MSSM $bb(H \rightarrow bb)$ semileptonic

Antonio Branca, Ugo Gasparini, Tommaso Dorigo, Kostya Kanishev,
Stefano Lacaprara, Paolo Ronchese, Mia Tosi, Alberto Zucchetta

INFN Padova

Higgs Review,
CERN, 16 May 2012

Toward pre-approval

Timescale

- Thu 31/5 freeze AN and PAS
  - deadline 24:00, submit at 23:57.
- Fri 8/6 pre-approval presentation
- $\approx$ 16/6 unblind signal region
- 20-22/6 approval (?)
- 4/7 ICHEP

Intro: motivations

- Search for Neutral SUSY Higgs, $H \to b\bar{b}$;
- Large $BR(H \to b\bar{b}) \approx 90\%$
- huge multijet QCD background:
  - Use associate production to reject hadronic background
  - $pp \to b\bar{b}H \to b\bar{b}b\bar{b}$
- Compete with $H \to \tau\tau$ channel:
  - larger yield,
  - larger background (QCD),
  - different channel.

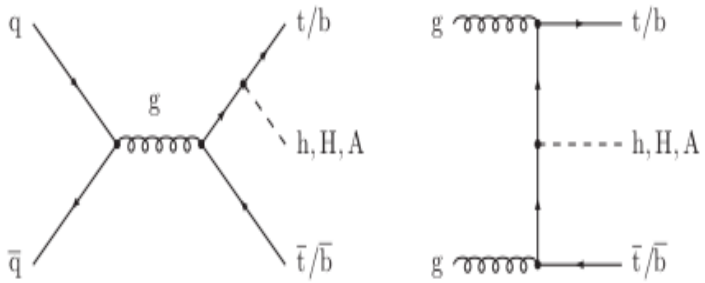
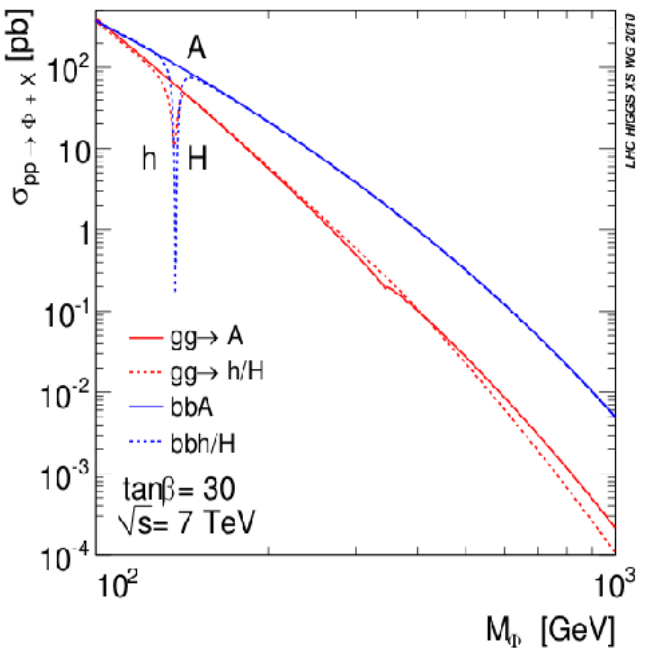


Fig. 17: Typical diagrams contributing to $q\bar{q}/gg \to Q\bar{Q} + h/H/A$ $(Q = t, b)$ at lowest order.

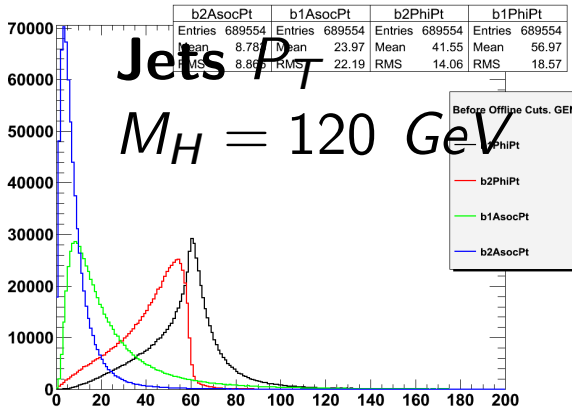


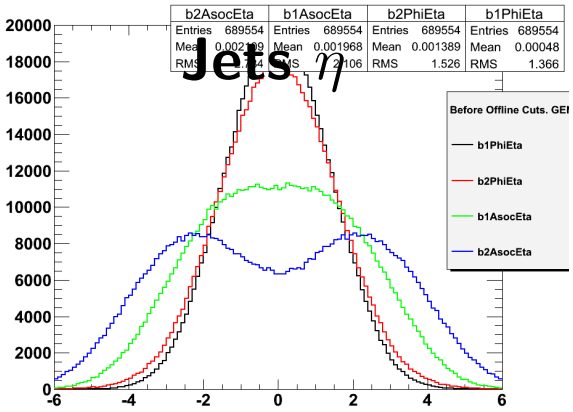
Analysis Strategy

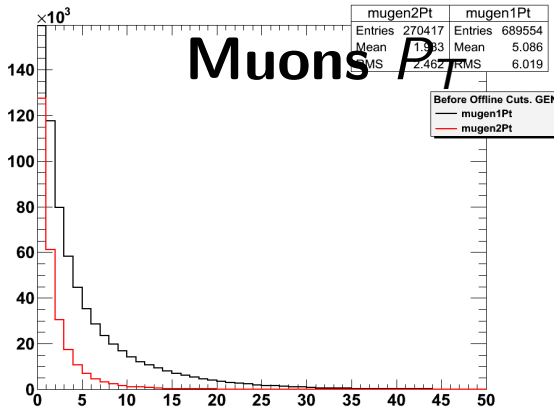
- Three b final state: $H \rightarrow bb$ plus additional associated b
- trigger is critical:
  - Use semi-leptonic (muon) b decay for trigger: muon+jets+b-tagging
- Major irreducible background source is multijet QCD
- Data driven background estimate from bbj sample
  - define signal-poor control sample;
  - B-tagging matrix method:
    - ★ get b/c-fraction of 3rd jet from mass & lifetime fits
    - ★ combine with MC b-tagging efficiency to derive B-tag probability
    - ★ weight bbj events to estimate number of 3-b-tags in signal region
  - Second approach with nearest-neighbour method (hyperball)
- Use reconstructed mass of leading jet pair as signal-sensitive variable in final fit
- Use only 2011 data

Trigger Strategy

- Use semi-leptonic b decay for trigger: **muon + jets + b-tagging**







Use different trigger Path in 2011 to cope with increasing $\mathcal{L}$

HLT paths	runs	triggers	$\int \mathcal{L}dt$ [pb^{-1}]
Mu12_CentralJet30_BtagIP	163738-165633	3 027 717	180.9
Mu12_DiCentralJet30_BtagIP3D	165970-166967	4 532 555	537.1
Mu12_DiCentralJet20_DiBtagIP3D1stTrack	167039-173198	2 244 550	1108.6
Mu12_eta2p1_DiCentralJet20_DiBtagIP3D1stTrack	173236-176469	1 237 147	652.2
Mu12_eta2p1_DiCentralJet20_DiBtagIP3D1stTrack*	176545-180252	5 690 304	2326.8
All		16 732 273	4805.7

Analysis selection

Selections

Baseline selections:

- Trigger
- at least 1 global muon $P_T^{\mu} > 15\ GeV$, no isolation required;
- at least 3 jets (PFak5, LooseId) $|\eta| < 2.6$, $P_T > (30, 30, 20)\ GeV$
  - $\Delta R_{ij} > 1$ for any pair $_{ij}$ of jets

bbj the first 2 jets must have b-tag $CSV > 0.8$

- the μ inside one of the two leading jets;

bbb last selection: third jet b-tag $CSV > 0.7$

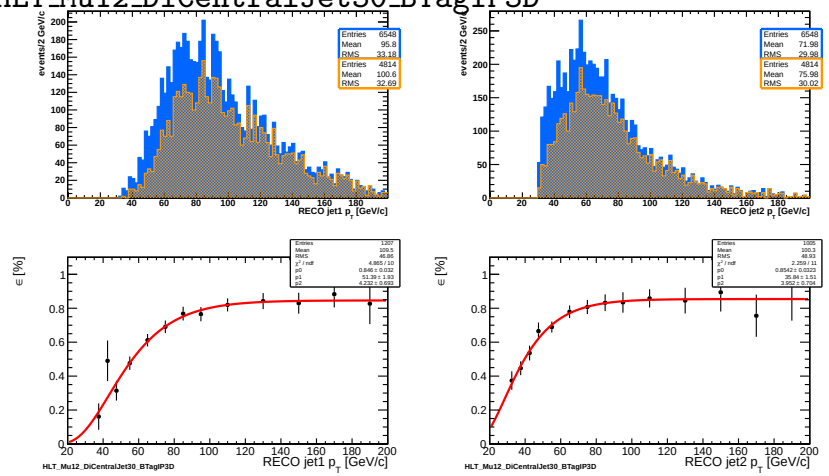
Trigger Efficiency vs Analysis

- $\epsilon_{trigger} = \epsilon(\mu) \times \epsilon(b - jets)$
- **SingleMu PD**, select all events passing a single muon path.
- Apply selection
- Build Turn on curves vs first and second B-jets Pt:

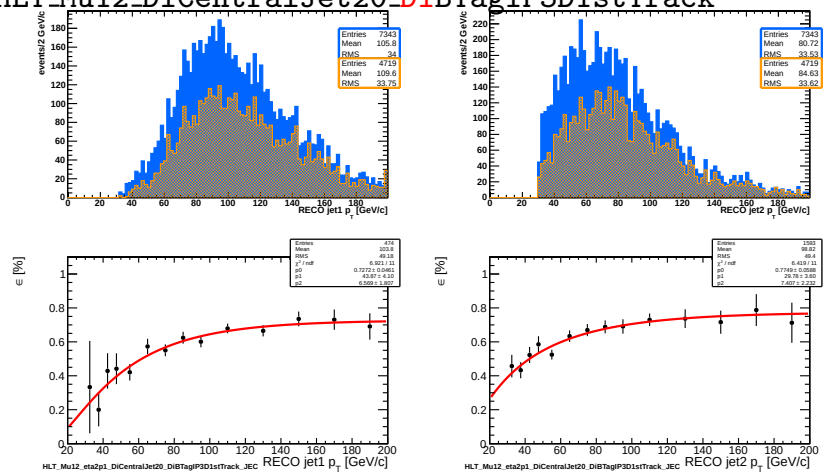
$$\epsilon = \frac{\text{Hbb path \& (pre)sel \& SingleMuHLT}}{\text{(pre)sel \& SingleMuHLT}}$$

- Turn-on stable wrt SingleMu threshold

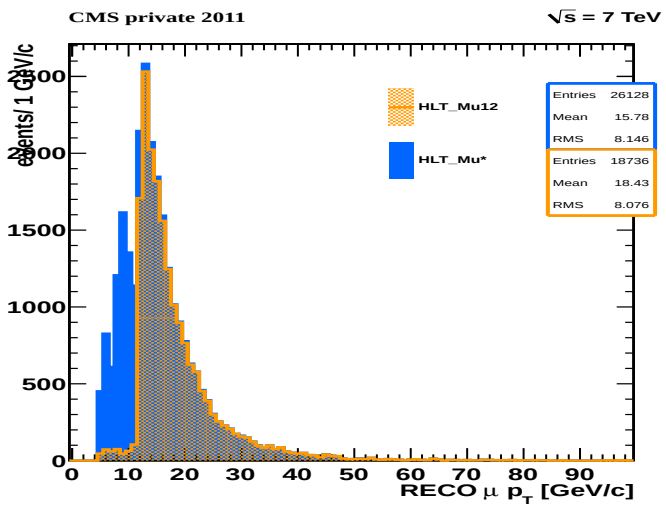
HLT_Mu12_DiCentralJet30_BTagIP3D

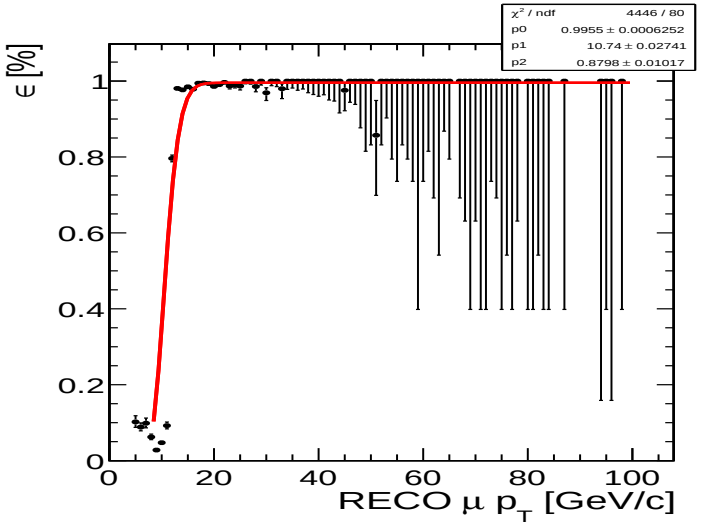


HLT_Mu12_DiCentralJet20_DiBTagIP3D1stTrack



Muon "leg": $\epsilon(\mu)$

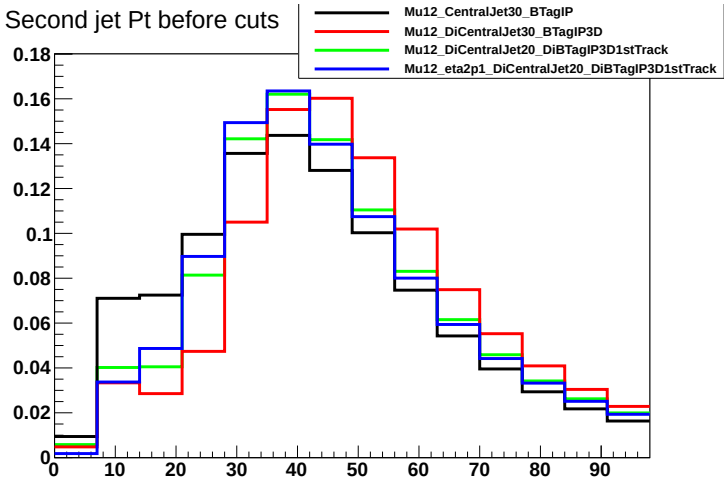


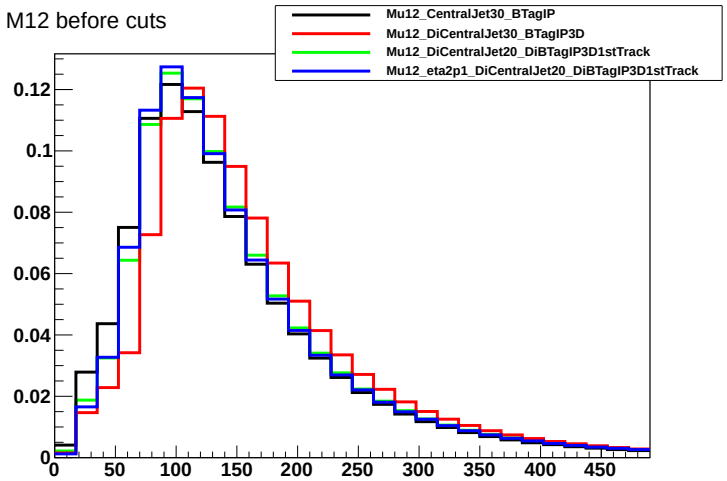


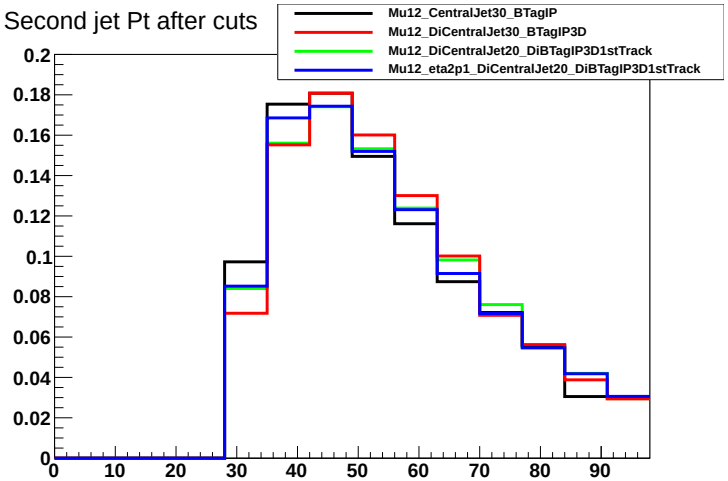
$\epsilon(\mu)$

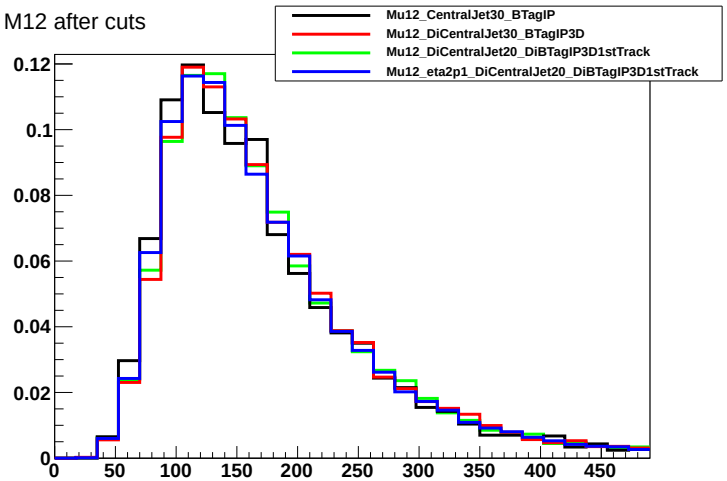
- HLT_Mu12
- Computed on data using pre-scaled singleMu trigger path
- as a function of recontructed muon p_{T}
- for $p_t > 15$ already in plateau at 99.x%.

Trigger paths vs Analysis









After pre-selection cut distributions for different path are similar: OK

Background determination: b-tag parametrization

- Define a control region using a likelihood ratio discriminator using the most discriminating variables (depends on M_H)
- Build B-tagging probability matrices $P^{3^{rd}jet}_{b-tag}(\ldots)$ in control region for third jet, as a function of 3^{rd} jet and event parameters;

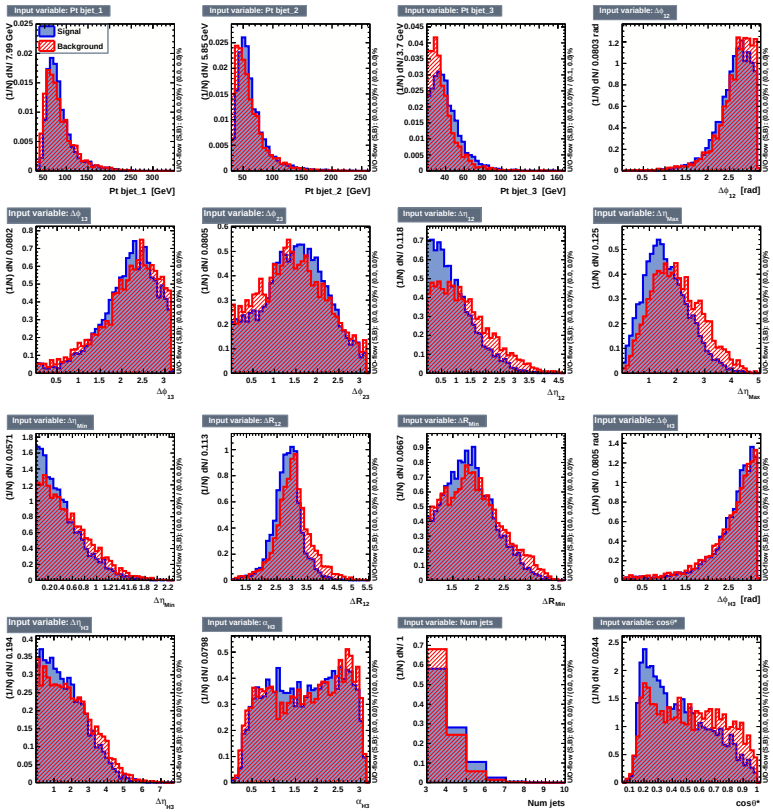
$$P^{3^{rd}jet}_{b-tag}(\ldots) = \epsilon_b \cdot f_b + \epsilon_c \cdot f_c + \epsilon_l \cdot f_l$$

  - b-tagging efficiencies $\epsilon' s$ from MC $\epsilon = \epsilon(E_T, |\eta|, N_{trk})$
  - flavour fractions $f_{b,c,l}$ from Data parametrization see next slides

- Estimate any bbb distribution $F(x; bbb)$ for variable x in signal region starting from same distribution for bbj;

$$F(x; bbb) = F(x; bbj) \otimes P^{3^{rd}jet}_{b-tag}(\ldots)$$

Discriminator: Low Mass $M_H \leq 200$ GeV

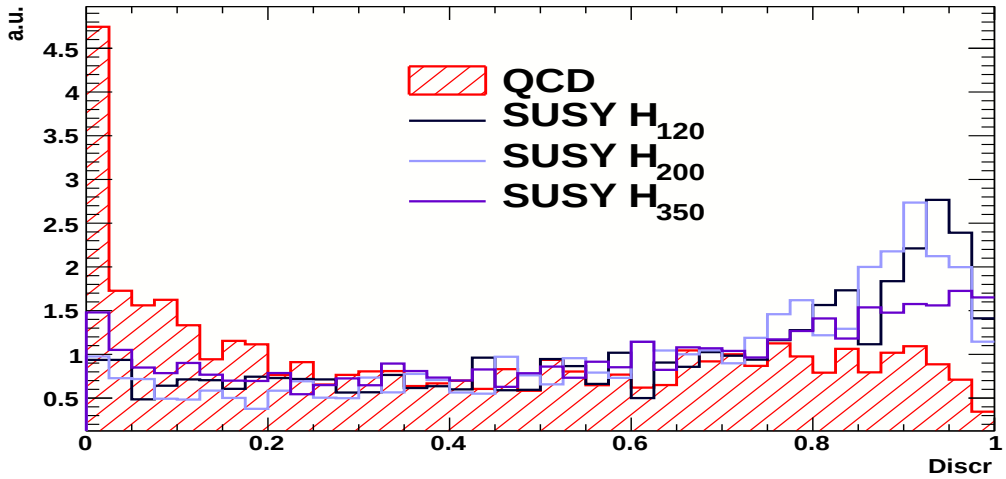


Likelihood ratio

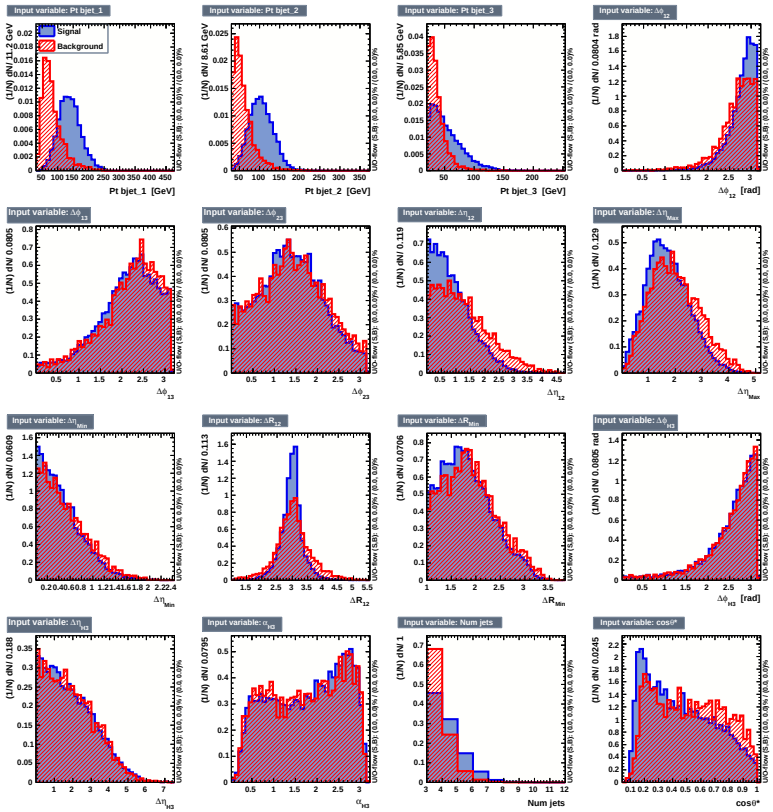
$$Discr = \frac{\Pi_i p_i^{(signal)(x_i)}}{\Pi_i p_i^{(signal)}(x_i) + \Pi_i p_i^{(QCD)}(x_i)}$$

Training $M_H = 120, 130, 140$

Used for test up to $M_H \leq 200$



Discriminator: High Mass $M_H > 200$ GeV

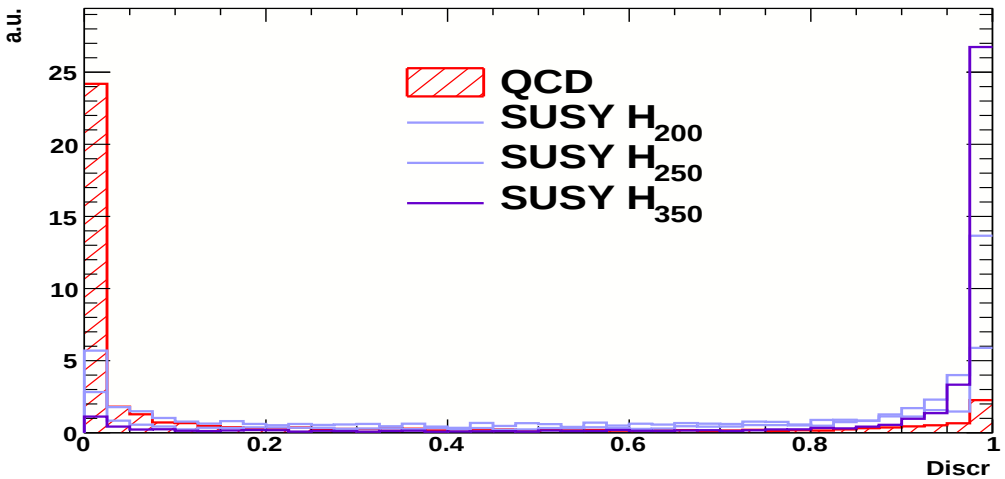


Likelihood ratio

$$Discr = \frac{\Pi_i p_i^{(signal)(x_i)}}{\Pi_i p_i^{(signal)}(x_i) + \Pi_i p_i^{(QCD)}(x_i)}$$

Training $M_H = 250, 300, 350$

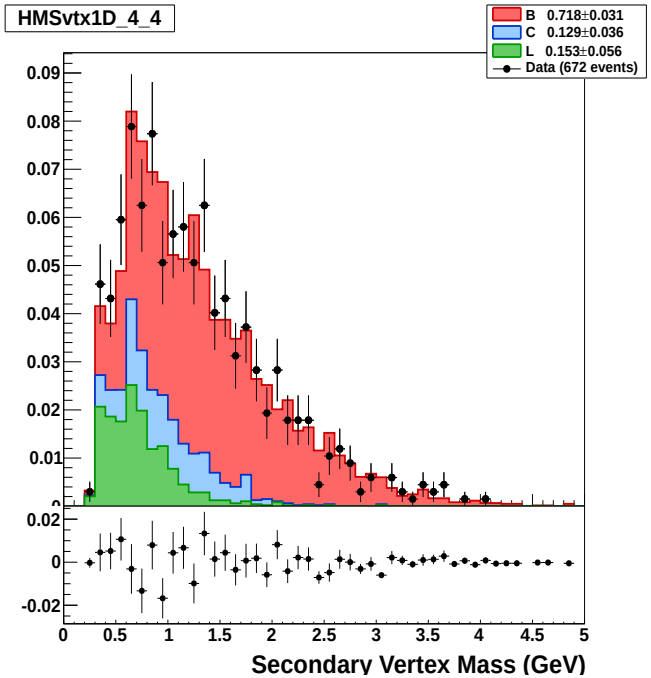
Used for test from $M_H > 250$

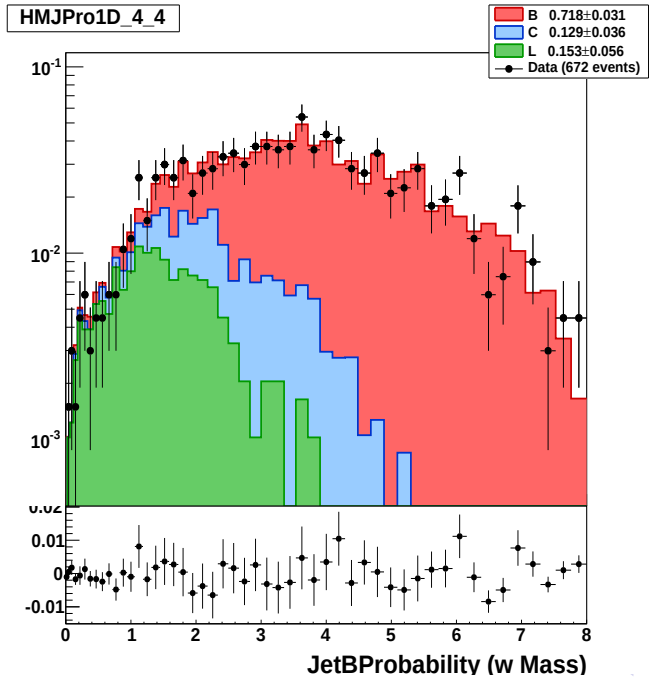


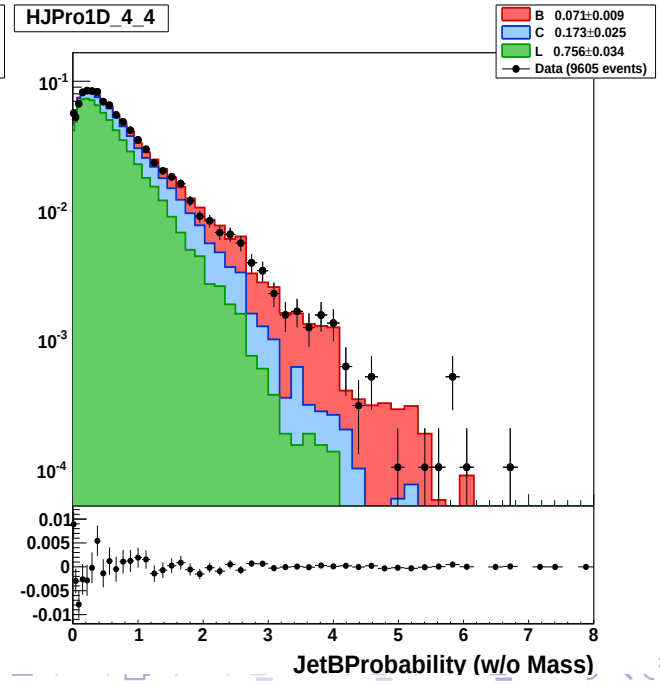
Heavy quark fraction

Consider Mass@Vertex and JetBProbabilty for third jet;

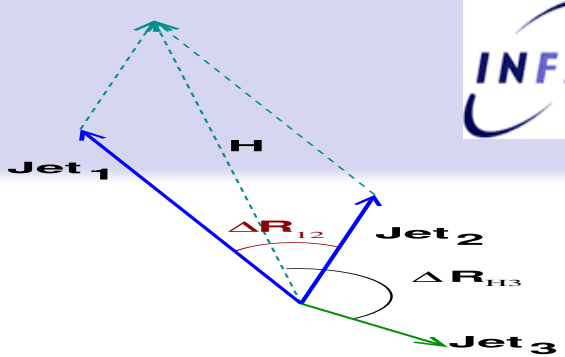
- build distribution templates from MC QCD, for B, C and Light;
- fit third jet distributions using templates and get $F_{b,c}$;
  - used only JetBProbabilty if Mass@Vertex not available.
- for Data, fit separately single B-tag and double B-Tag HLT paths.







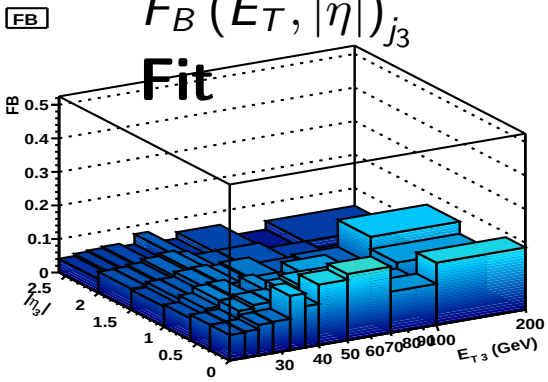
F_B Fit results vs MC truth

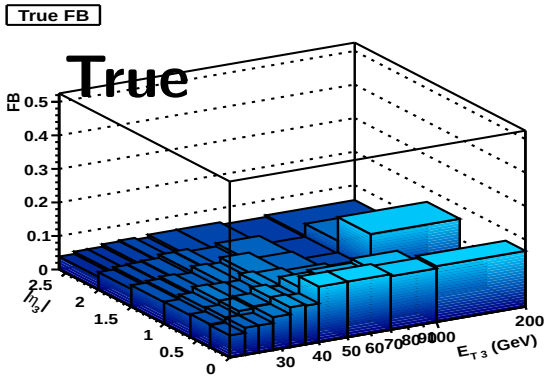


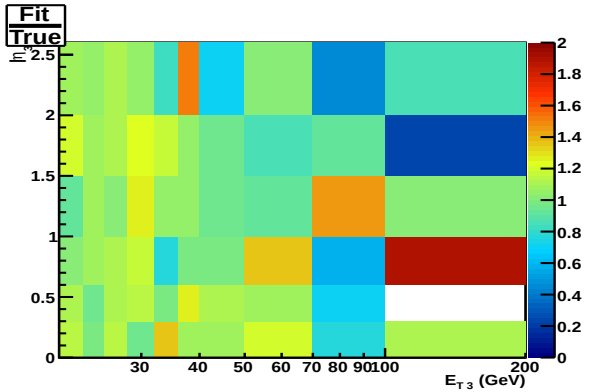
$$F_B = F_{B,C}\left(E_T^{(j_3)}, \left|\eta^{(j_3)}\right|\right) \times F_{B,C}\left(\Delta R_{H,j_3}, \Delta R_{j_1,j_2}\right)$$

assuming no correlation. Second factor only for shape.

$F_B\left(E_T, |\eta|\right)_{j_3}$

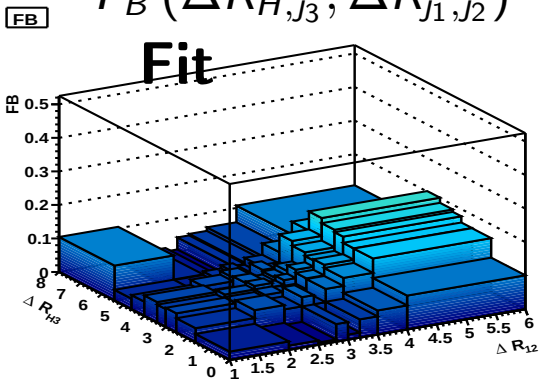


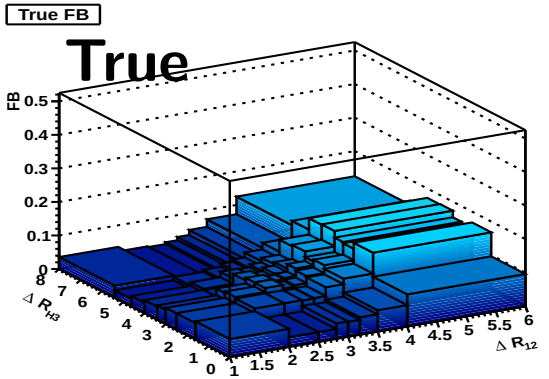


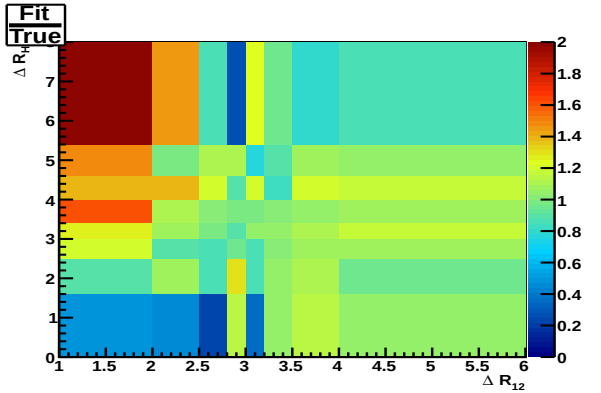


Overall bias for F_b around $\approx +6\%$

$F_B\left(\Delta R_{H,j_3}, \Delta R_{j_1,j_2}\right)$

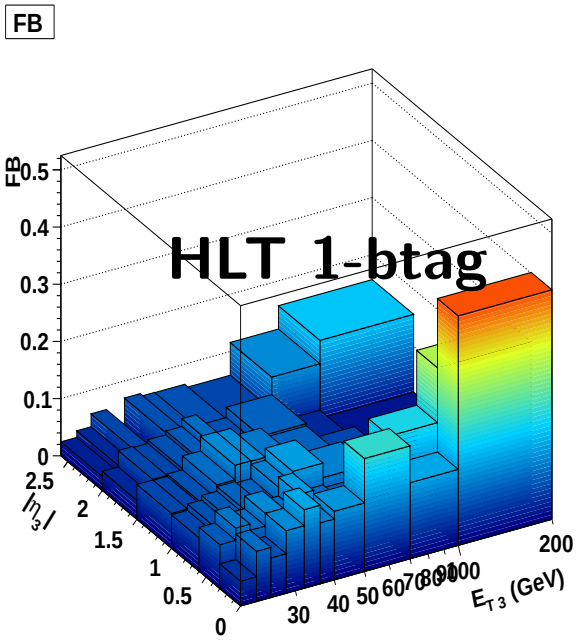


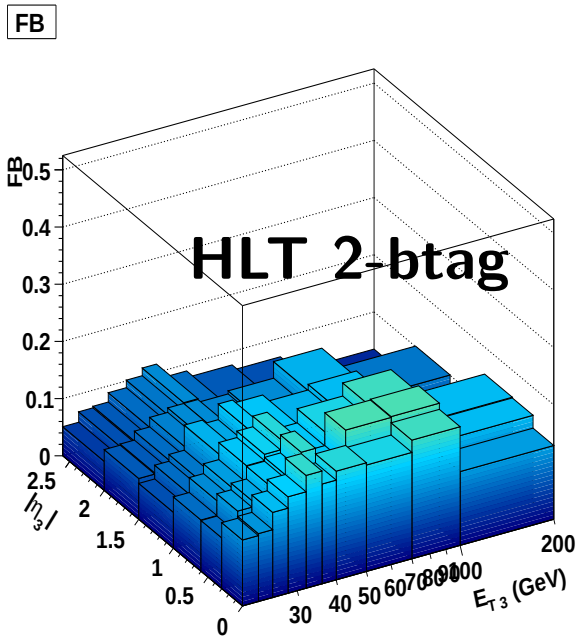


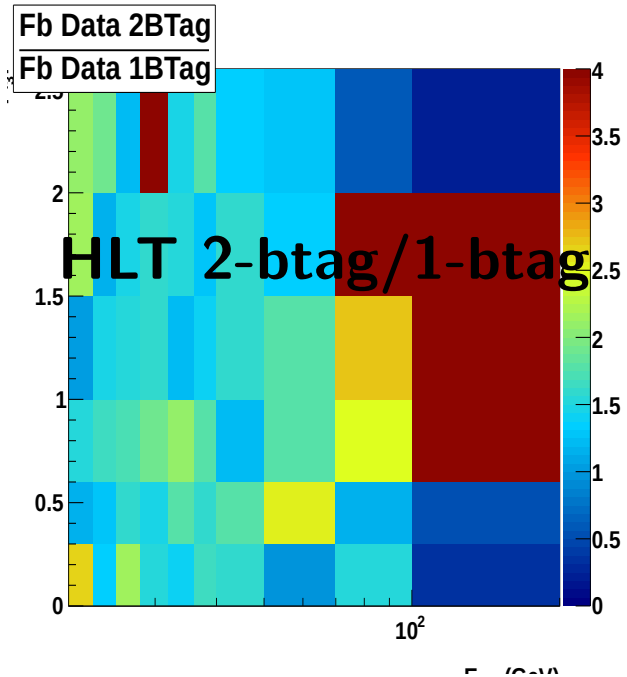


F_B Fit results Data

$$F_B\left(E_T^{(j_3)}, \left|\eta^{(j_3)}\right|\right)$$



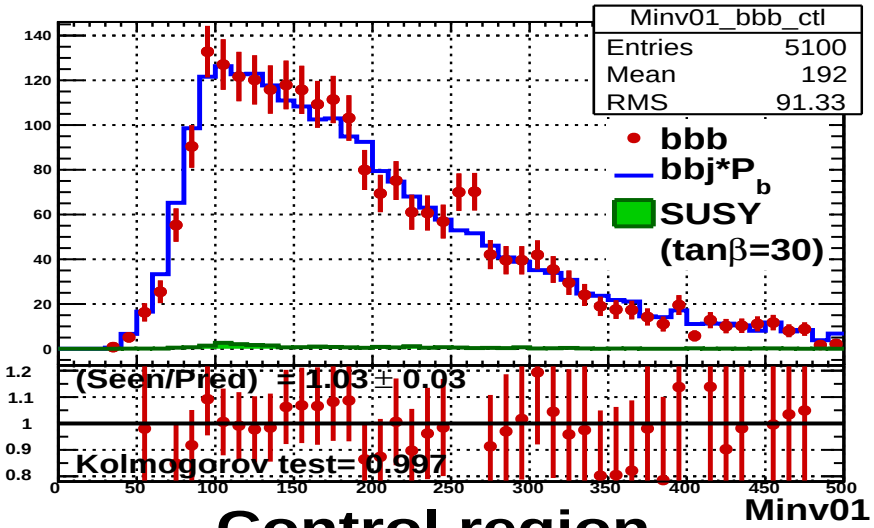


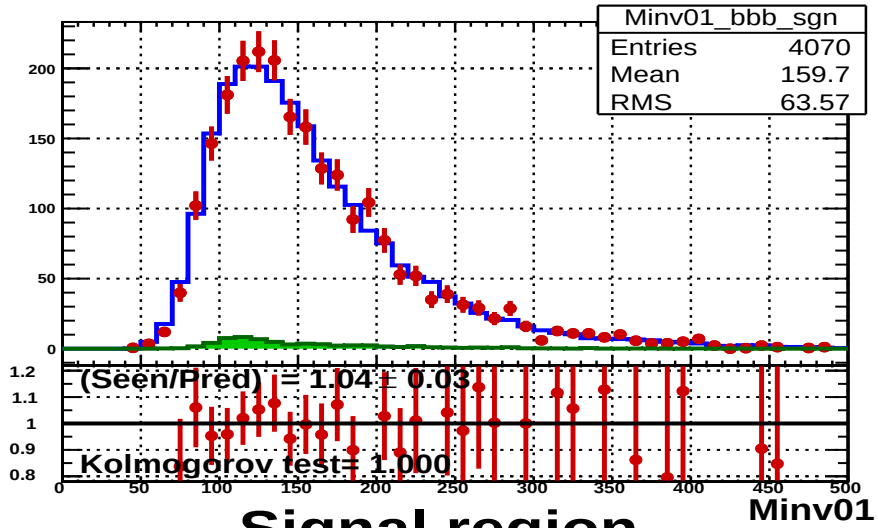


Enanchement of b in third jets, due to the online double b-tag trigger, is clearly visible.

M_{jj} prediction vs *bbb* in MC and Data

in control and signal region: Low Mass ($M_H \leq 200$) susy $M_H = 120$ *GeV*

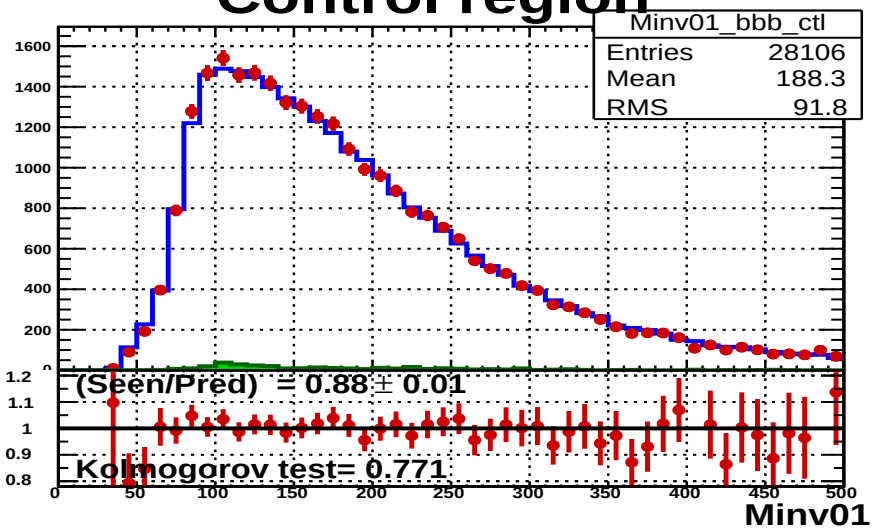


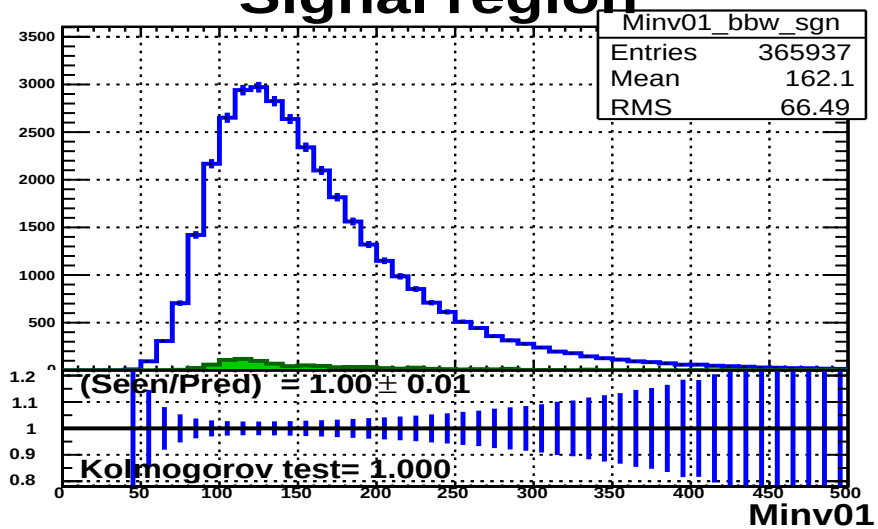


MC

Control region

Signal region



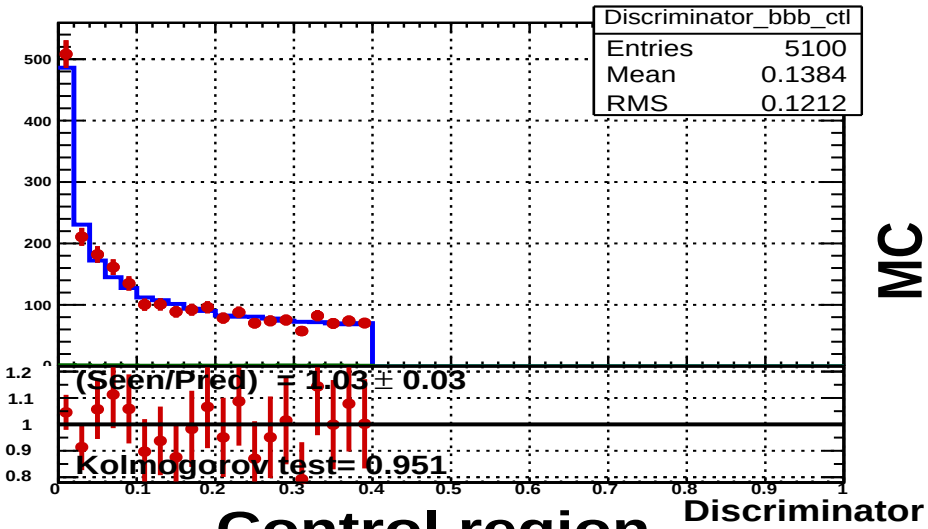


Data

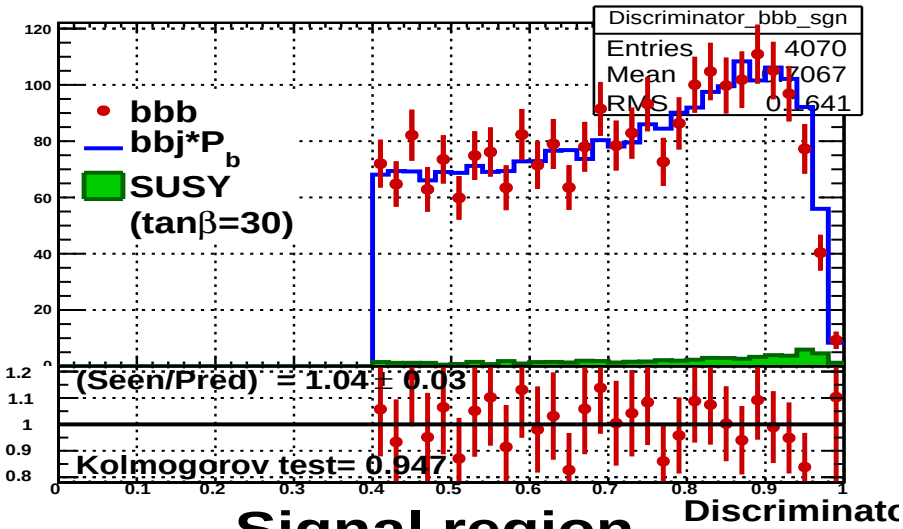
For Data (control region) shape fine, normalization $\sim 0.88\%$

Discriminator prediction vs *bbb* in MC and Data

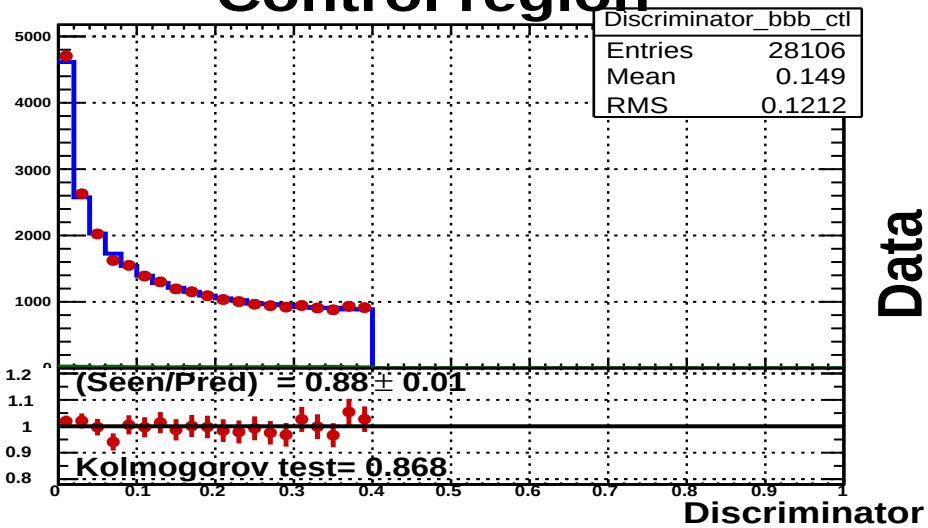
in control and signal region: Low Mass ($M_H \leq 200$) susy $M_H = 120$ *GeV*



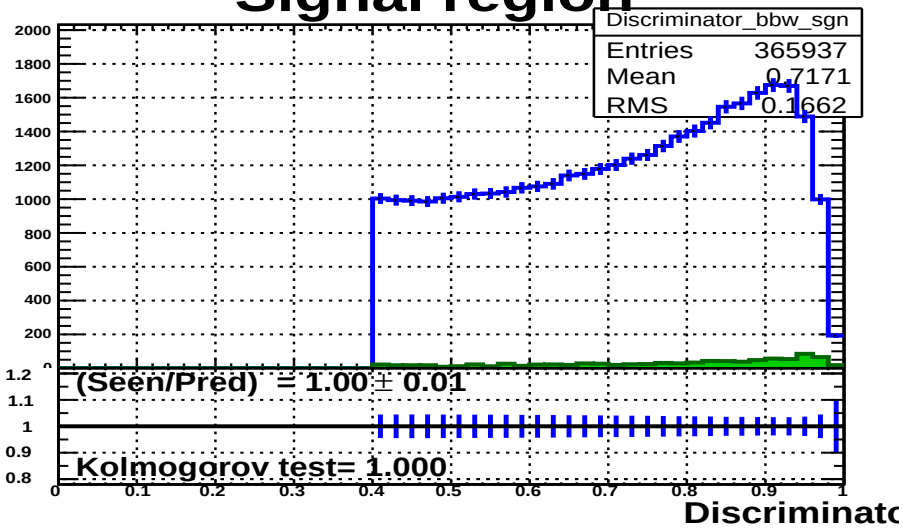
MC



Control region



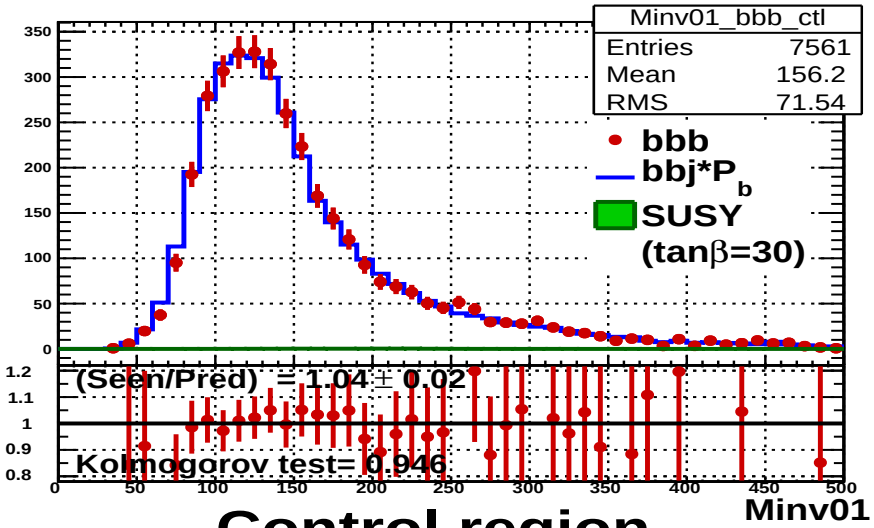
Data

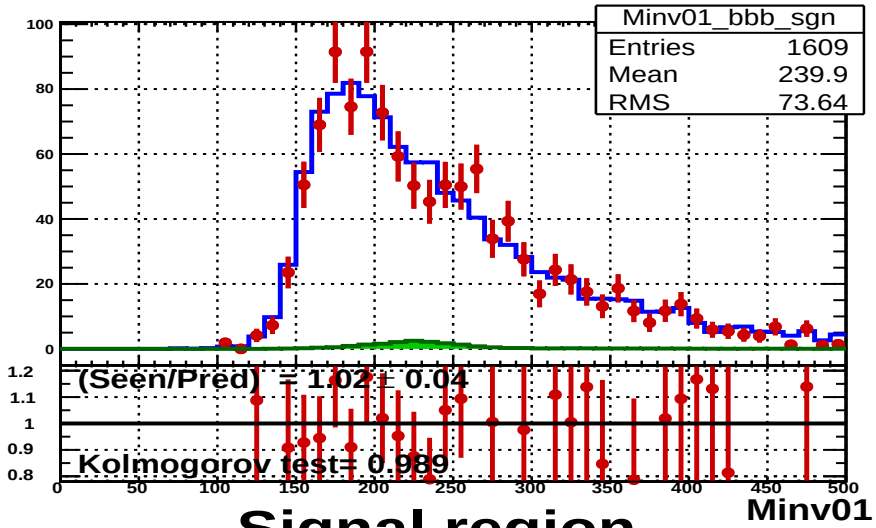


Signal region

M_{jj} prediction vs *bbb* in MC and Data

in control and signal region: High Mass ($M_H > 200$) susy $M_H = 250$ *GeV*

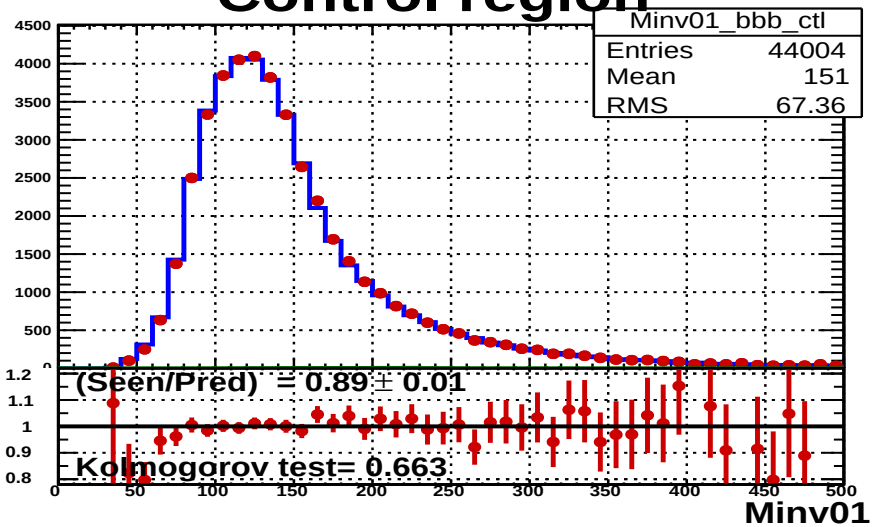


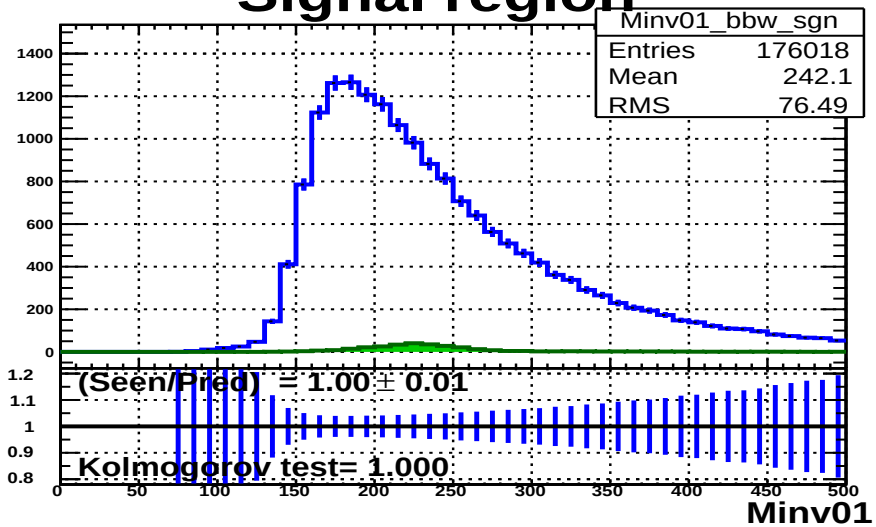


MC

Control region

Signal region



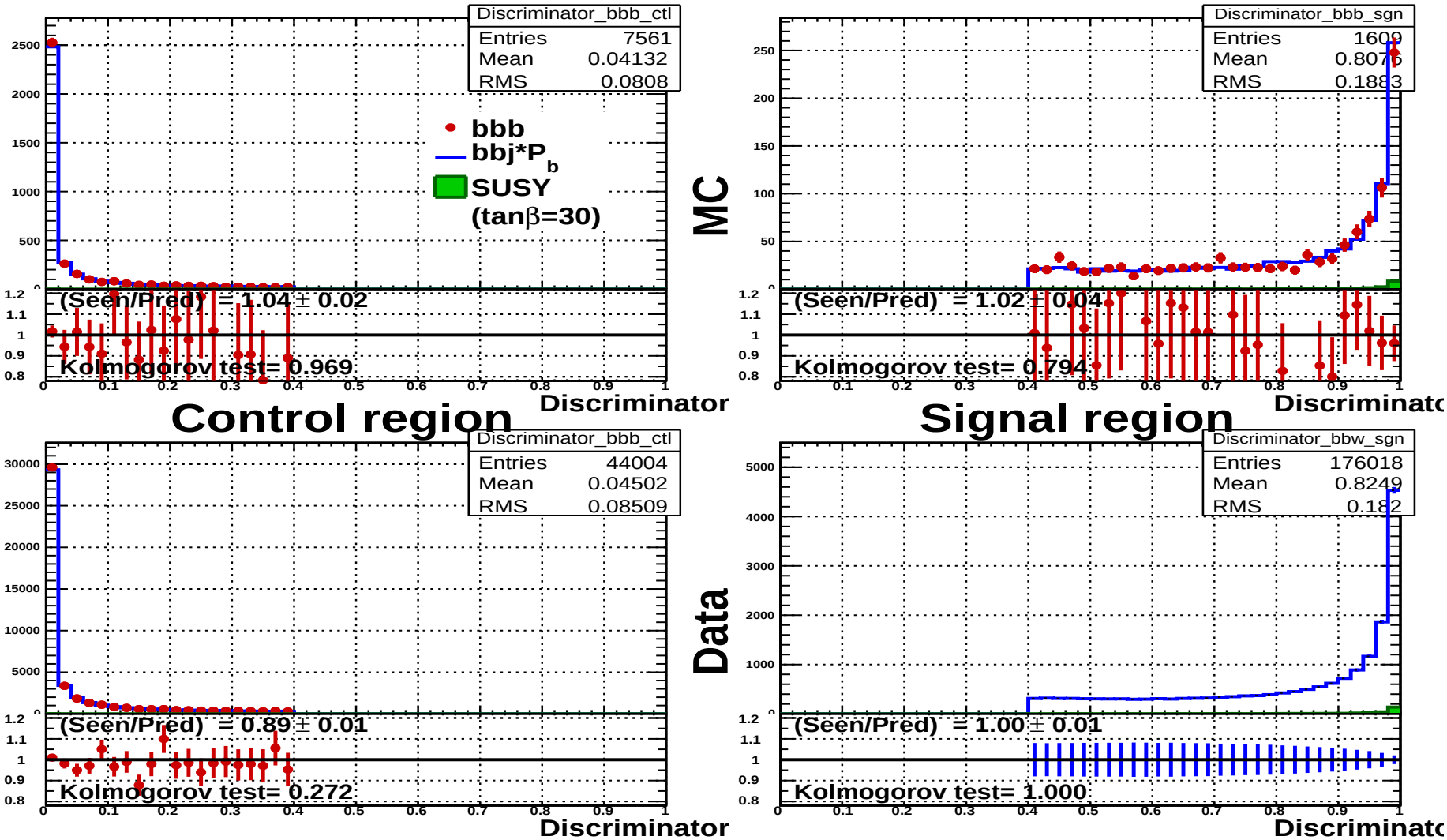


Data

For Data (control region) shape fine, normalization $\sim 0.88\%$

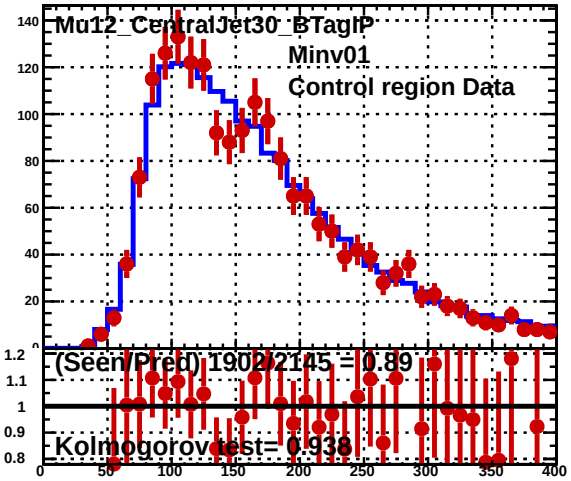
Discriminator prediction vs *bbb* in MC and Data

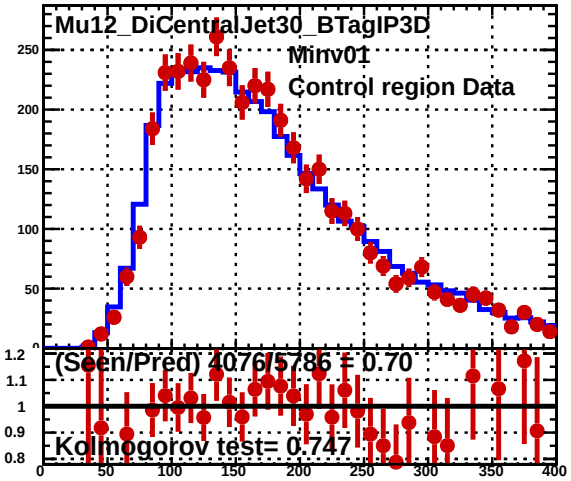
in control and signal region: High Mass ($M_H > 200$) susy $M_H = 250$ *GeV*

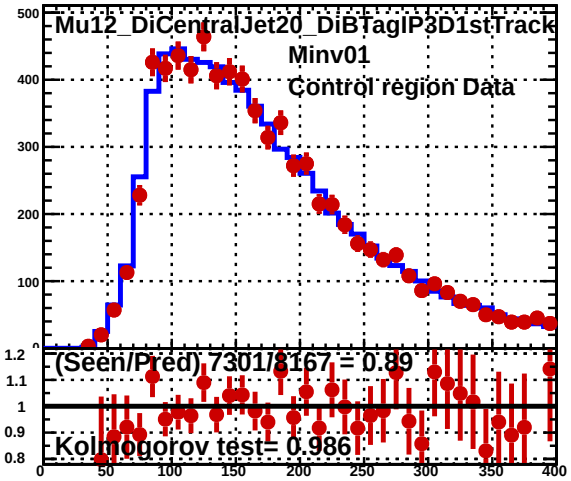


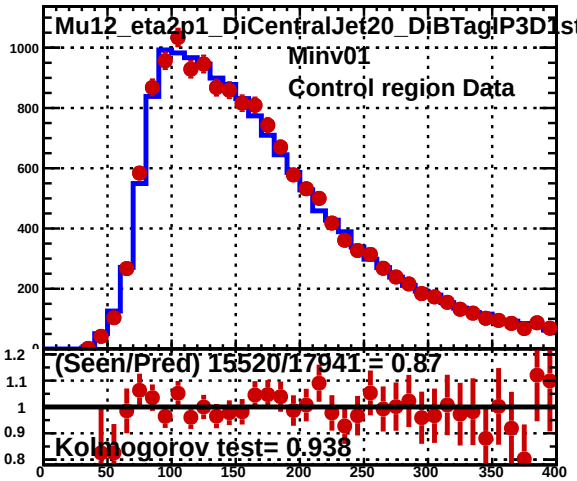
M_{jj} prediction vs *bbb* in MC and Data vs HLT paths

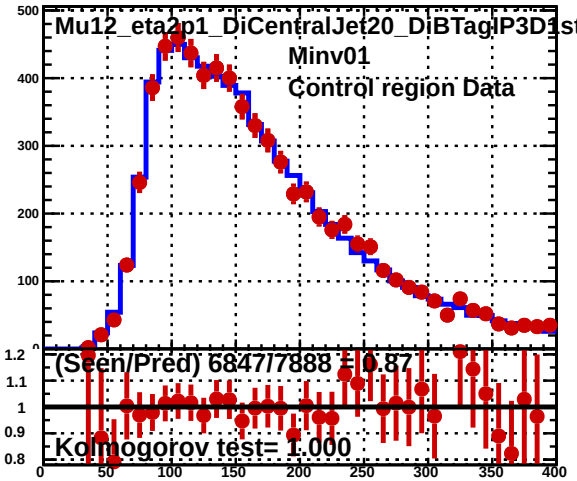
in control and signal region: Low Mass ($M_H \leq 200$) susy $M_H = 120$ *GeV*











Shape fine for all HLT paths

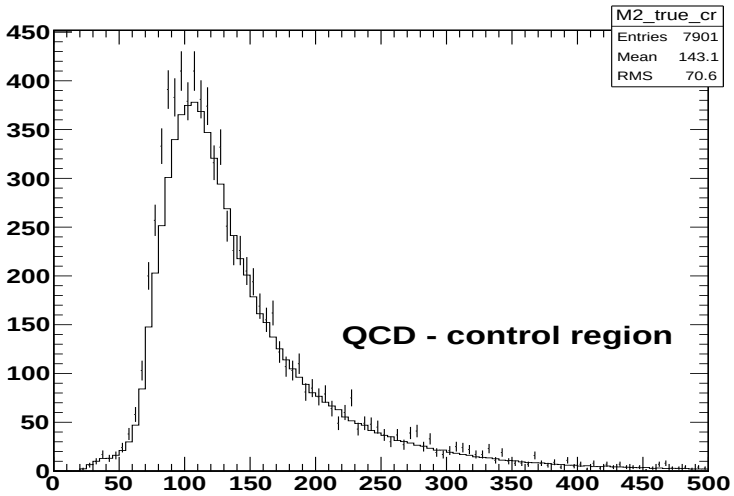
Hyper Ball

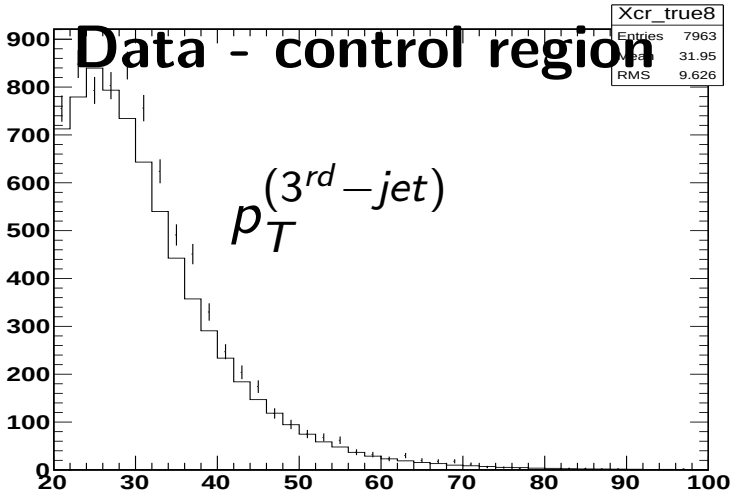
Second approach, independent background estimation.

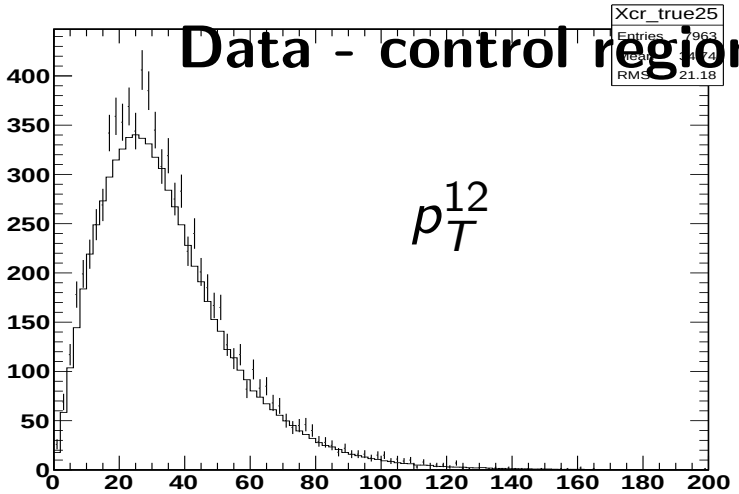
General idea

- Start from *bjj* sample, control region;
- For each event in $(bjj)_{CR}$ select a set of *similar* events $\mathcal{O}(1000)$
- Compute the fraction of these events passing full selection (bbb);
- *similarity* is defined by *distance* between events in hyperspace $d = \Sigma_i^n \left(w_i \left(x_i - y_i\right)\right)^2$
  - with x_i, y_i jet or event variables (p_T, η, $\Delta\phi_{ij}$, ...);
  - w_i weight to account for variability of *fraction* vs a given variable;
- Use the average fraction to weight events and predict *bbb* distribution.

Hyper Ball: prelim results on QCD MC and Data



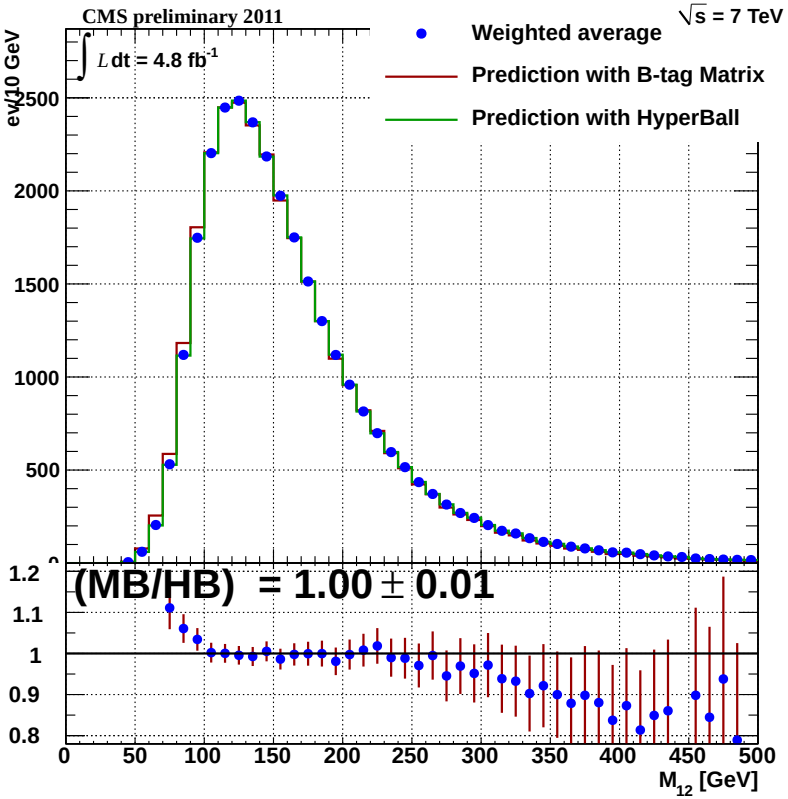




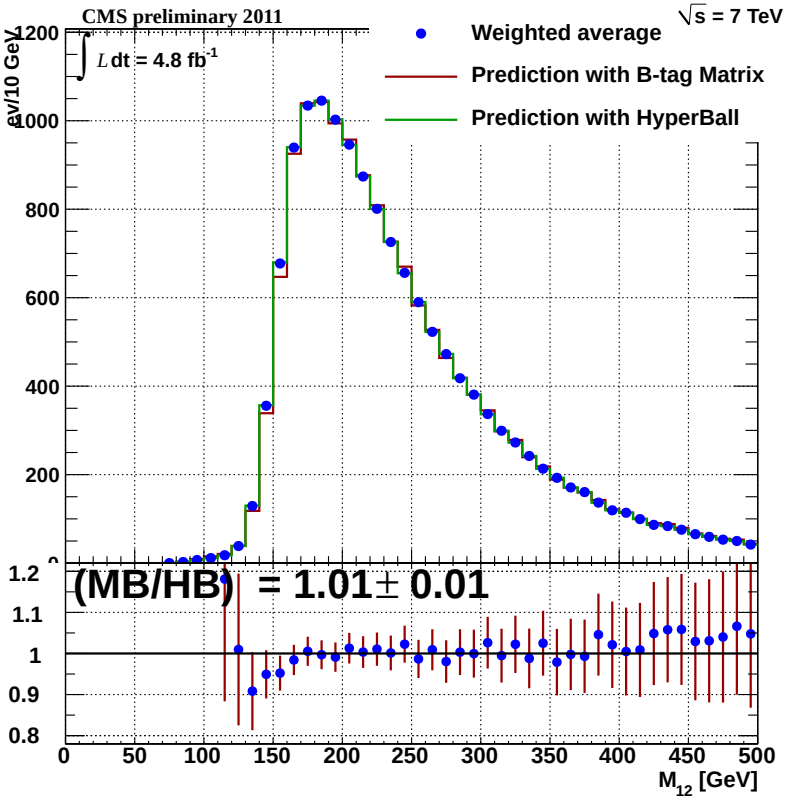
Comparison of B-Matrix and Hyperball prediction

NOT NORMALIZED

[Low Mass $M_H < 200\ GeV$]



[High Mass $M_H > 200\ GeV$]



Systematics

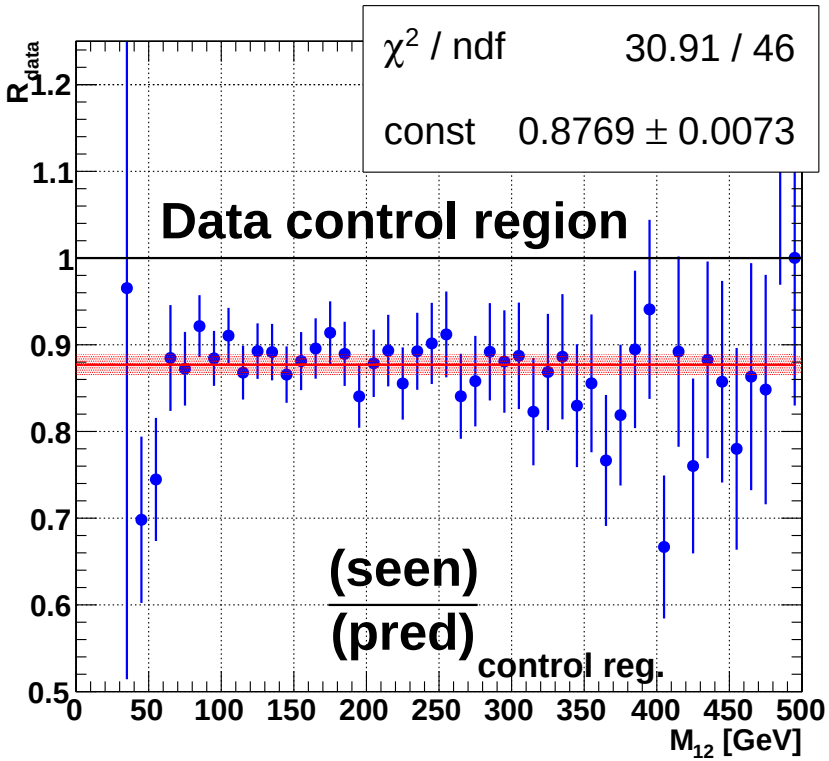
- Trigger syst: $\approx 3-5\%$ from data driven ϵ estimate;
- Physics object syst:
  - B-tagging eff. BTV-12-001 $\approx 4\%$ per BJet: $\approx 12\%$ for three jets.
  - JEScale $^{+2.5}_{-3.1}\%$
  - JEResolution $\pm 1.9\%$
  - Mu momentum scale $\approx 0.2\%$ and resolution $\approx 0.6\%$
- pdf For $M_H = 120\ GeV$: $^{+2.5}_{-2.7}\%$; for $M_H = 250\ GeV$: $^{+4.7}_{-4.4}\%$;
- Integrated Lumi syst: $\approx 2.2\%$
- Background normalization syst $\approx 5\%$:

Background Systematics

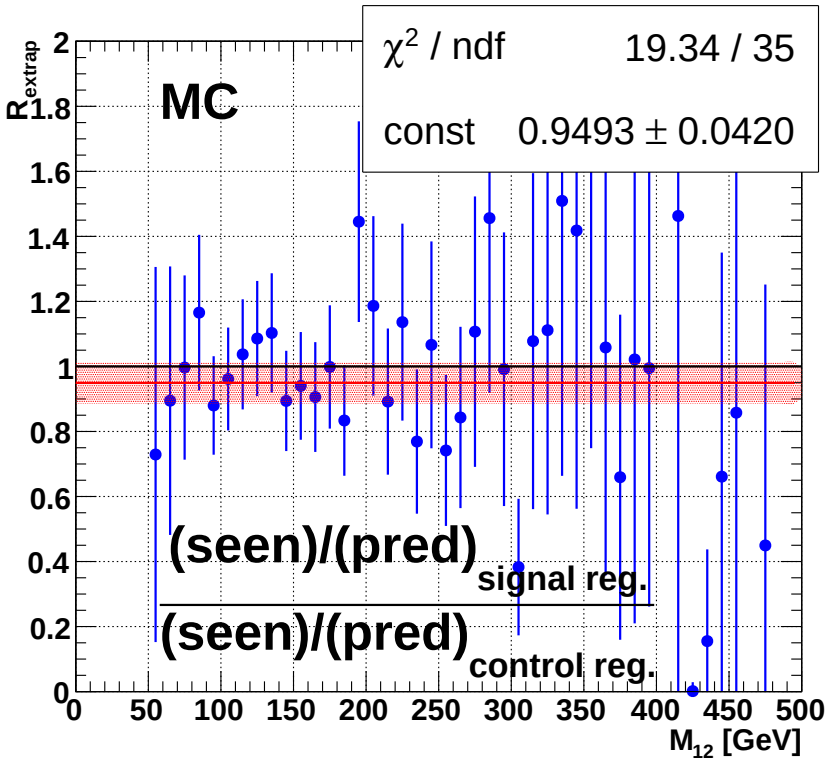
- Two major source of systematics for the predicted *bbb* in signal region:
  1. Systematics from *bbb* prediction from DATA control region
     - ★ compare *bbb* and $bbj \times P$ in DATA control region;
     - ★ use normalization in signal region;
     - ★ use fit error as systematics;
  2. Systematics due to extrap. from control to signal region from MC
     - ★ get ratio of ratios from MC (signal/control) and fit it;
     - ★ use fit results to correct extrapolation bias;
     - ★ and fit errors to estimate systematics for extrapolation;
- both can be used bin per bin when computing CL's
- we can use *no man's land* in data to check extrapolation systematics and bias in data;

Systematics test

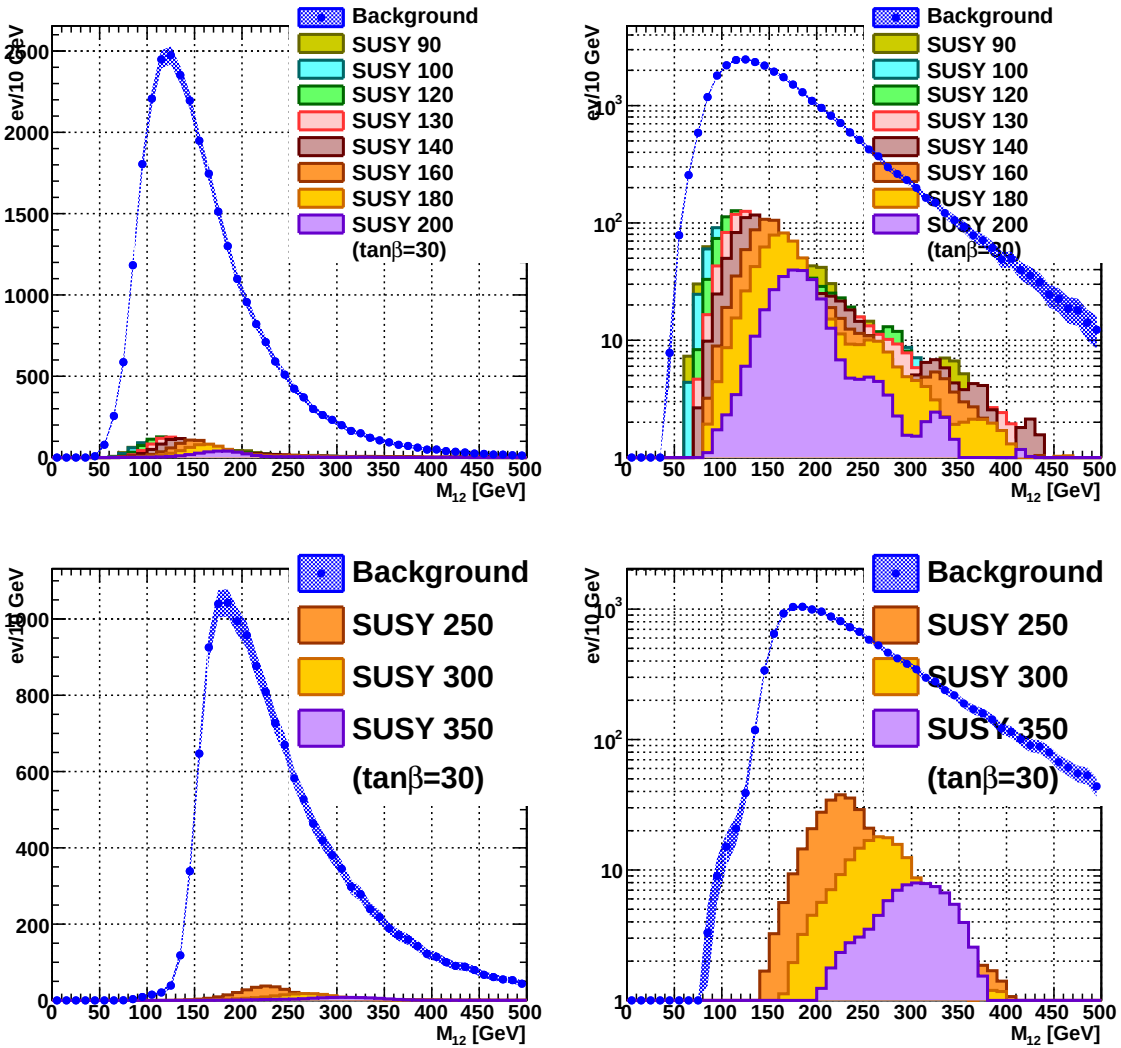
[Data control region]



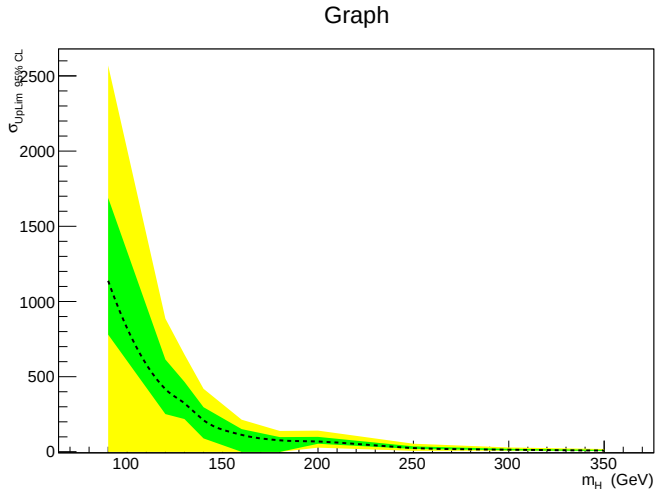
[MC]



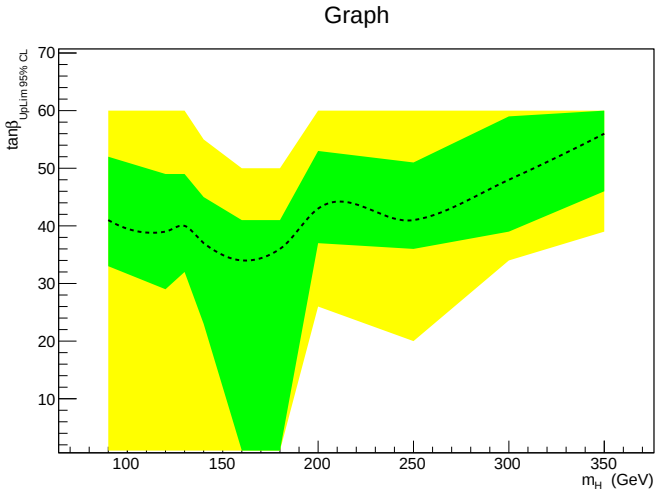
Final prediction and signal



Sensitivity

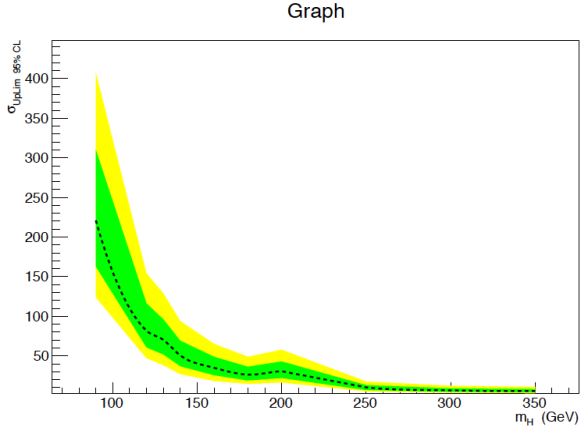


CLs limits on Xsec

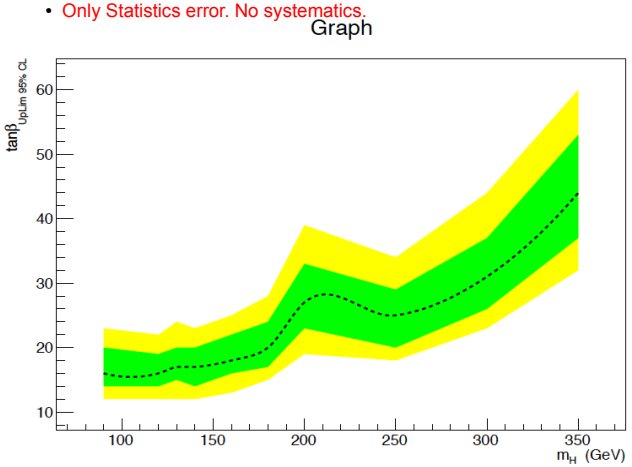


CLs limits on TanB

- Only Statistics error. No systematics.



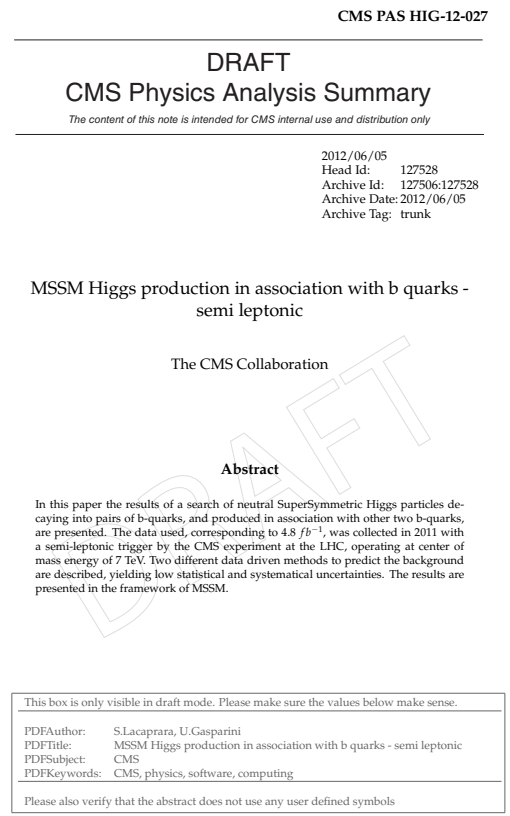
- Only Statistics error. No systematics.



Summary

Issues (not public!)

- we do have some problem in PU reweight;
- Discriminator distribution is different in training and test in signal region (which makes use of the same QCD MC!);
- not able to fit template for $F_{b,c}$ in QCD MC if trigger and PU weights are included;
- trigger systematics are estimated by eye;
- unbliding procedure is just "open the box" or we should open it in two steps?
- Final limit seems to be too high.
- By comparison the MSSM Hbb full hadronic channel excludes starting from tan $\beta \approx 20$



CMS PAS HIG-12-027

DRAFT
CMS Physics Analysis Summary

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

2012/06/05
Head Id: 127528
Archive Id: 127506:127528
Archive Date: 2012/06/05
Archive Tag: trunk

MSSM Higgs production in association with b quarks - semi leptonic

The CMS Collaboration

Abstract

In this paper the results of a search of neutral SuperSymmetric Higgs particles decaying into pairs of b-quarks, and produced in association with other two b-quarks, are presented. The data used, corresponding to 4.8 fb^{-1}, was collected in 2011 with a semi-leptonic trigger by the CMS experiment at the LHC, operating at center of mass energy of 7 TeV. Two different data driven methods to predict the background are described, yielding low statistical and systematical uncertainties. The results are presented in the framework of MSSM.

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

PDFAuthor: S.Lacaprara, U.Gasparini
PDFTitle: MSSM Higgs production in association with b quarks - semi leptonic
PDFSubject: CMS
PDFKeywords: CMS, physics, software, computing

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

BACKUP

Physics Objects

- CMSSW 4_2_7 including JetMet suggested tags.
- AK5 ParticleFlow Jets, JEC applied:
  - L1FastJet, L2Relative, L3Absolute, L2L3Residual (only for Data)
  - Global Tag: `FT_R_42_V20A` and `START42_V17` for Data and MC
  - PU treatment: PF Charged Hadron Subtraction and Area Method;
  - Loose JetId selections;
- Jet b-tagging used is Combined Secondary Vertex (CSV);
- Standard Global Muon (no isolation requirements)

- JES and JER from POG (CERN-PH-2011/102 and update JetMET presentation 9/1/12);
- BTag efficiency studies on top samples (BTV-12-001);
- Muon (non isolated) efficiency on J/ψ MUO-10-004;

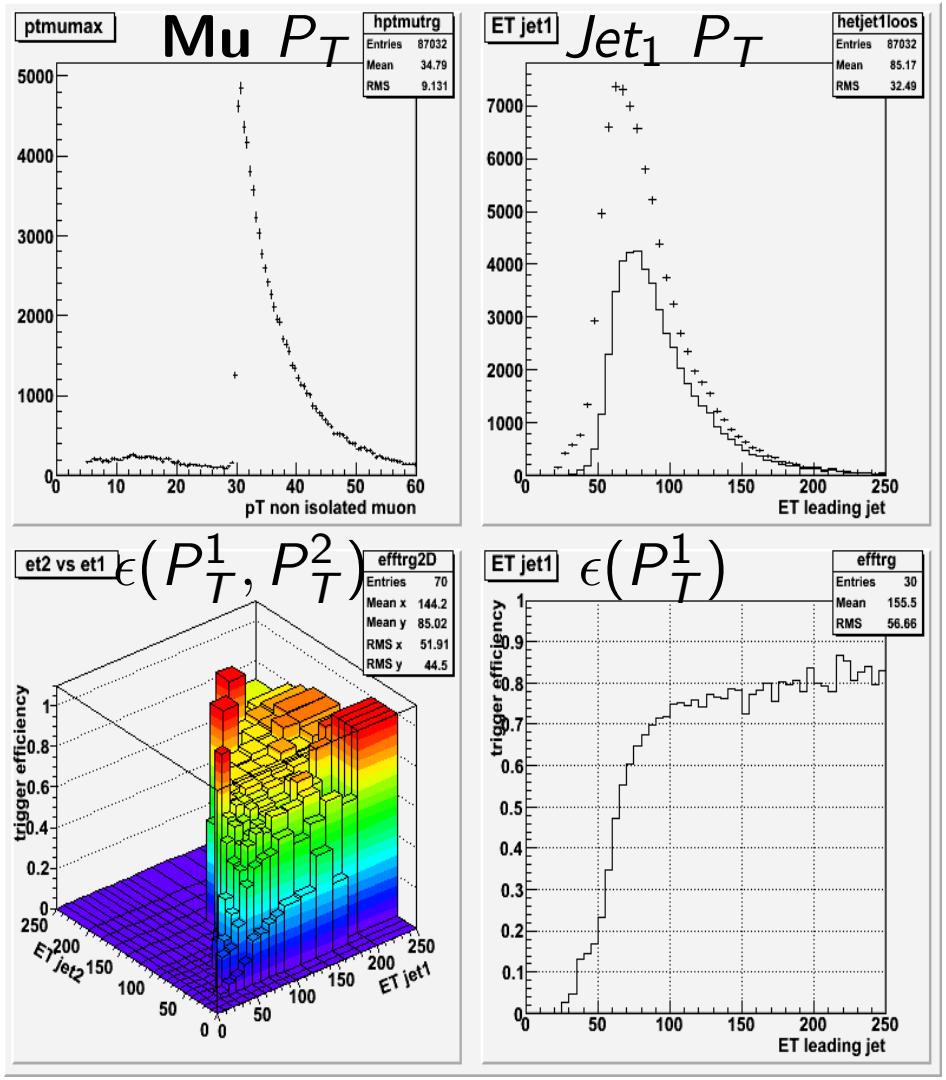
Numerology

HLT paths *(L1 seed)*	run range	triggered events	$\int \mathcal{L} dt$ $[\mathrm{pb}^{-1}]$
HLT_Mu12_CentralJet30_BtagIP			
L1_SingleMu7	163738-165633	3 027 717	183.815
HLT_Mu12_DiCentralJet30_BtagIP3D			
L1_SingleMu10	165970-172952	4 532 555	524.904
HLT_Mu12_DiCentralJet20_DiBtagIP3D1stTrack			
L1_SingleMu10	167039-173198	1 873 247	780.378
HLT_Mu12_eta2p1_DiCentralJet20_DiBtagIP3D1stTrack			
L1_Mu10_Eta2p1_DoubleJet_16_8	173236-178380	4 762 858	1944.527
HLT_Mu12_eta2p1_DiCentralJet20_DiBtagIP3D1stTrack*			
L1_Mu10_Eta2p1_DoubleJet_16_8	178420-180252	2 164 634	811.861
All		16 361 011	4245.485

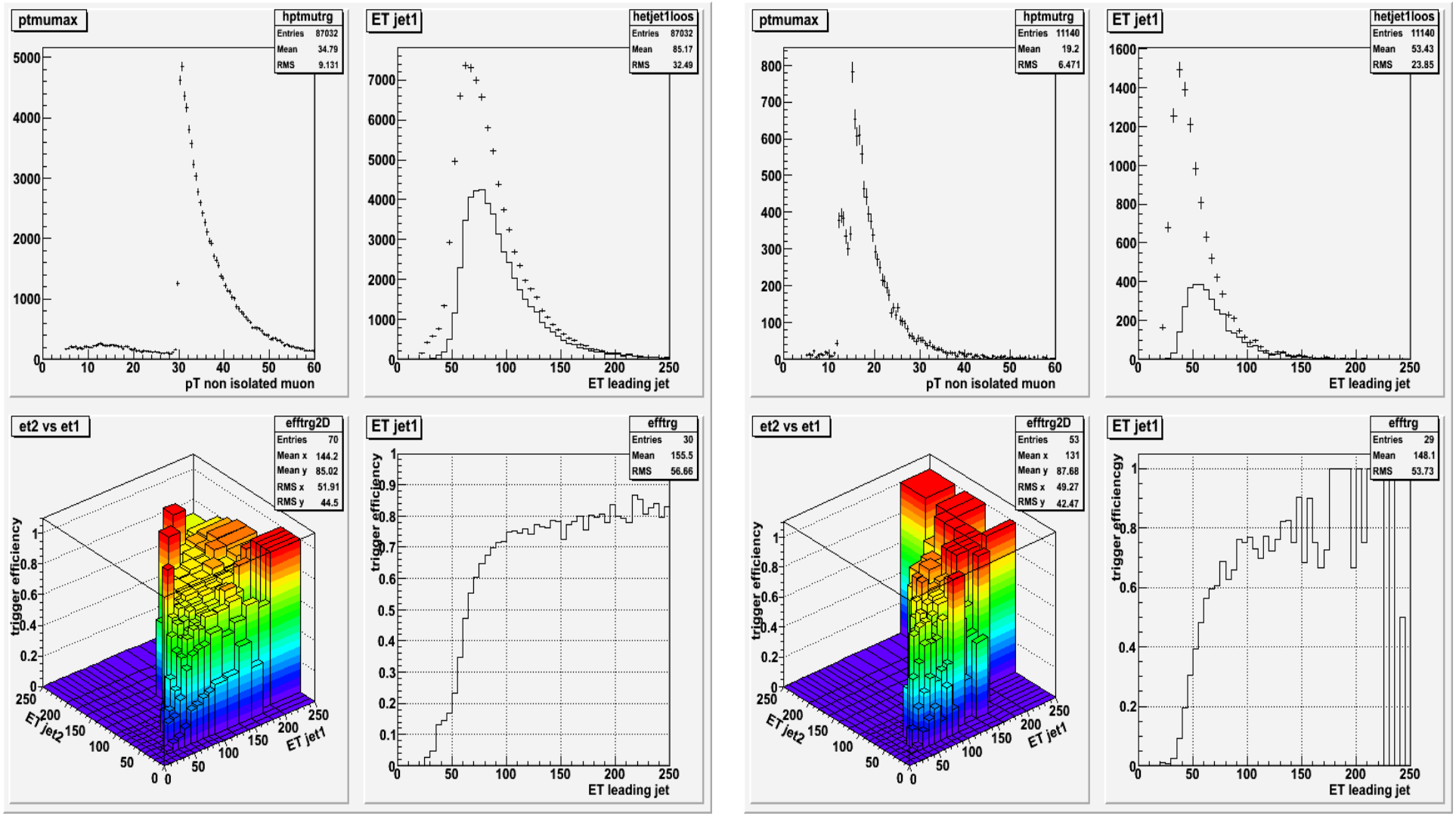
Warning: still `lumiCalc2.py`

Trigger Efficiency vs Analysis

- Use **SingleMuon PD**, select all events passing a single muon path.
- Apply preselection (2 bjets)
- Build Turn on curves vs first and second B-jets Pt:
  $\epsilon = \frac{\text{Hbb path \& presel \& SingleMuHLT}}{\text{presel \& SingleMuHLT}}$
- Here for `HLT_Mu12_DiCentralJet30_BtagIP3D` and `SingleMu30`
- Turn-on stable wrt SingleMu threshold



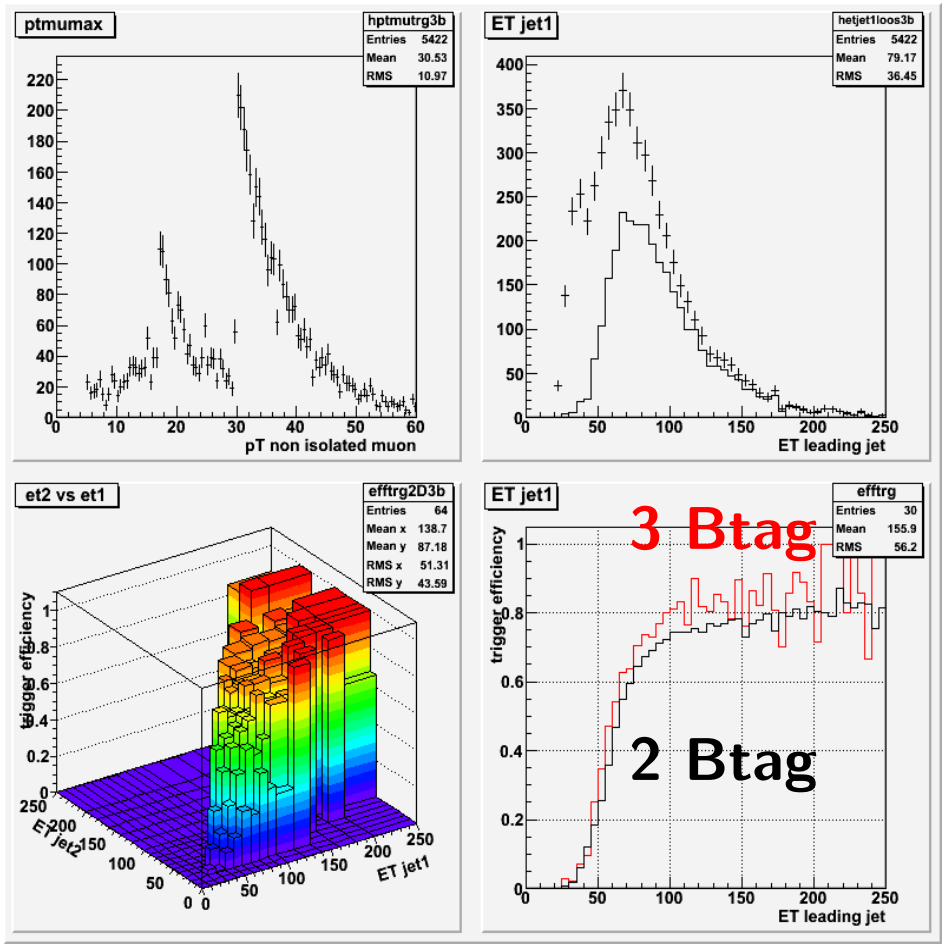
Trigger Efficiency (I)



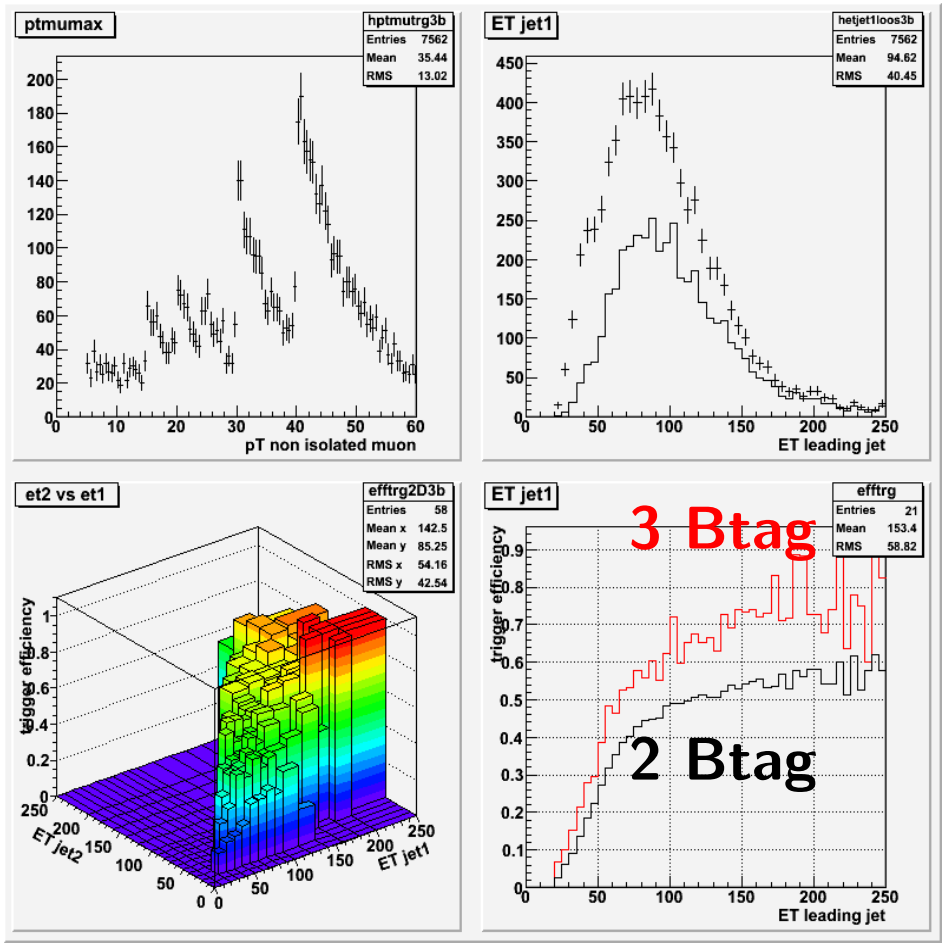
Turn-on stable for different SingleMu threshold (left Mu30, right Mu12 &

Trigger Efficiency (II) 2 btags vs 3 btags

HLT_Mu12_DiCentralJet30_BtagIP3D

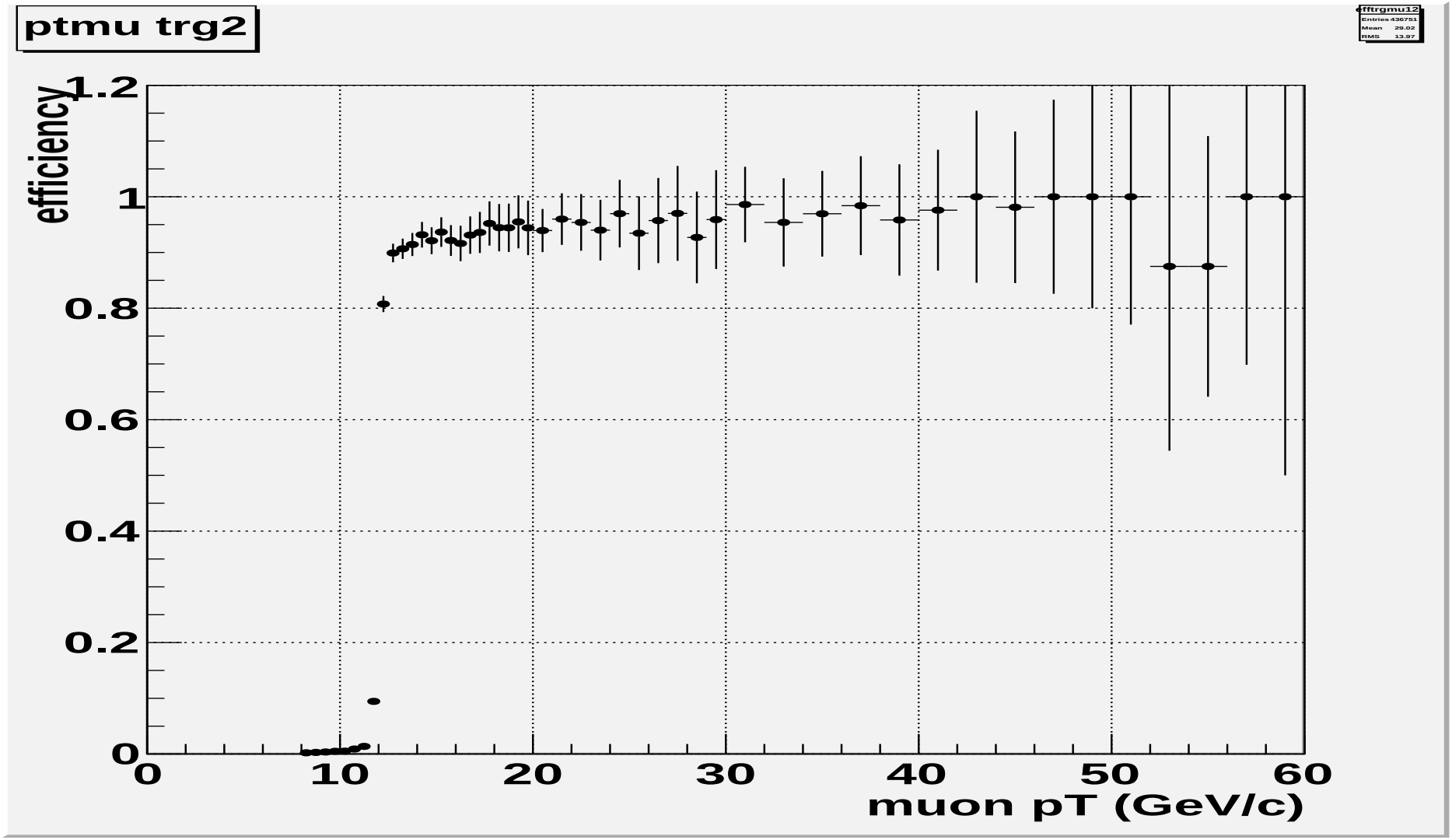


HLT_Mu12_DiCentralJet20_DiBtagIP3D1stTrack



Statistics bit low but still affordable

HLT Mu12 trigger efficiency wrt Analysis

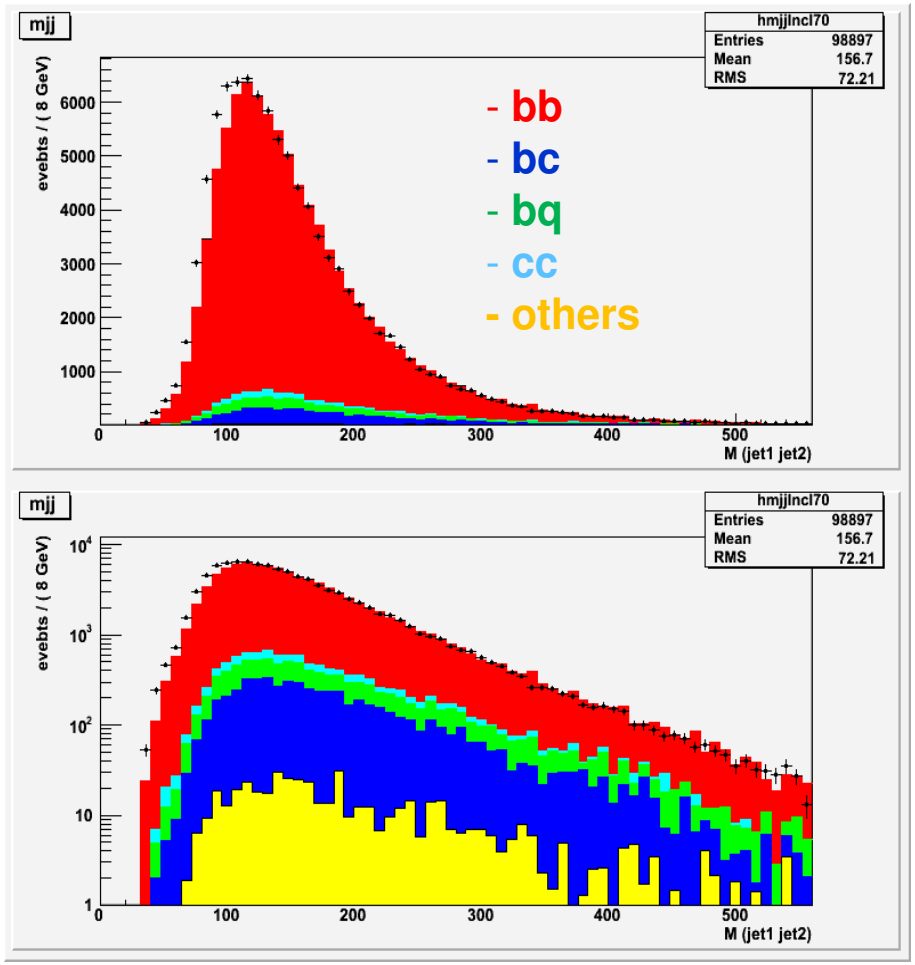


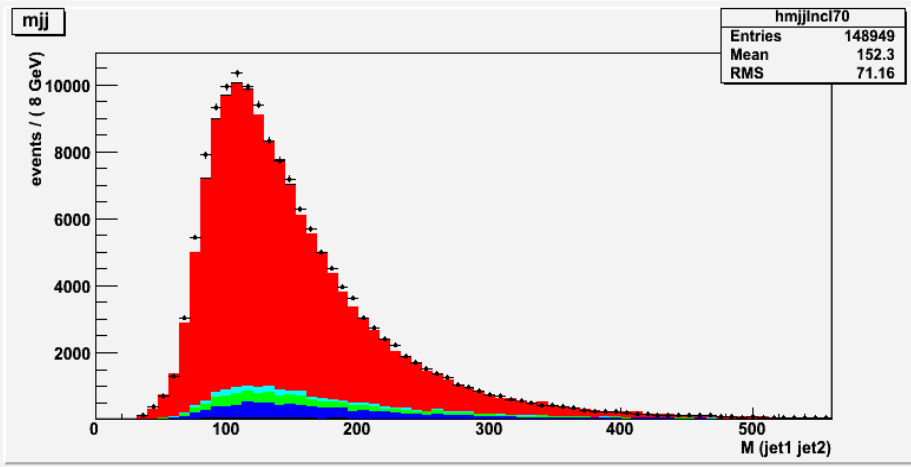
Mass spectra, trigger corrected

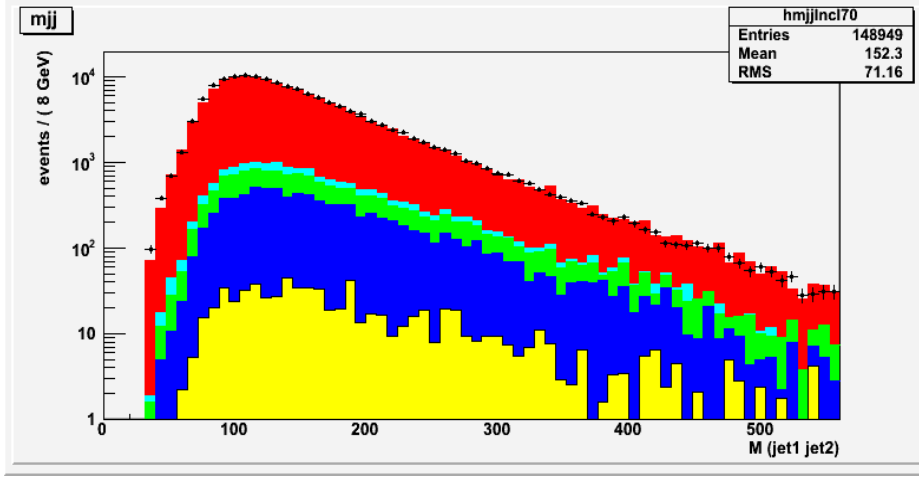
HLTMu12_DicentralJet30_BtagIP3D — **L = 525 pb-1**

2 btags

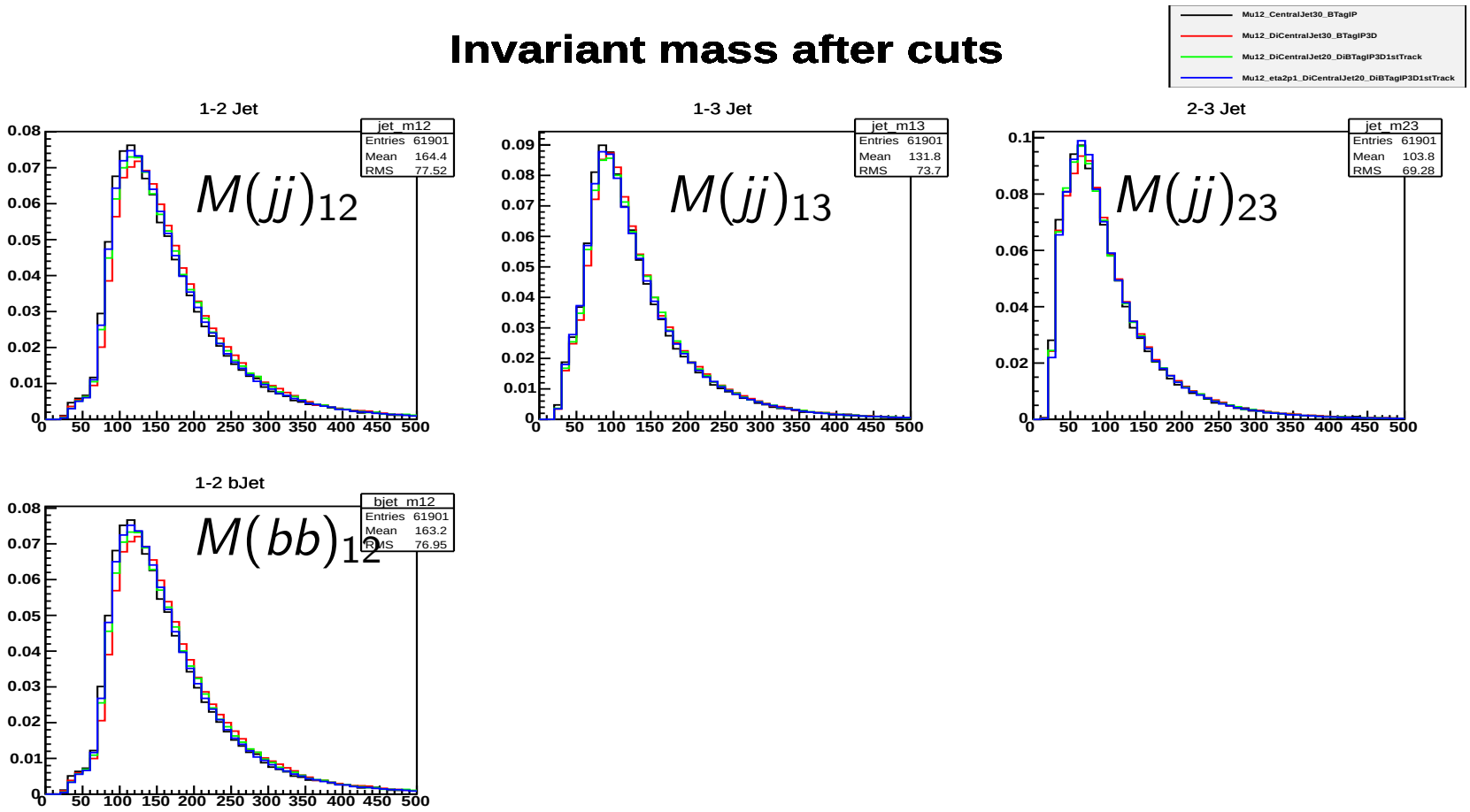
HLTMu12_DicentralJet20_DiBtagIP3D1stTk — **L = 1002 pb-1**





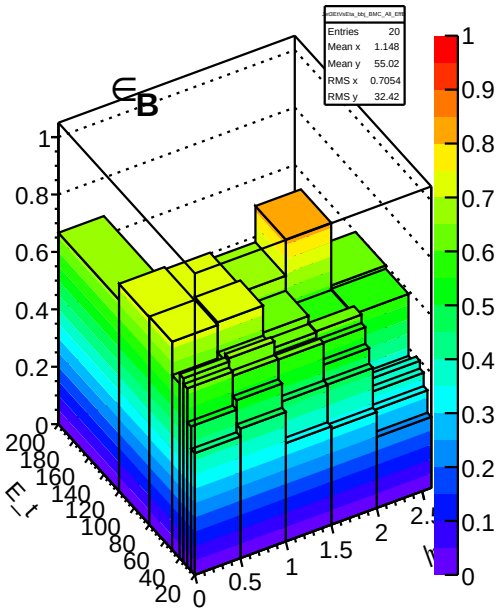


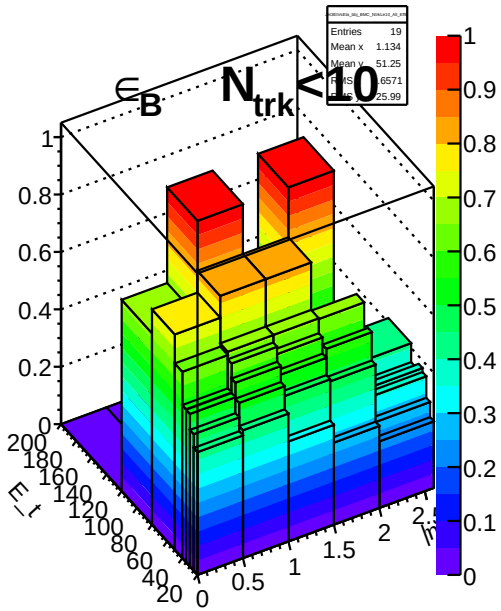
M_{jj} and M_{bb} for different HLT paths

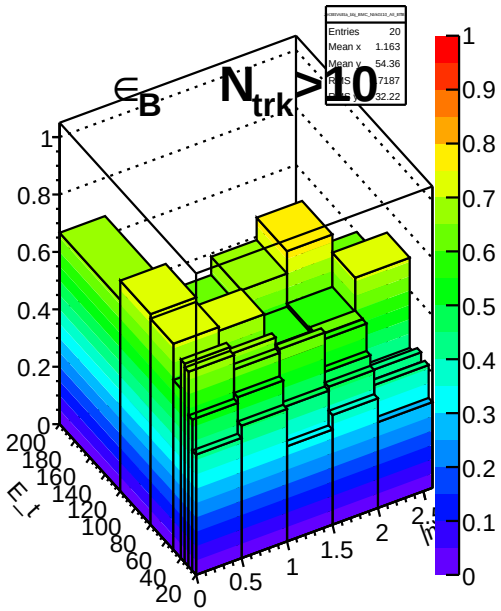


Applying pre-selections (2 b-tag) only: no bias in M_{jj} nor M_{bb}

ϵ_b for bbj vs $|\eta|$, E_t vs N_{trk}



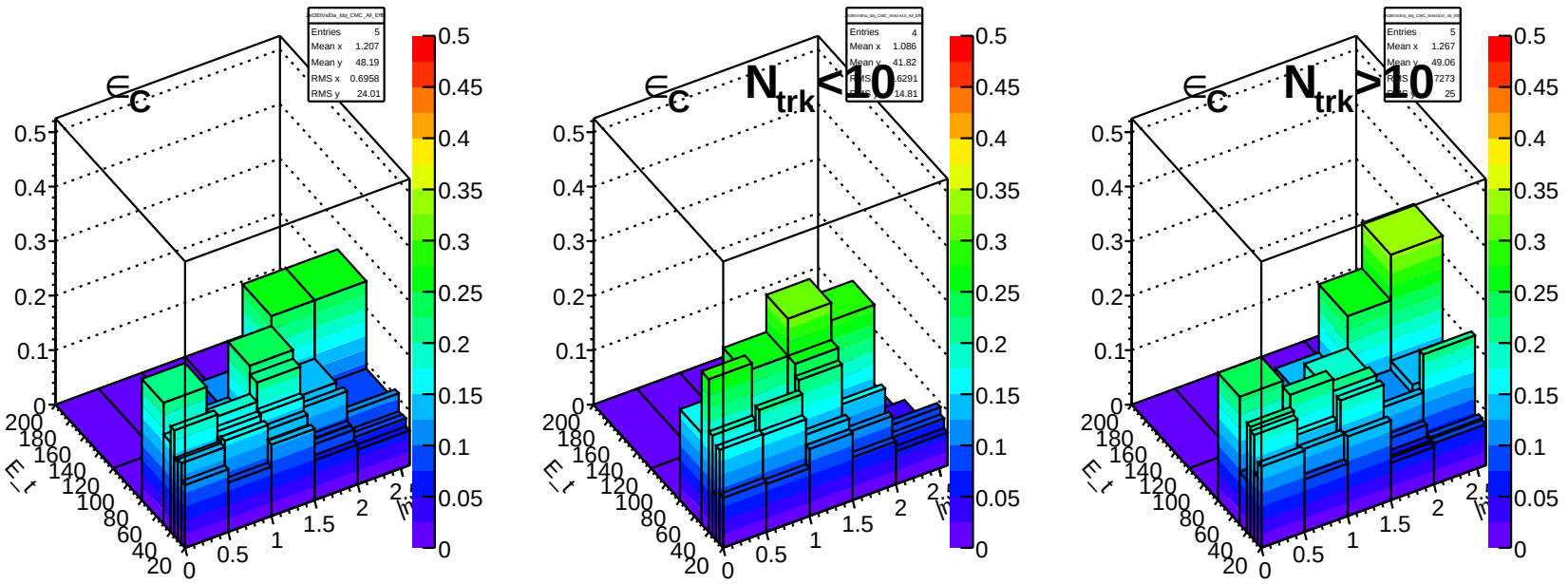




Left to Right:

ϵ_B All, $N_{trk} < 10$, $N_{Trk} \geq 10$

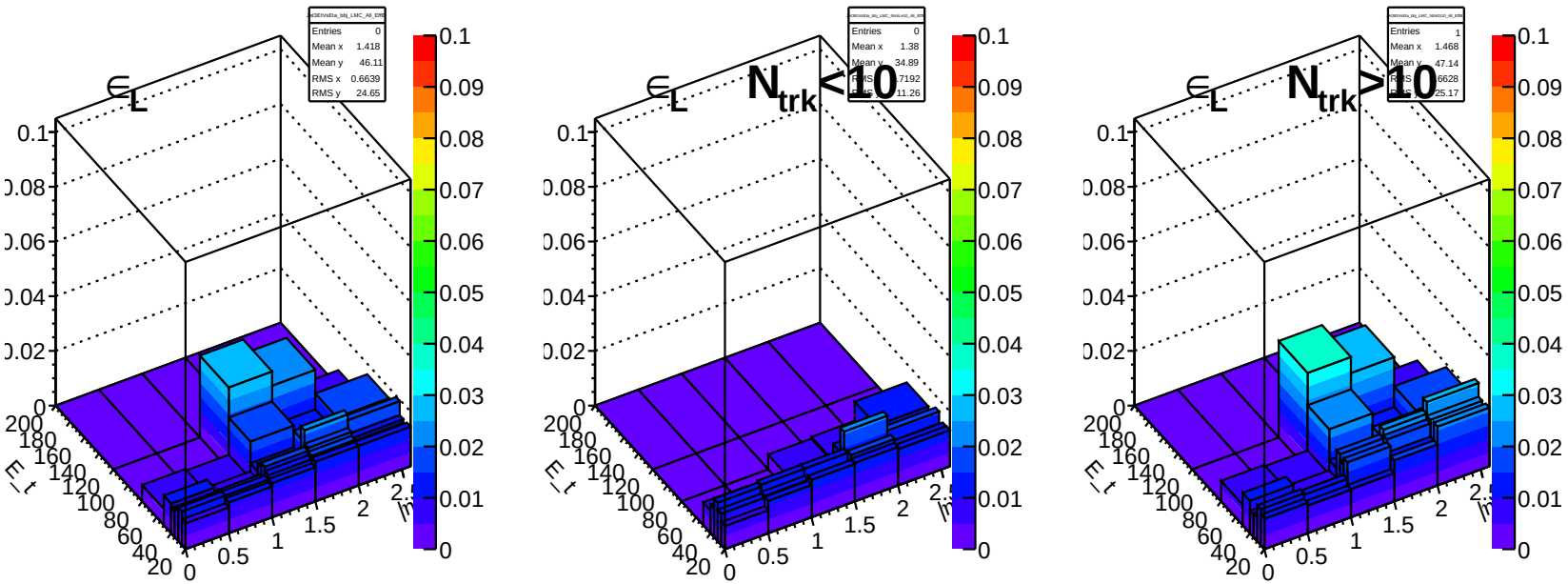
ϵ_c for bbj vs $|\eta|$, E_t vs N_{trk}



Left to Right:

ϵ_C All, $N_{trk} < 10$, $N_{Trk} \geq 10$

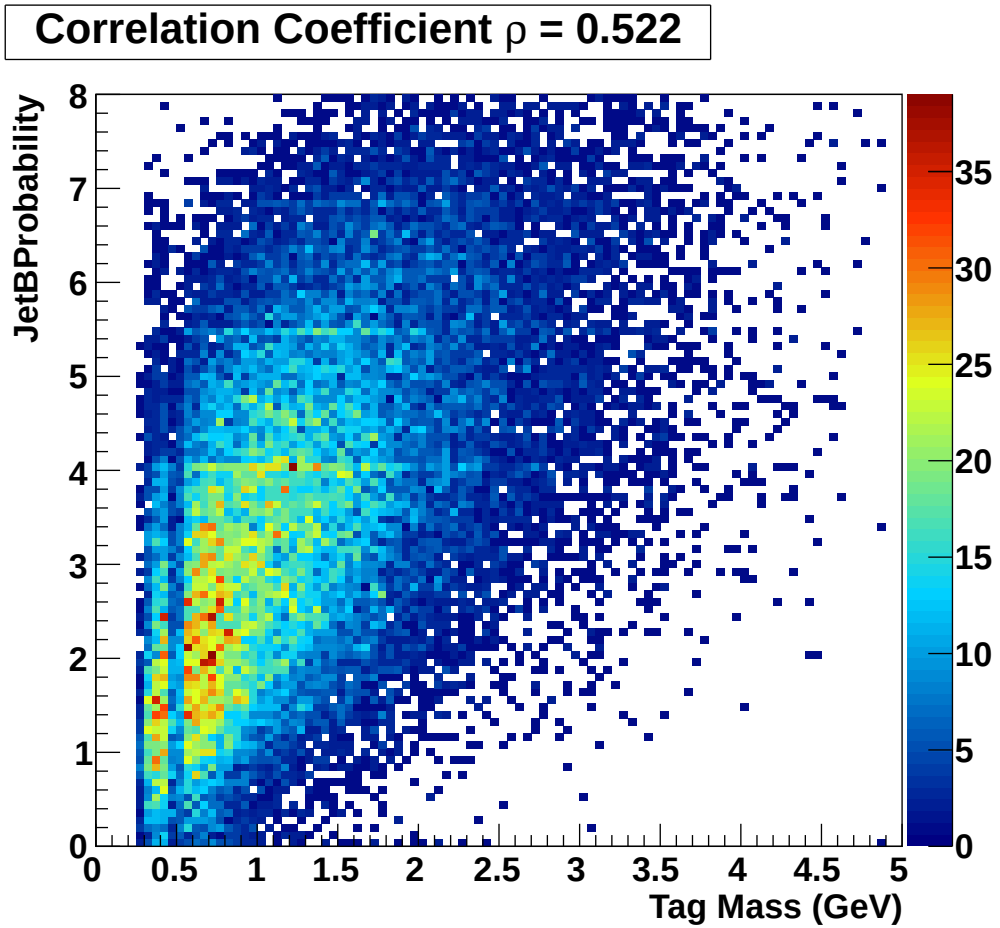
ϵ_{light} for bbj vs $|\eta|$, E_t vs N_{trk}



Left to Right:

ϵ_{Light} All, $N_{trk} < 10$, $N_{Trk} \geq 10$

Mass@Vertex and JetBProbabiliy correlation



$F_{b,c}$ parametrization

$$F_{B,C}\left(E_T^{(j_3)}, \left|\eta^{(j_3)}\right|\right) \times F_{B,C}\left(\Delta R_{H,j_3}, |\Delta R_{j_1,j_2}|\right)$$

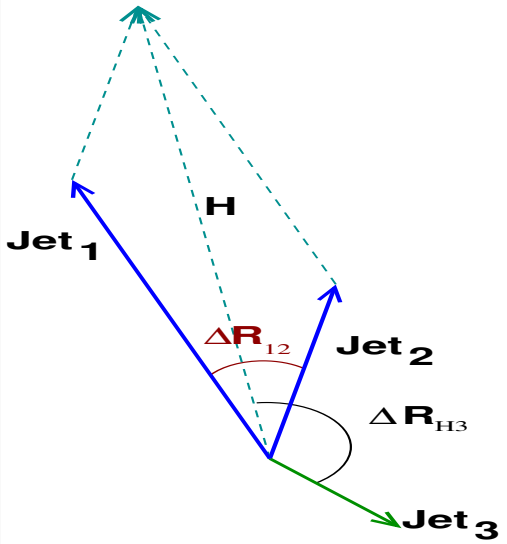
assuming no correlation.

One factor for the third jet features, one for the event topologies.

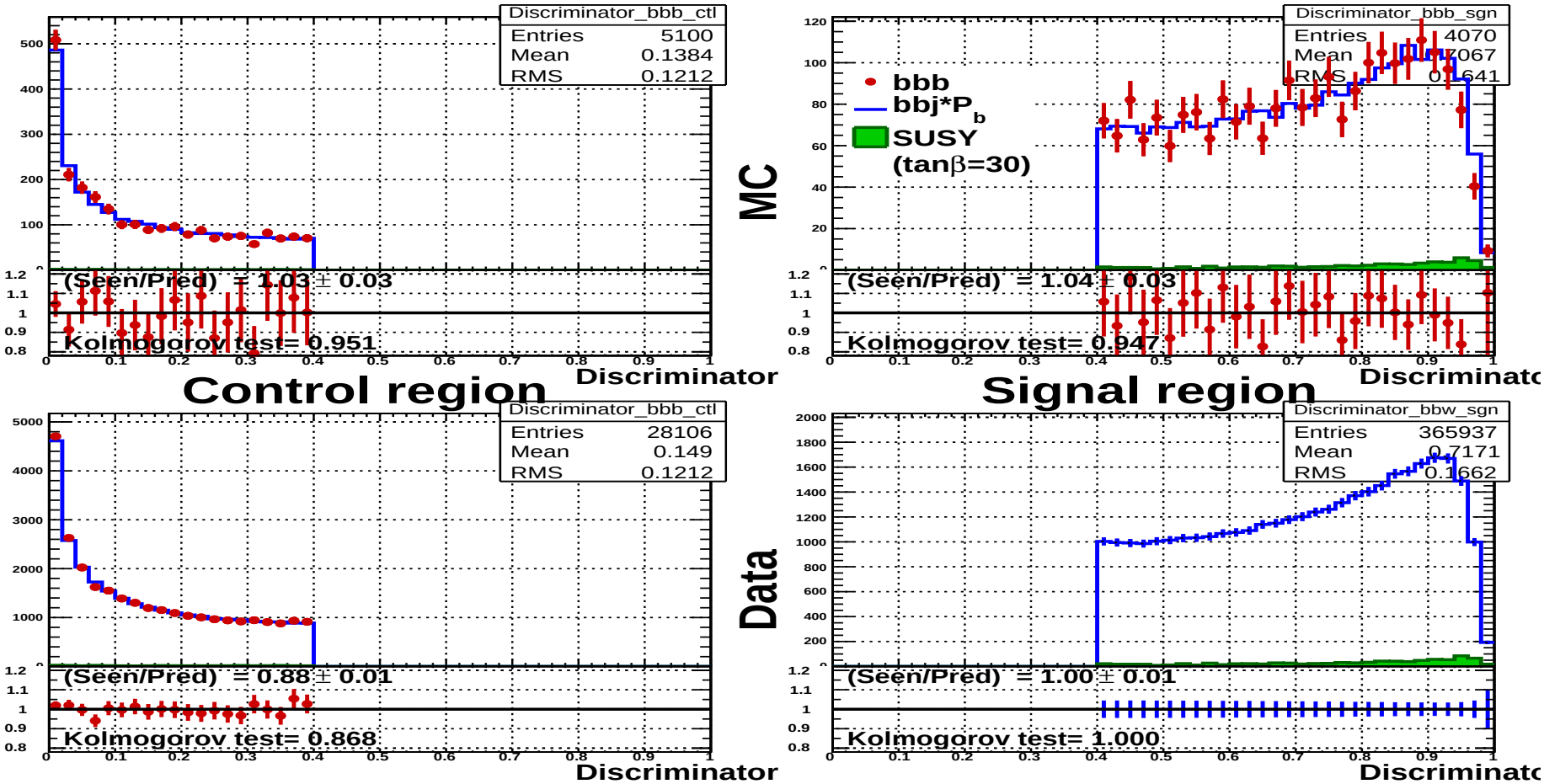
Use $F_{B,C}(\Delta R_{H,j_3}, |\Delta R_{j_1,j_2}|)$ only for shape:

$F_{B,C}$ average, weighted to *bbj* distribution, is normalized to unity.

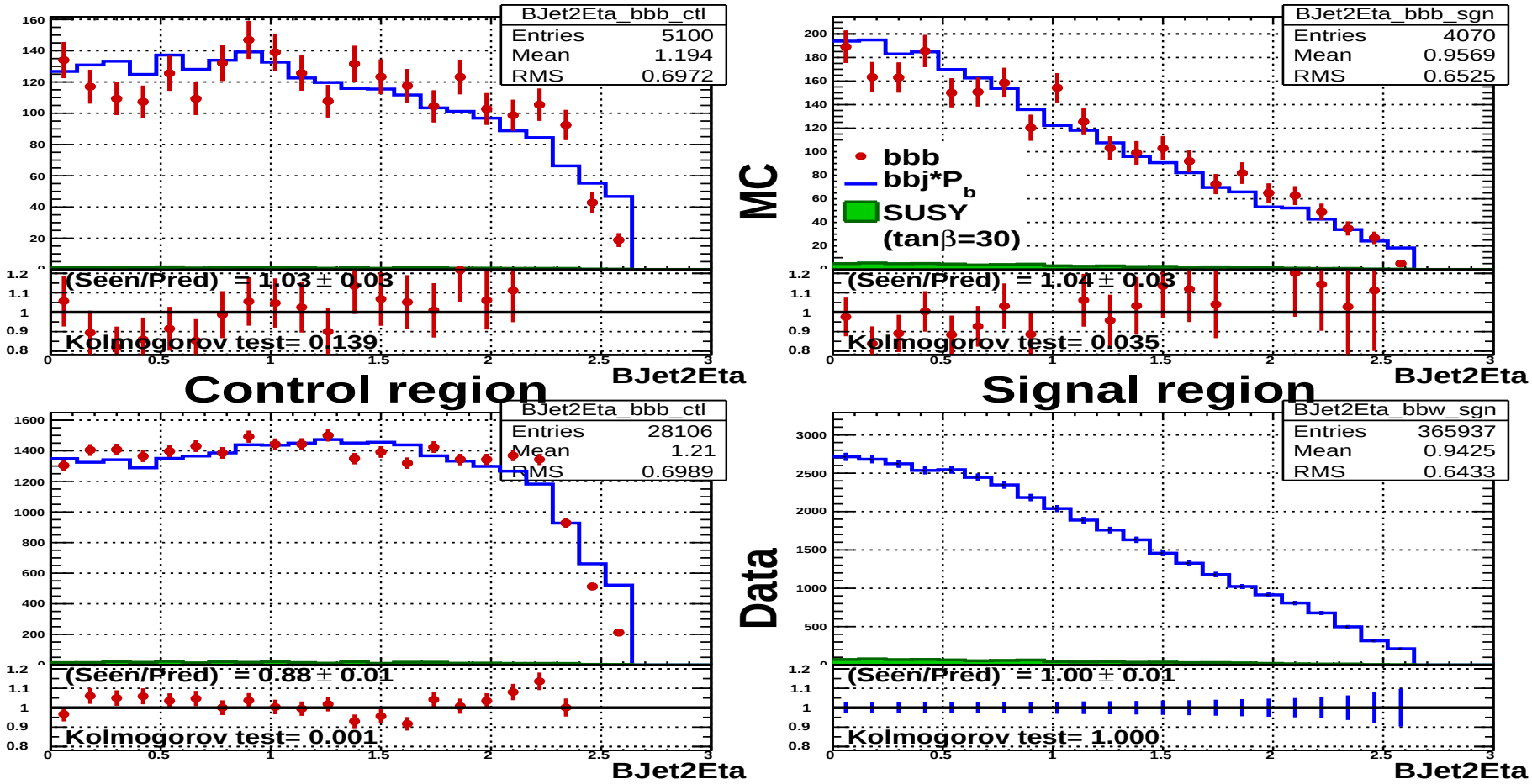
$$\int_{C\ reg.} \frac{dN}{d\Delta R_{12} d\Delta R_{H,j3}} \cdot F_{B,C}\left(\Delta R_{H,j3}, \Delta R_{12}\right) d\Delta R_{12} d\Delta R_{H,j3} = \int_{C\ reg.} \frac{dN}{d\Delta R_{12} d\Delta R_{H,j3}} d\Delta R_{12} d\Delta R_{H,j3}$$



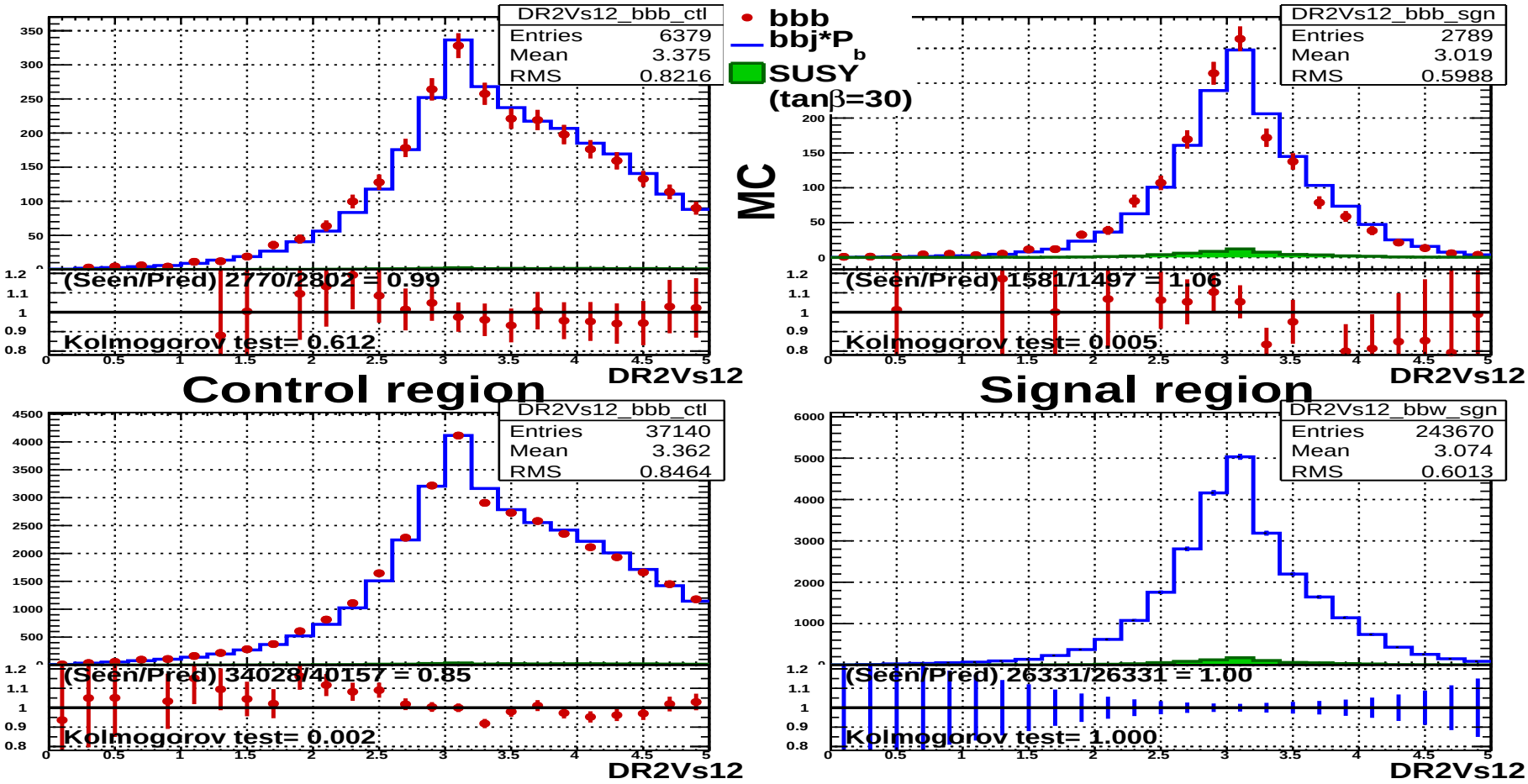
Prediction vs *bbb* in MC and Data



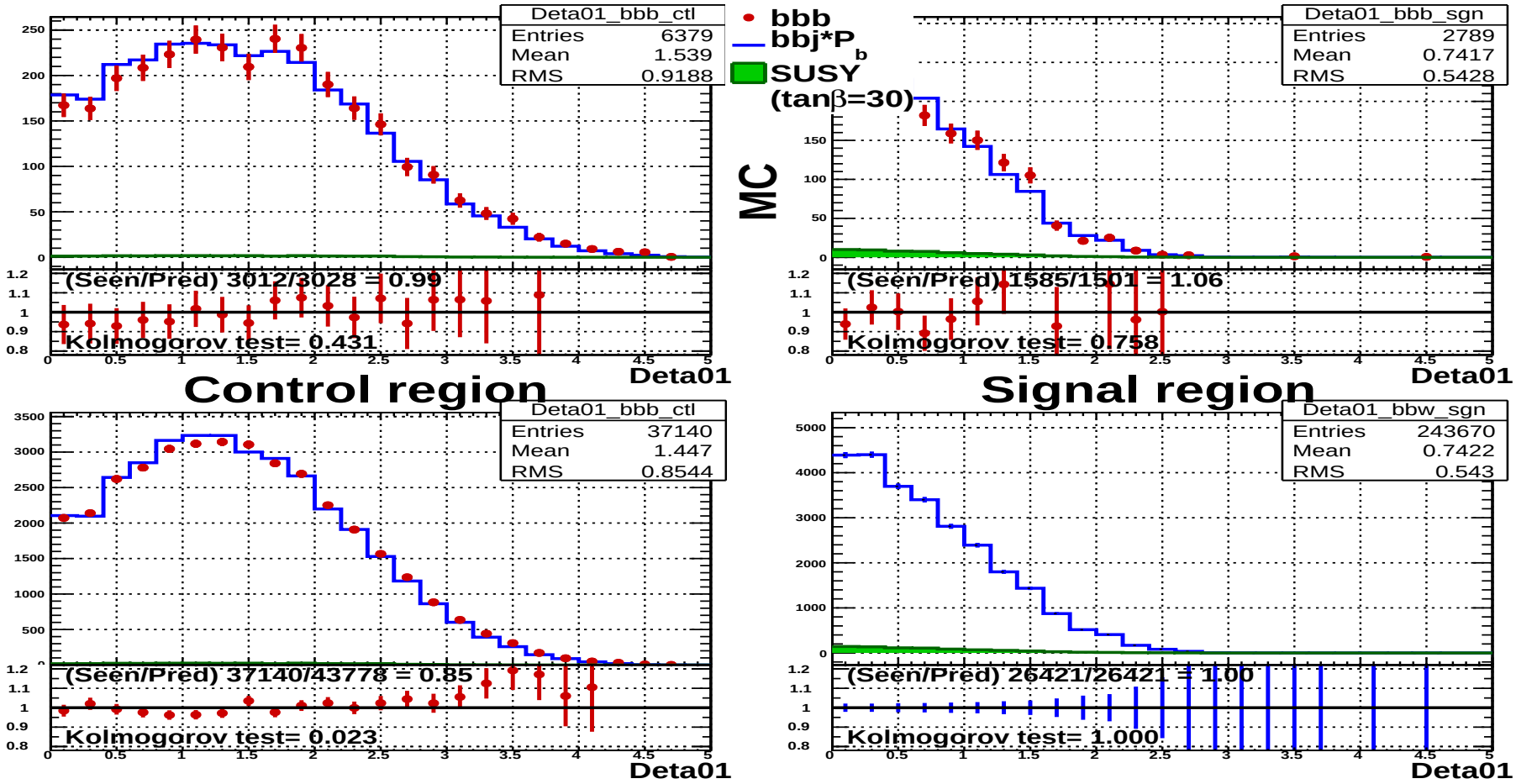
Prediction vs *bbb* in MC and Data



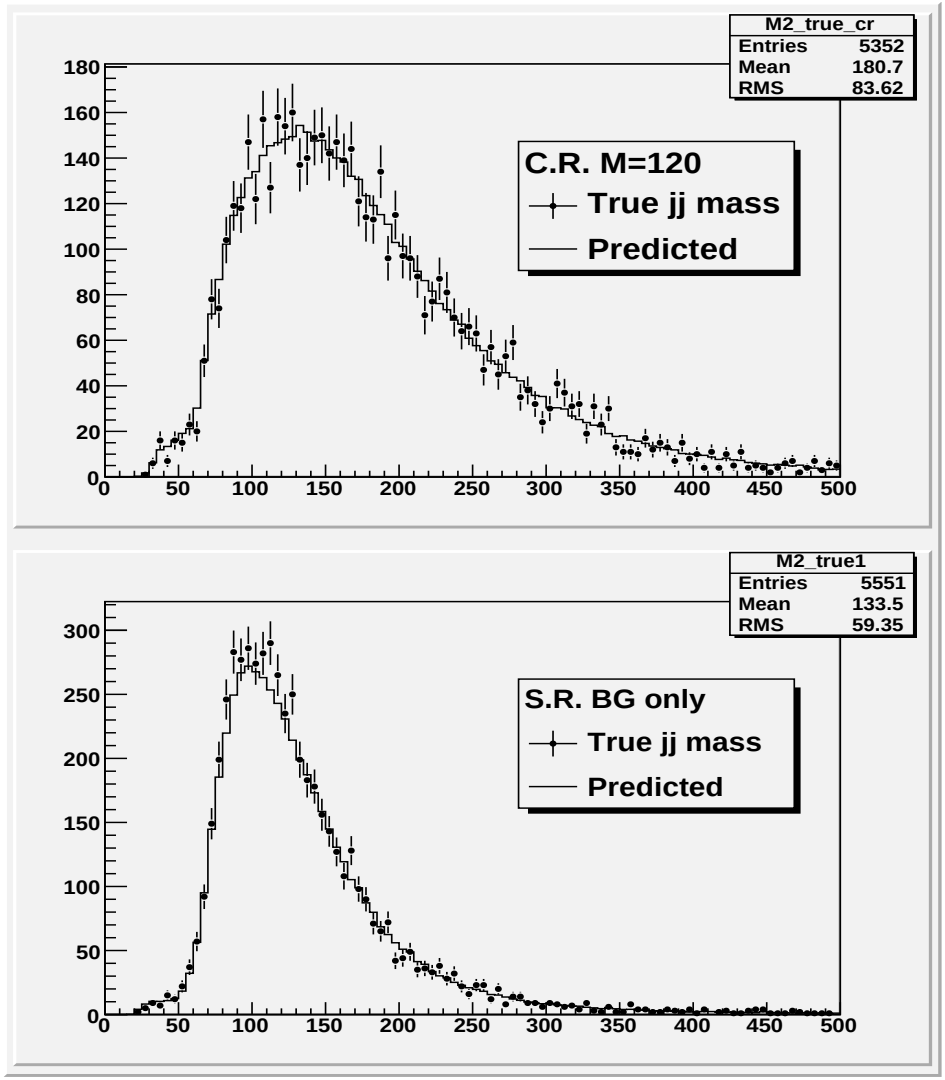
Prediction vs *bbb* in MC and Data

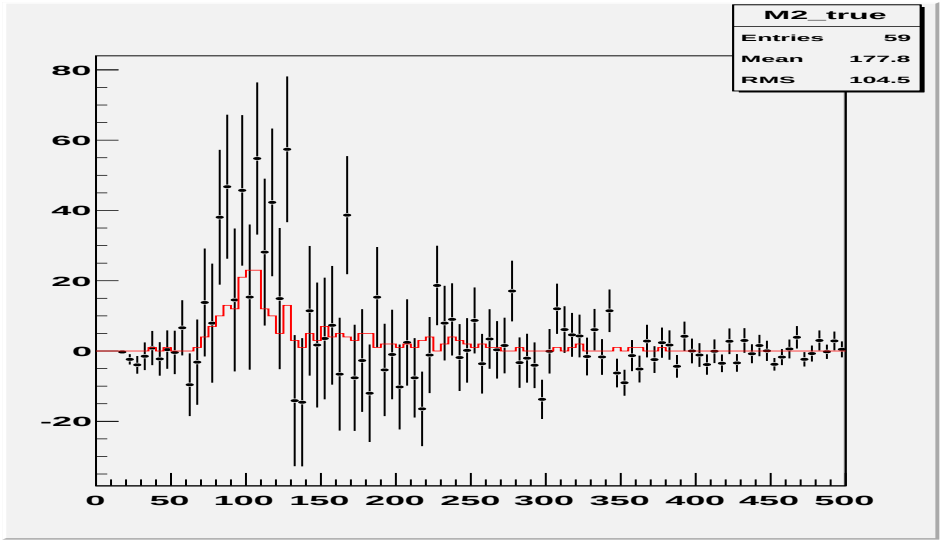


Prediction vs *bbb* in MC and Data



Hyper Ball: prelim results on QCD MC





Example of signal injection and extraction $M_H = 120$ GeV