LHCb
6.1 Fit to the signal region
6.2 Systematic uncertainties
6.3 Upper limits results

7 Z to bb and Z to cc cross section determination
7.1 Event selection and yield prediction
7.2 Test of the signal extraction
7.3 Fit to the signal region
7.4 Systematic uncertainties
7.5 Results

8 The Muon collider machine and detector
8.1 Overview of the facility
8.2 The interaction region and machine-detector interface
8.3 Beam induced background
8.4 Muon Collider detector
8.5 Overview of the detector
8.6 Vertex detector and tracking system
8.6.1 Vertex detector
8.6.2 Tracker System
8.7 Calorimetry
8.8 Electromagnetic and hadronic calorimeter
8.9 Calorimeter system performance
8.10 Solenoid and Muon detector
8.10.1 Vertex detector and tracking system
8.10.2 Calorimeter system
8.10.3 Solenoid and muon detector
8.10.4 Effects of BIB on muon collider detector and mitigation strategies

9 Event reconstruction at Muon Collider
9.1 Track reconstruction
9.1.1 Tracking performance
9.2 Particle flow
9.3 Jet reconstruction
9.3.1 Jet reconstruction efficiency and resolution
9.4 b tagging algorithm

10 Higgs coupling to b and Higgs self-coupling at muon collider
10.1 Reconstruction of Higgs to bb at 3 TeV muon collider
10.2 Double Higgs reconstruction in 4b
10.2.1 Signal and background events generation
10.2.2 Event selection
10.3 Determination of the HH cross section precision
10.3.1 Study of kinematic properties
10.3.2 Multivariate analysis
10.3.3 Event classification
10.3.4 Determination of the HH cross section precision
10.4 Determination of the sensitivity on the Higgs self-coupling
10.4.1 Study of kinematic properties
10.4.2 Multivariate analysis
10.4.3 Event classification
10.4.4 Scan of the likelihood
10.4.5 Determination of the uncertainty on the trilinear Higgs self-coupling

11 Conclusions and Future Prospects
11.1 Prospects on H to cc at LHCb at HL-LHC

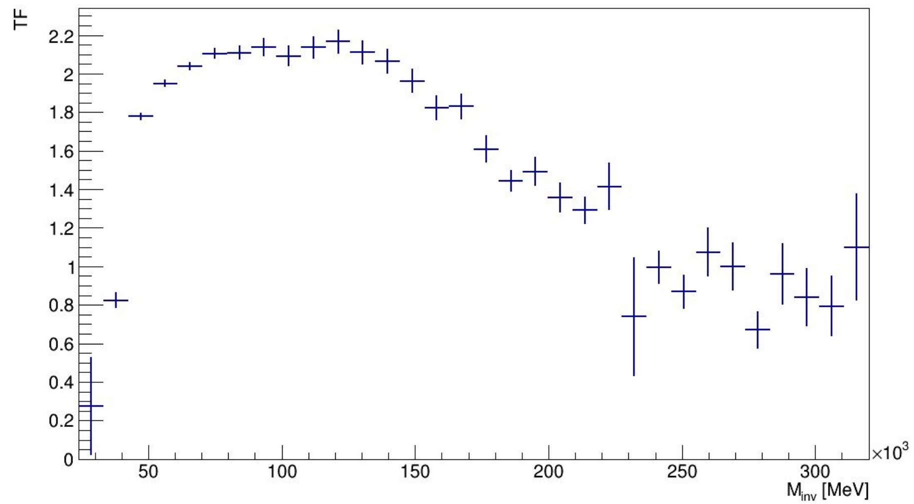
References

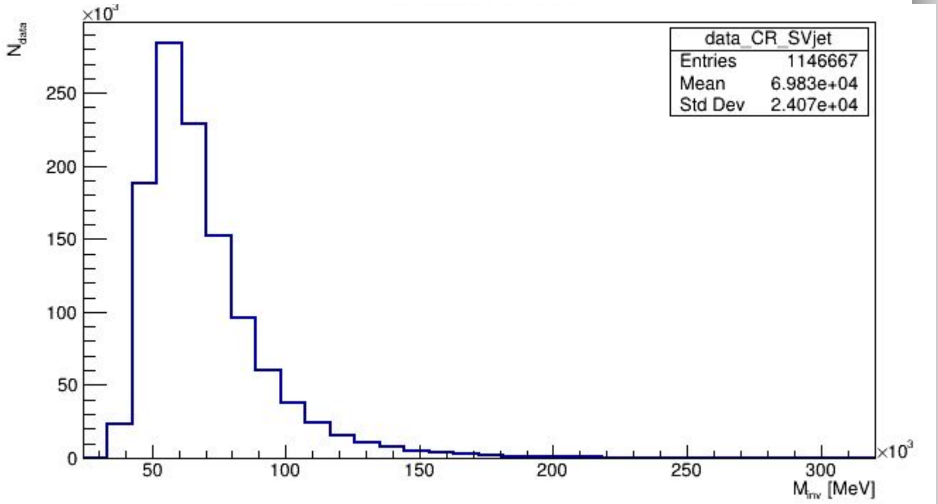
BACKUP

H →bb: transfer function SVjet

- QCD bb background sample used to calculate the Transfer Function as a function of the dijet invariant mass

$$\mathrm{TF} = \frac{n_{events,SR}}{n_{events,CR}}$$



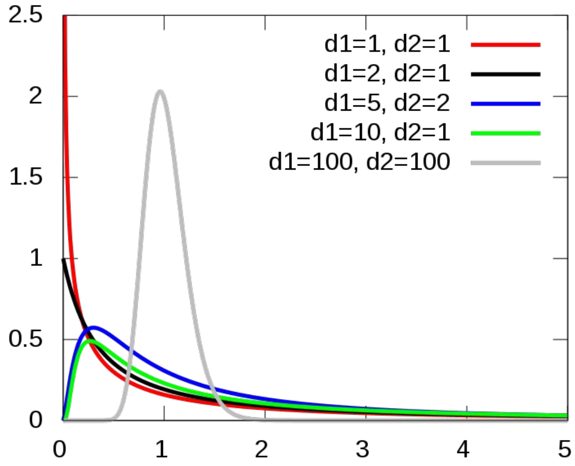


F-test

1. Define a larger **full model**. ("larger," with more parameters.)
2. Define a smaller **reduced model**. ("smaller," with fewer parameters.)
3. Use an ***F*-statistic** to decide whether or not to reject the smaller reduced model in favor of the larger full model
    a. Size of the test: alpha=0.05
    b. Null hypothesis: additional parameter is useless
    c. Null hypothesis always pertains to the reduced model, the alternative hypothesis always pertains to the full model.
    d. F is distributed as a Fisher-Snedecor distribution
    e. Calculate the F statistic from data ("Ftest") and p as 1-FDistI(Ftest,p2-p1,N-p2)
    f. reject the null hypothesis if p < alpha

$$F = \frac{\dfrac{\sum_i (y_i - f_1(x_i))^2 - \sum_i (y_i - f_2(x_i))^2}{p_2 - p_1}}{\dfrac{\sum_i (y_i - f_2(x_i))^2}{n - p_2}}$$

f1: reduced model
f2: full model
p1: N parameters in f1
p2: N parameters in f2
N: Nbins

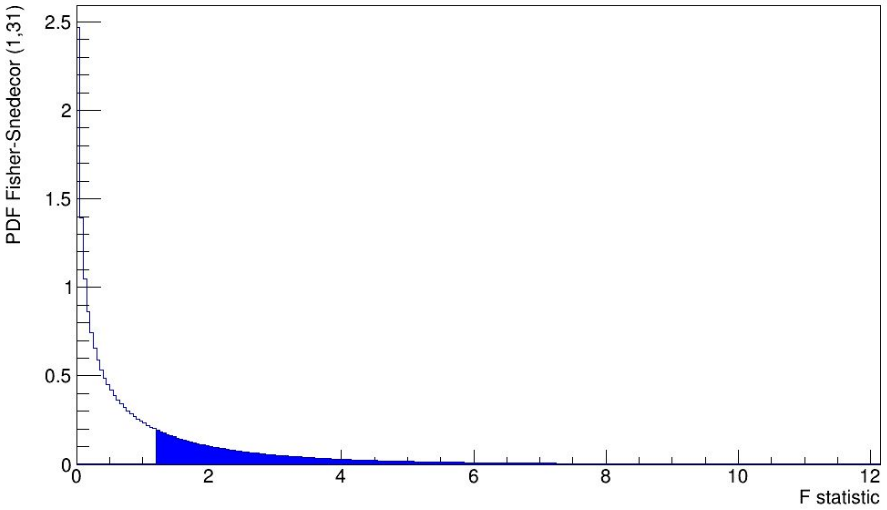


Fisher test H to bb

1-FDistI(Ftest,p2-p1,N-p2)

F value Bernst(degree 2 vs 3): 305
F value Bernst(degree 3 vs 4): 772
F value Bernst(degree 4 vs 5): 1.19

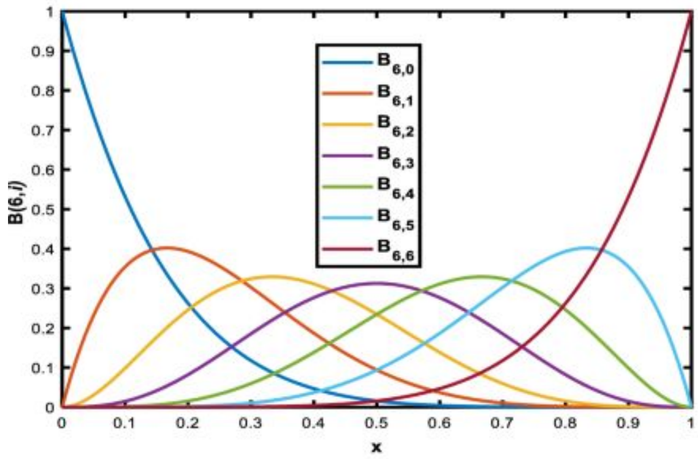
p value Bernst(degree 2 vs 3): 1.1e-16
p value Bernst(degree 3 vs 4): 1.1e-16
p value Bernst(degree 4 vs 5): 0.3



Bernstein polynomial

$$PDF(x, c_0, \ldots, c_n) = \mathcal{N} \cdot \sum_{i=0}^{n} c_i \cdot B_{i,n}(x).$$

$$B_{i,n}(x) = \binom{n}{i} x^i \cdot (1-x)^{n-i}$$



$$B_{6,0} = (1-x)^6$$

$$B_{6,1} = 6\,x(1-x)^5$$

$$B_{6,2} = 15\,x^2(1-x)^4$$

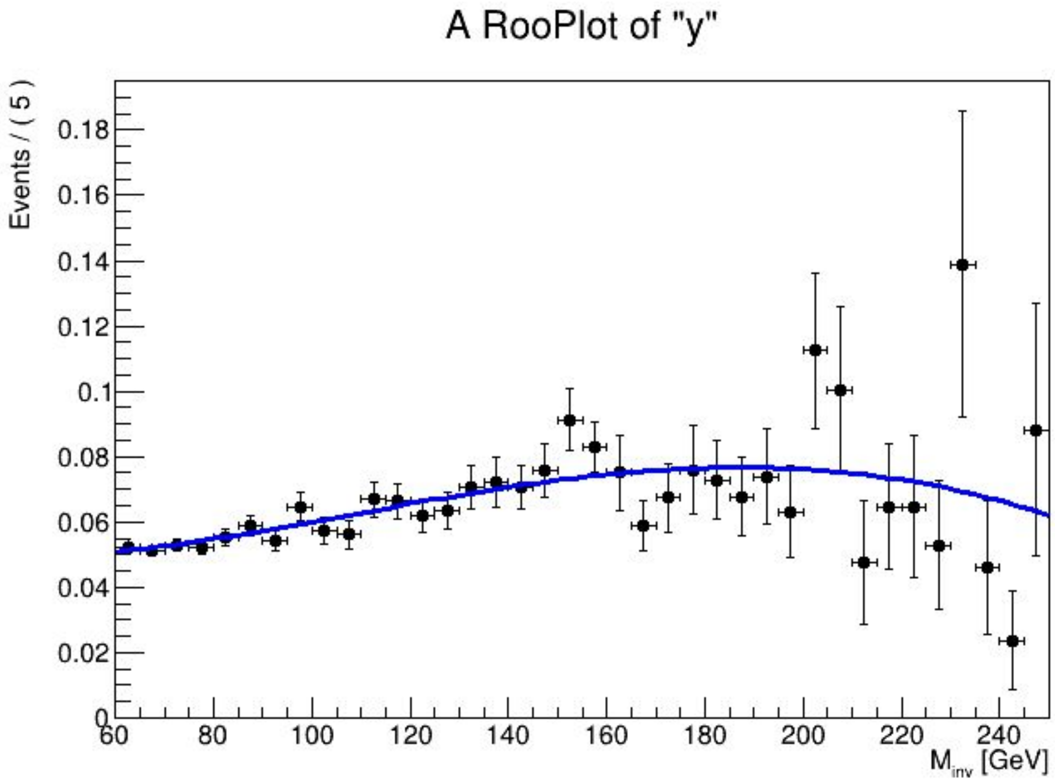
$$B_{6,3} = 20\,x^3(1-x)^3$$

$$B_{6,4} = 15\,x^4(1-x)^2$$

$$B_{6,5} = 6\,x^5(1-x)$$

$$B_{6,6} = x^6$$

Correction function H to cc

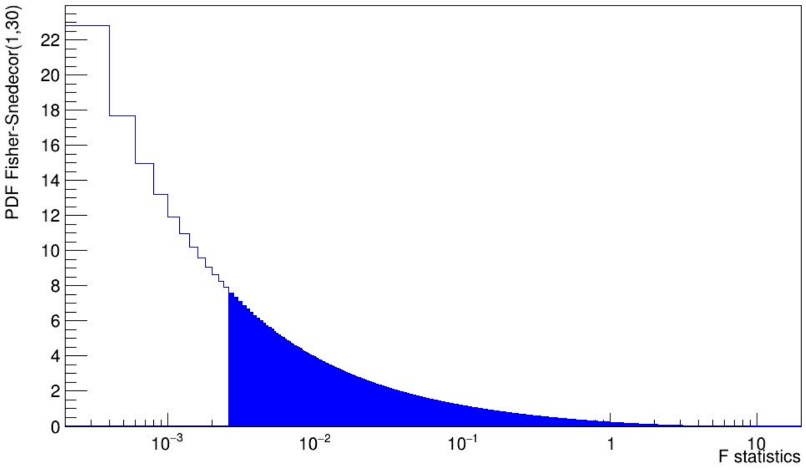


$$\frac{(SVSV)_{SR}}{(SVjet)_{CR} \cdot TF}$$

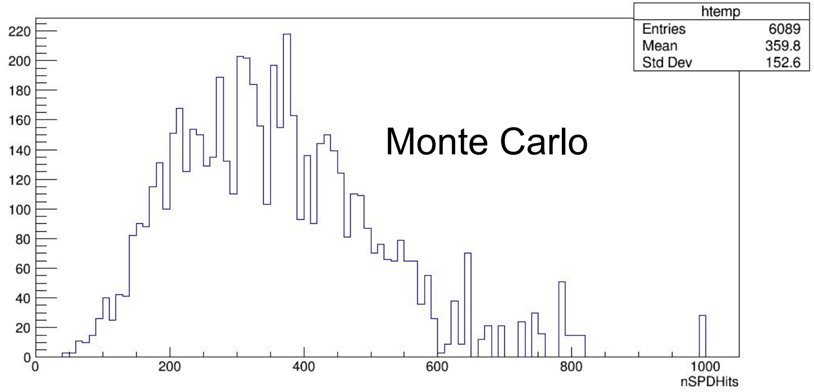
Fisher test H to cc

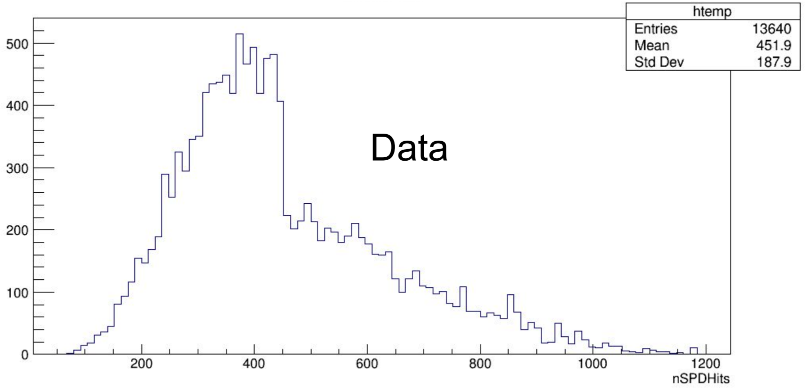
F value Bernst(degree 2 vs 3): 68.1678
F value Bernst(degree 3 vs 4): 52.7894
F value Bernst(degree 4 vs 5): 17.9335
F value Bernst(degree 5 vs 6): 0.00553859

p value Bernst(degree 2 vs 3): 1.55447e-09
p value Bernst(degree 3 vs 4): 2.95368e-08
p value Bernst(degree 4 vs 5): 0.00018944
p value Bernst(degree 5 vs 6): 0.941169



Global Event Cut





PT Higgs and Z