Semi-leptonic MSSM $bb(H \rightarrow bb)$ status update

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

INFN Padova

Hbb meeting,
Hbb, 23 March 2012

Intro

Intro

- Neutral SUSY Higgs, $H \to b\bar{b}$
- Use associate production to reject hadronic background $pp \to b\bar{b}H \to b\bar{b}b\bar{b}$
- Use semi-leptonic b decay for trigger;
- Baseline selections:
  - at least 1 global muon $P_T^{\mu} > 15\ GeV$;
  - at least 3 jets (PFak5, LooseId) $|\eta| < 2.6$, $P_T > (30, 30, 20)\ GeV$
  - the first 2 jets must have b-tag $CSV > 0.8$, the μ inside one of the two;
  - last selection: third jet b-tag $CSV > 0.7$

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`

Selections

Cut	Mu12_CentralJet30 _BTagIP	Mu12_DiCentralJet30 _BTagIP3D	Mu12_DiCentralJet20 _DiBTagIP3D1stTrack	Mu12_eta2p1_DiCentralJet20 _DiBTagIP3D1stTrack	Mu12_eta2p1_DiCentralJet20 _DiBTagIP3D1stTrack_NewJEC	All
All	3027717	4532555	1873247	4762858	2164634	16361011
$p_{\mathrm{T}}^{\mu} > 15$ GeV	1757902	2678935	1118307	2741523	1223401	9520068
$\#jets >= 3$	665203	1244561	537517	1365228	594070	4406579
$CSV(1^{st} - jet) > 0.8$	309544	631637	310742	774241	335506	2361670
$CSV(2^{nd} - jet) > 0.8$	64072	138772	167319	414414	179859	964436
μ in 1^{st} or 2^{nd} jet	61901	133538	160688	399079	173417	928623
$CSV(3^{rd} - jet) > 0.7$	4295	9744	14552	34306	14685	77582
$\int \mathcal{L} dt$ [pb^{-1}]	183.8	524.9	780.4	1944.5	811.9	4245.5

Strategy

Get QCD background from data

- Major background source is QCD
  - second source $t\bar{t} + jet(s)$ estimated (MC) roughly 500 ev;
- Define a control region using a likelihood ratio discriminator using the most discriminating variables (depends on M_H)
- Build B-tagging matrixes for $bb + j$ sample in control region;
- Estimate bbb background in signal region starting from bbj:

$$F(bbb) = F(bbj) \times P_b^{3rd-j}(j)$$

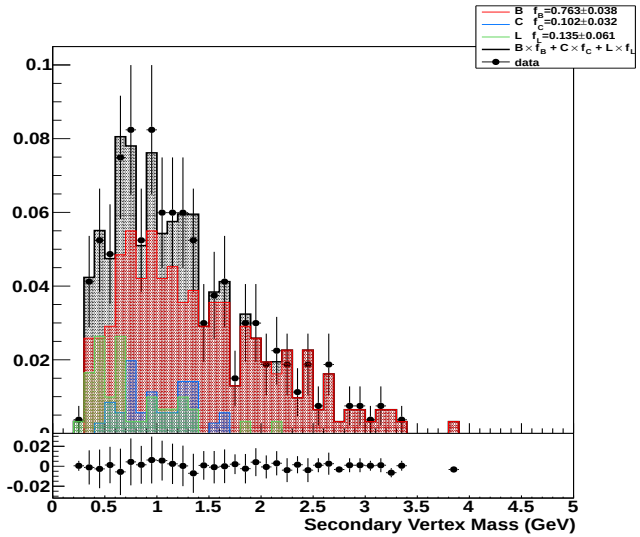
$$P_b^{3rd-j}(j) = \epsilon_b \cdot f_b + \epsilon_c \cdot f_c + \epsilon_l \cdot f_l$$

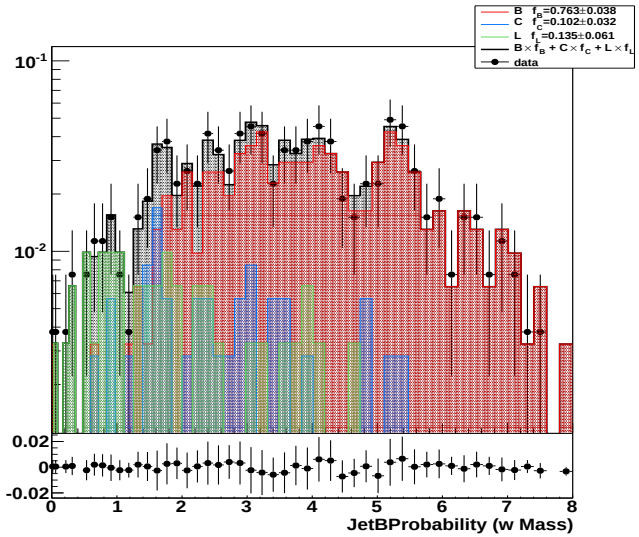
- Get $\epsilon's$ from MC and $f_{b,c,l}$ from Data;
- Compare single distribution (M_{bb}) or MVA variable;

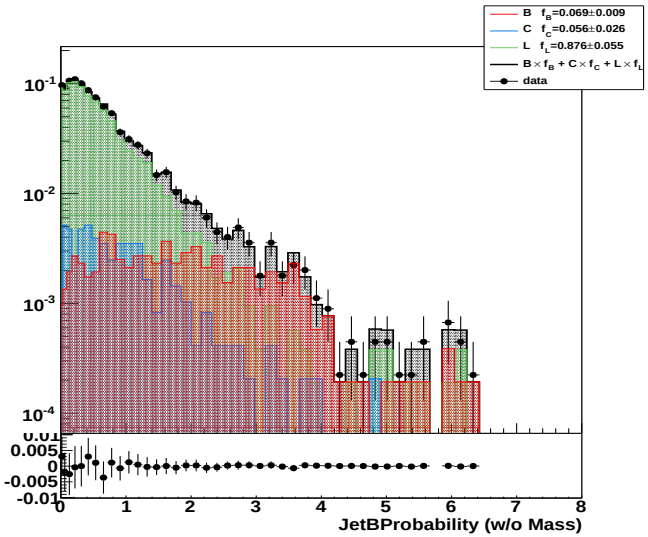
Get $F_{b,c}$

Consider Mass@Vertex and JetBProbabilty (new) for third jet; (before: TCHP. Tried also JetProbability;)

- build templates from MC QCD;
- fit templates and get $F_{b,c}$;
- used only JetBProbabilty if Mass@Vertex not available.

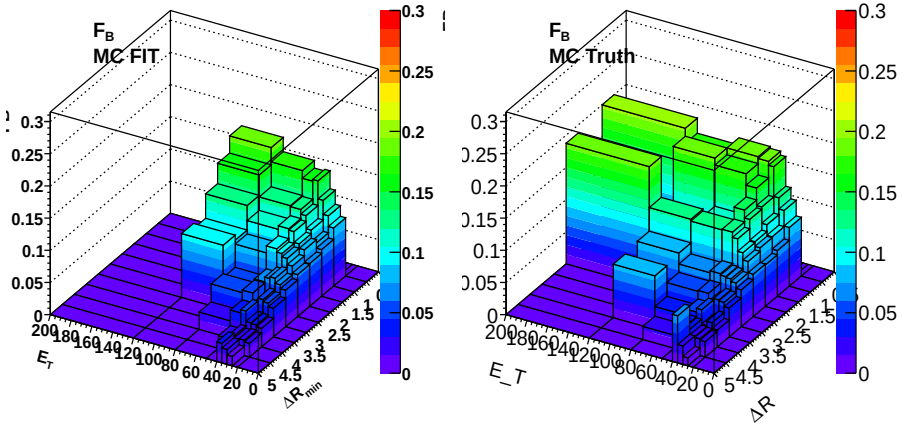




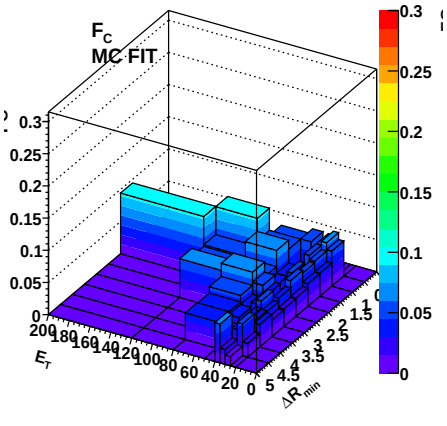


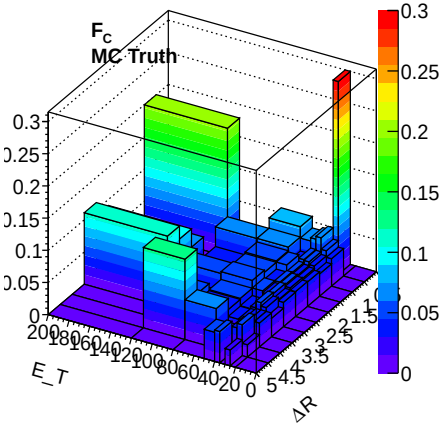
$F_{b/c}$ in QCD MC: Fit vs MC truth

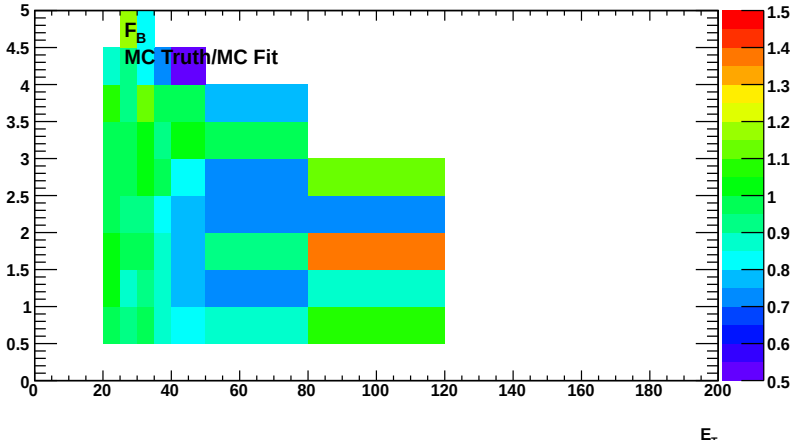
F_b vs $\left(\Delta R_{min}, E_T^{3rd-j}\right)$

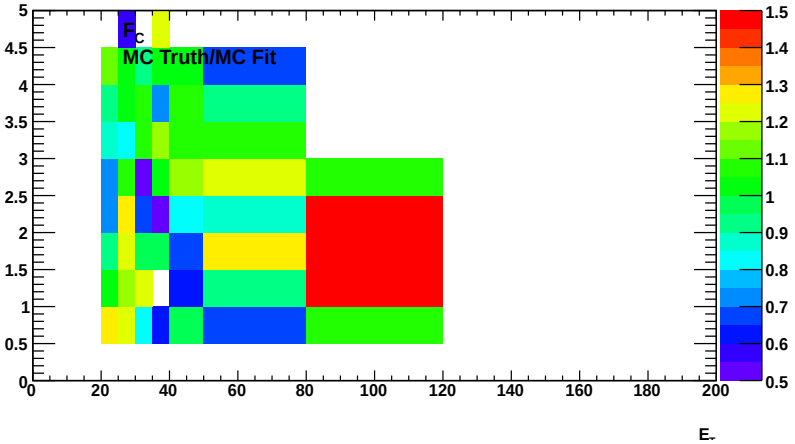


F_c vs $\left(\Delta R_{min}, E_T^{3rd-j}\right)$



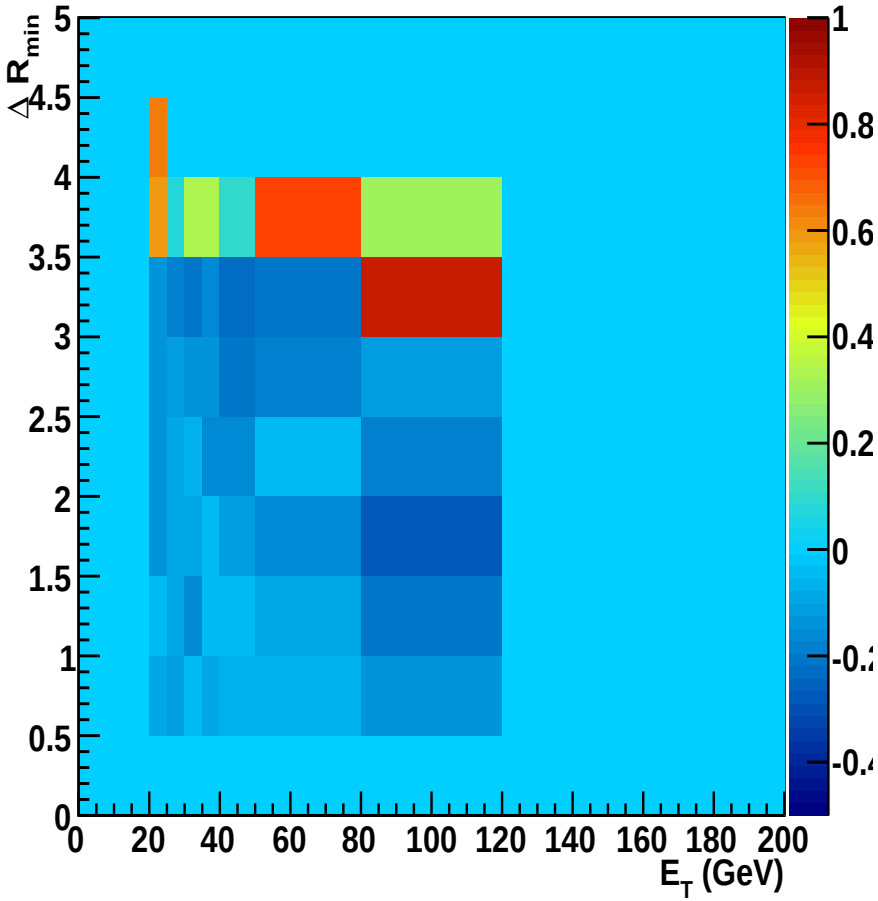




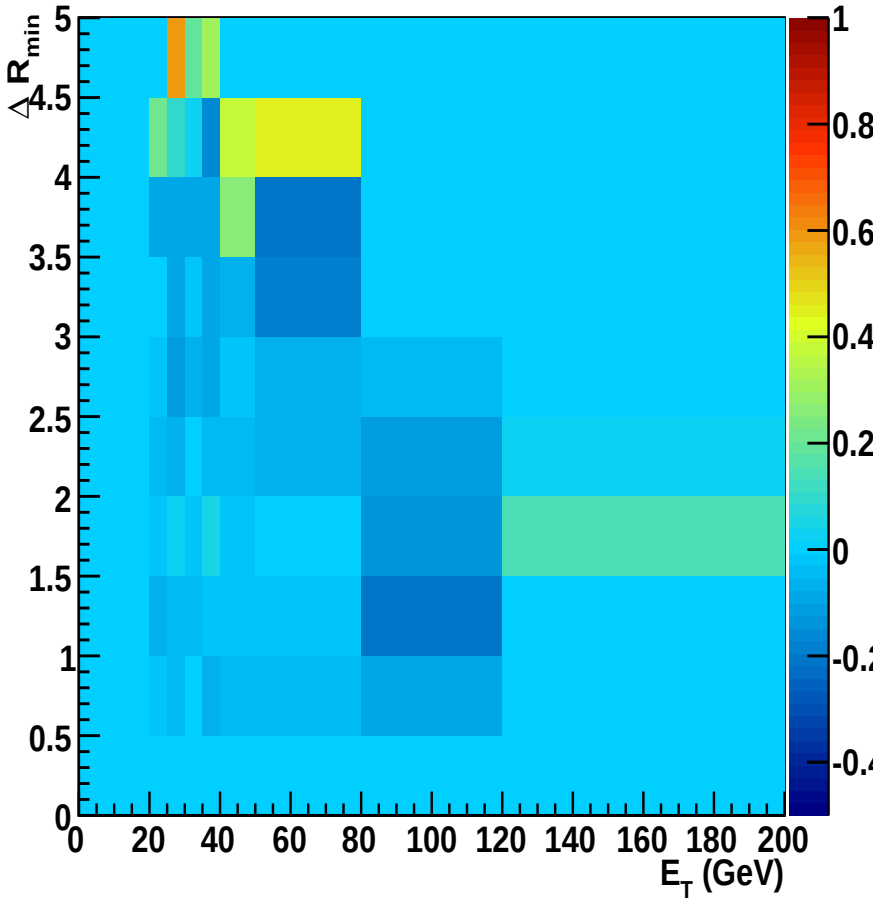


$(F_b^{Fit} - F_b^{True})/F_b^{True}$

Fit Mass@Vtx & JetBProbability



Fit only JetBProbability



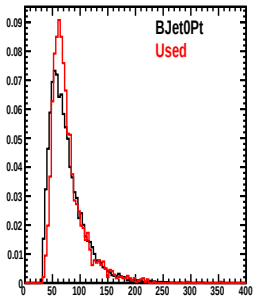
Possible (small) bias?

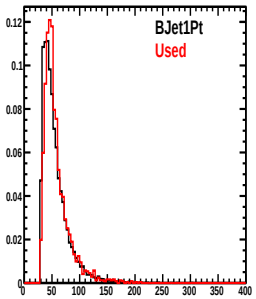
Discriminator optimization

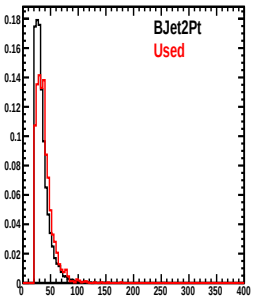
$$Discr = \frac{\Pi_i p_i^{(signal)(x_i)}}{\Pi_i p_i^{(signal)}(x_i) + \Pi_i p_i^{(QCD)}(x_i)}$$

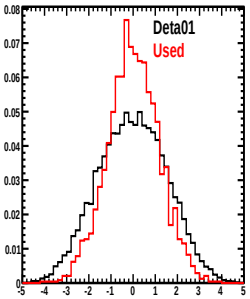
- For $M_H = 120$ GeV we compared different set of discriminator's variables, and see the effect on the shape of M_{bb}.
- in particular check the effect of inclusion of invariant mass in the discriminator.
- Higgs Mass 120 GeV (σ for $\tan\beta = 30$).
  - Discriminator 0 baseline so far, not optimized
  - Discriminator 1 optimized, w/ $M_{inv}(jj)$
  - Discriminator 2 same as 1 but w/o $M_{inv}(jj)$

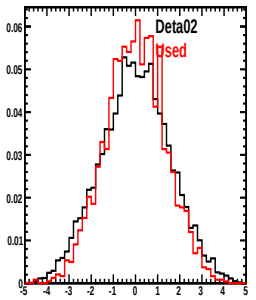
Discriminator 0: baseline so far

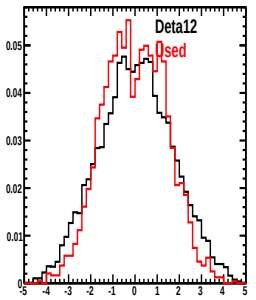


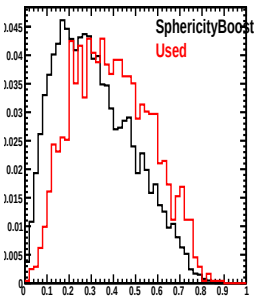


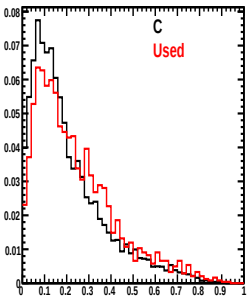


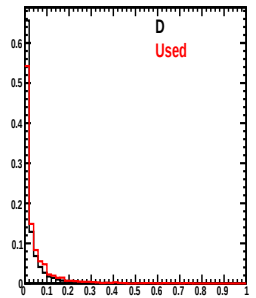


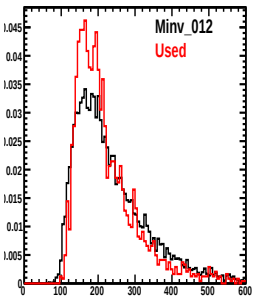


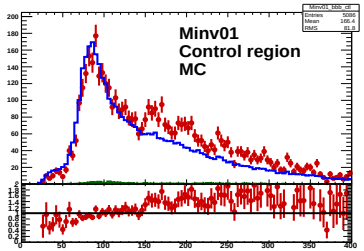


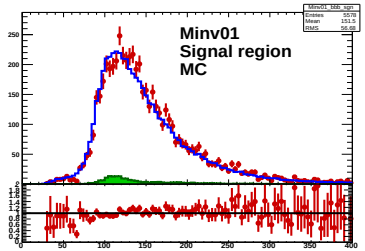


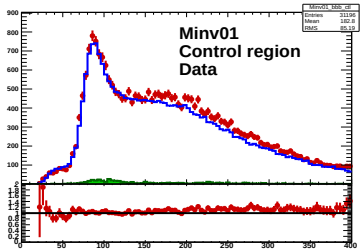


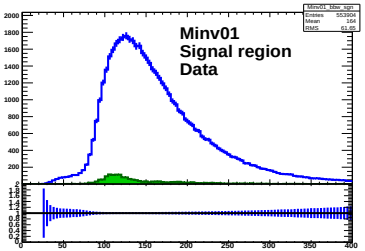






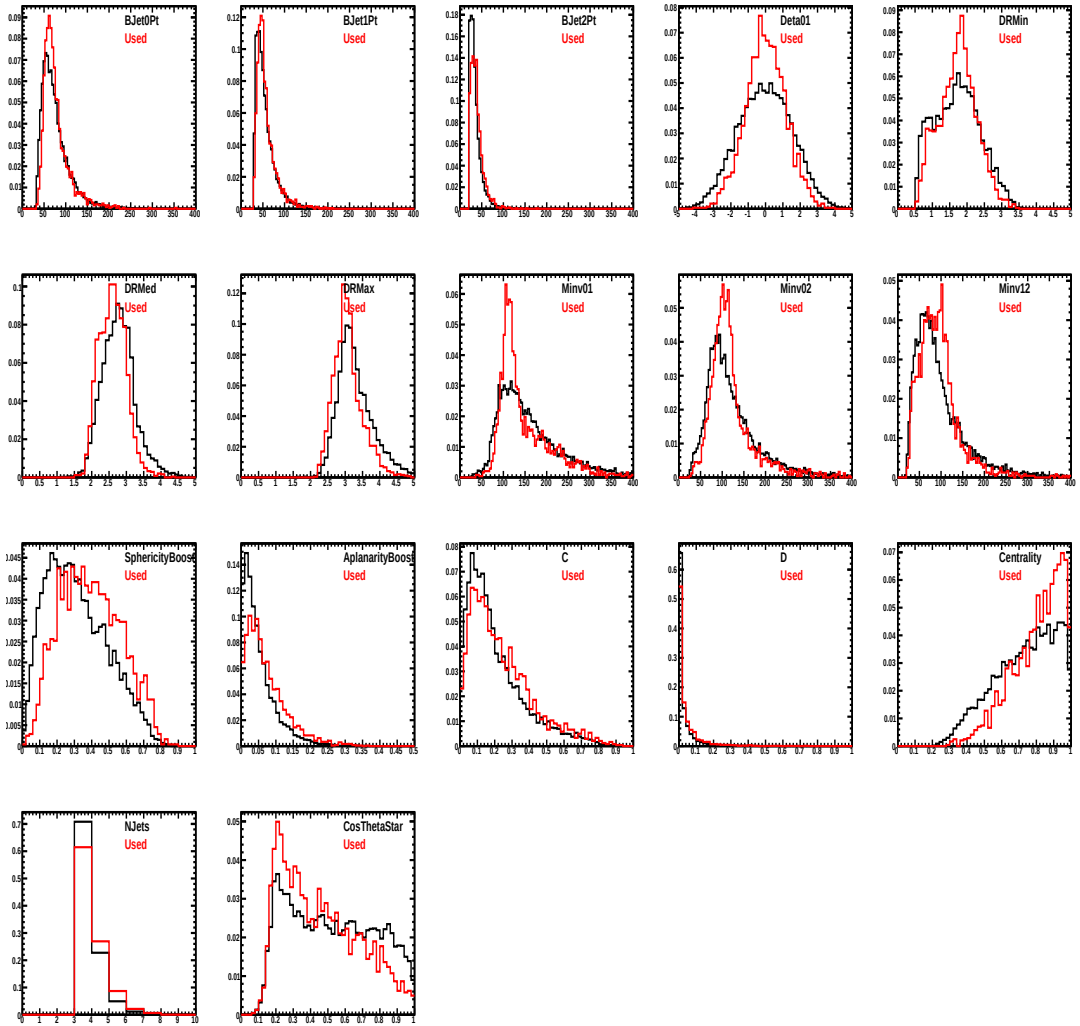


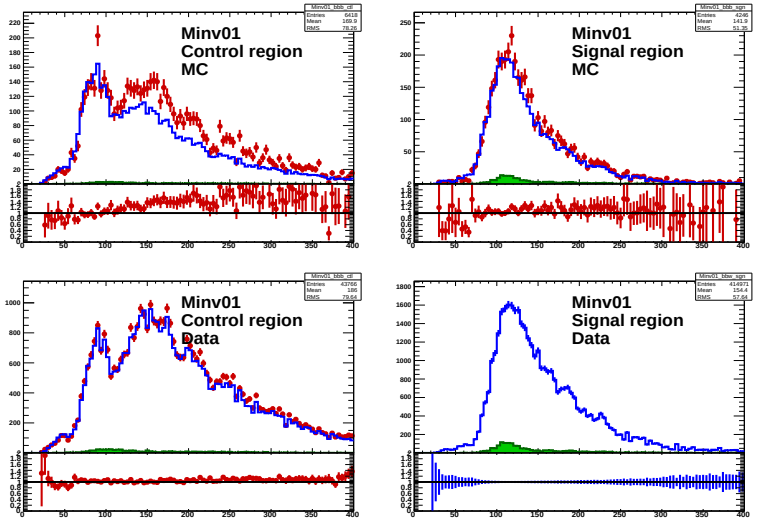




- peak at $M_{bb} \approx 90$ GeV in control region;
- underestimate in prediction for higher M_{bb} both data & MC.

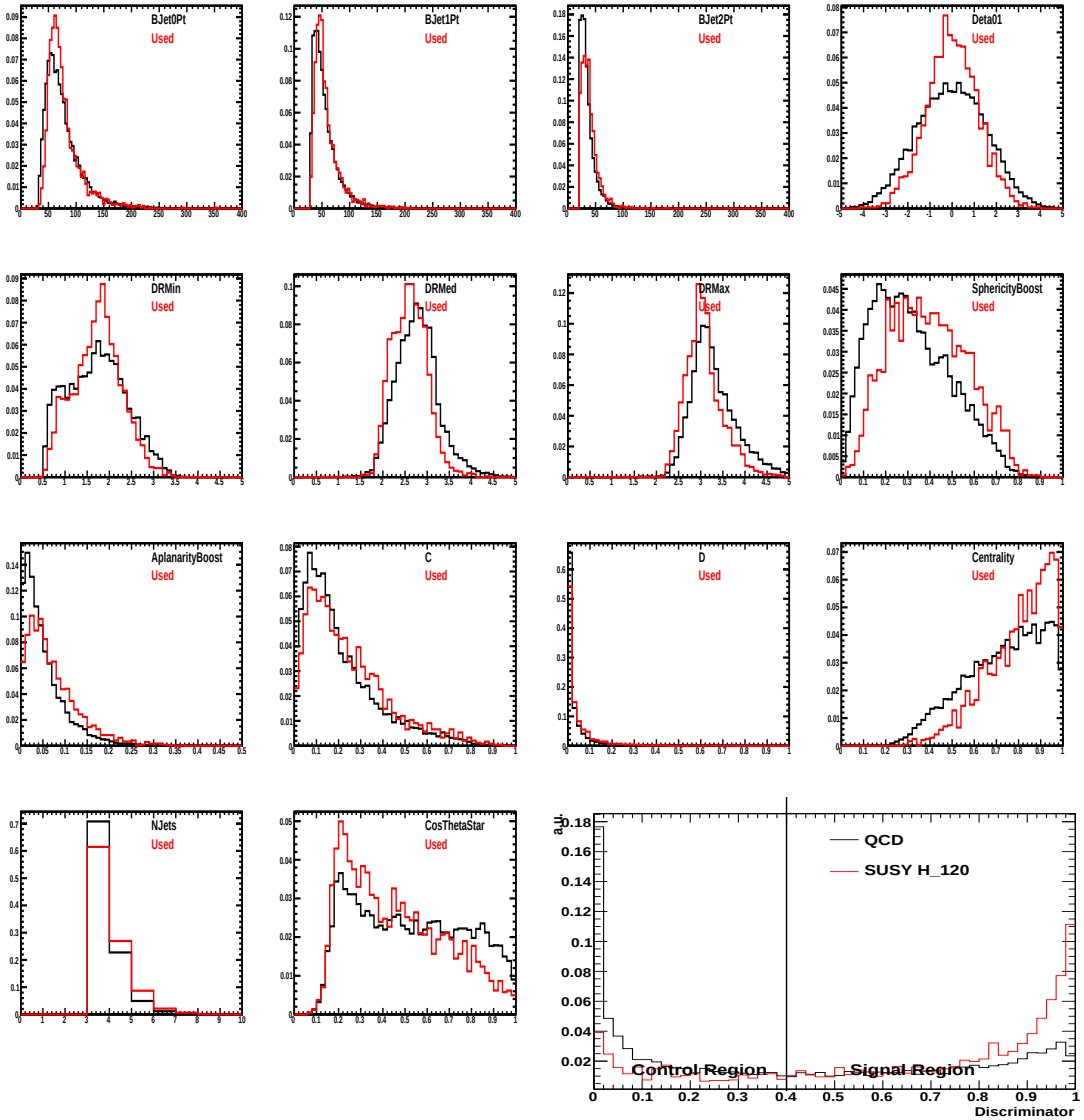
Discriminator 1: optimized with $M_{inv}(jj)$

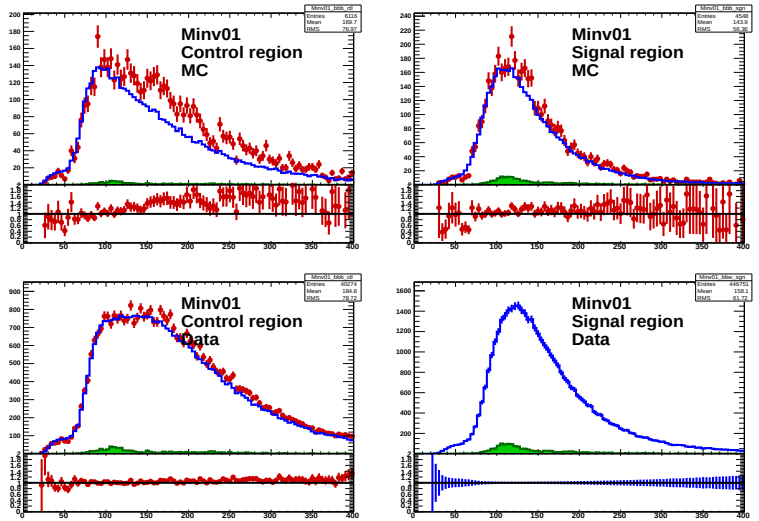




- better signal/QCD discrimination;
- very funny structure for M_{bb};
- underestimate in prediction for higher M_{bb} only in MC.

Discriminator 2: optimized w/o $M_{inv}(jj)$





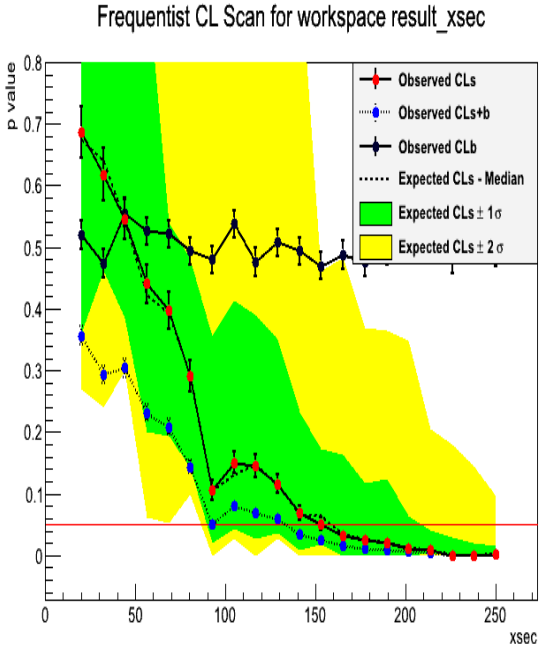
- still good signal/QCD discrimination;
- much nicer shape for M_{bb};
- underestimate in prediction for higher M_{bb} only in MC.

Comparing CLs for different discriminators

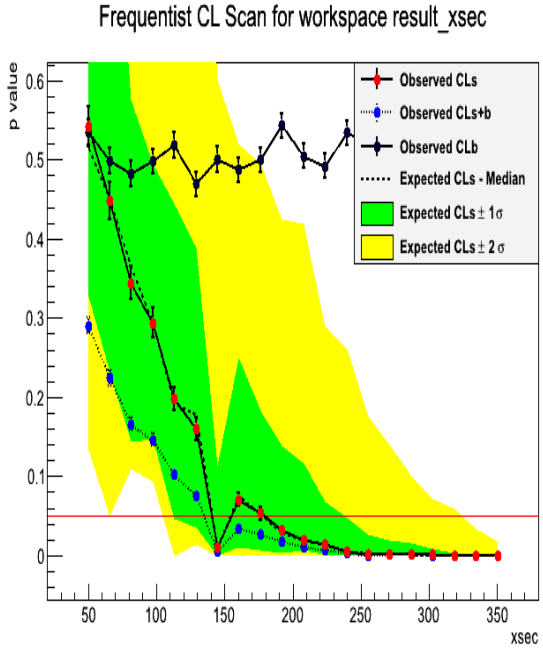
CLs calculated as shown by Antonio on 24/2/2012 (M_{bb} shape analysis, assuming $\left(\int \mathcal{L}dt\right)_{sys} = 4.5\%$, $Bkg_{syst} = 5\%$ $Eff_{syst} = 5\%$)

observed datasample is the predicted one

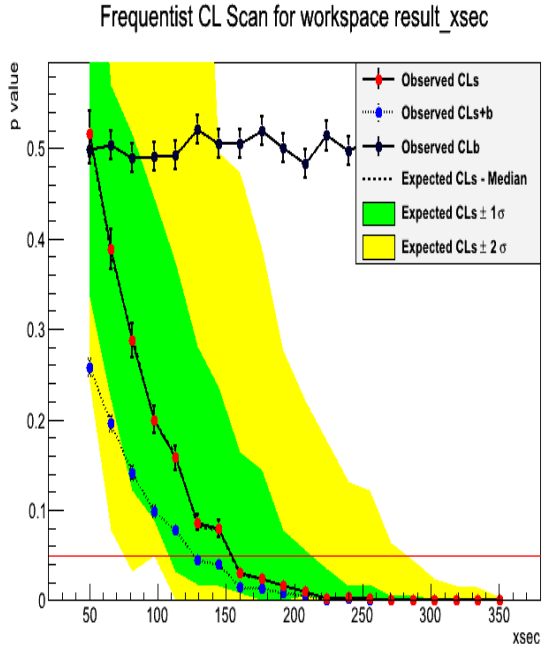
Discriminator 0



Discriminator 1

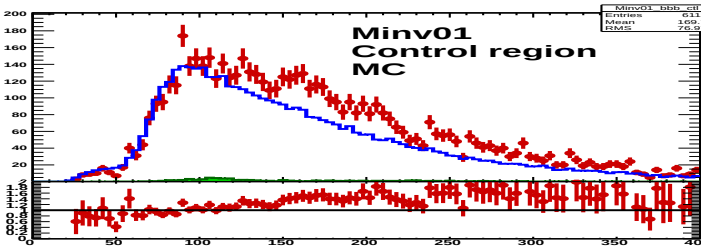


Discriminator 2



Optimized discriminator w/o invariant masses seems the better one.

So, what's wrong in the MC prediction?



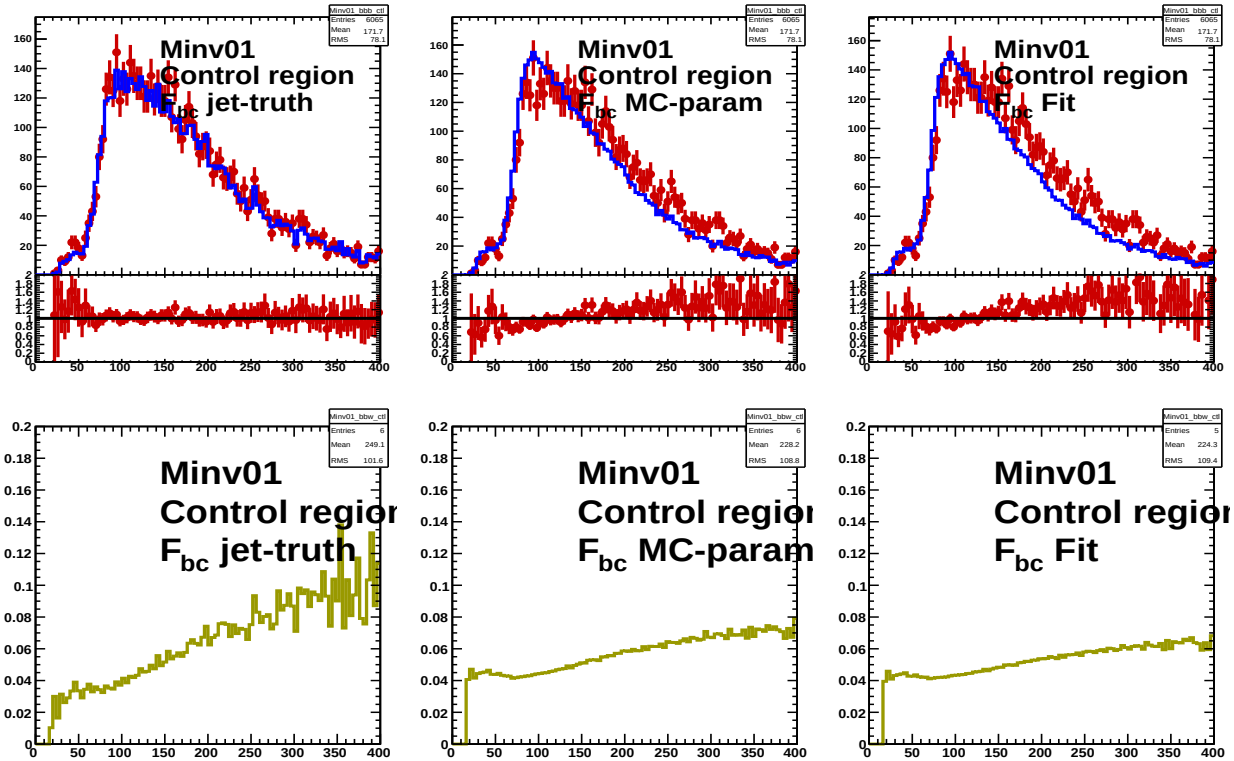
Improvement

- code review (and fix of a set of bugs, of course);
- move to Fall11 production (previously Summer11);
- get $F_{b,c}$ from control region only;
- try ϵ_X only, namely take jet flavour from MC truth jet per jet;
- try different parametrization of $F_{b,c}$ ($E_T, \Delta R_{min}$ vs E_T, η)
- try flat $F_{b,c}$ (pure numbers)
- etc etc ...

After all this work

Look only at MC, control region

Top: $(bbj) \cdot P_b^{3rd-j}(j)$ vs (bbb). **bottom:** just $P_b^{3rd-j}(j)$



$$\frac{(bbb)}{(bbj) \cdot P_b^{3rd-j}(j)}$$

JetTruth

$5788/5581.15 = 1.03706$

MCTruth

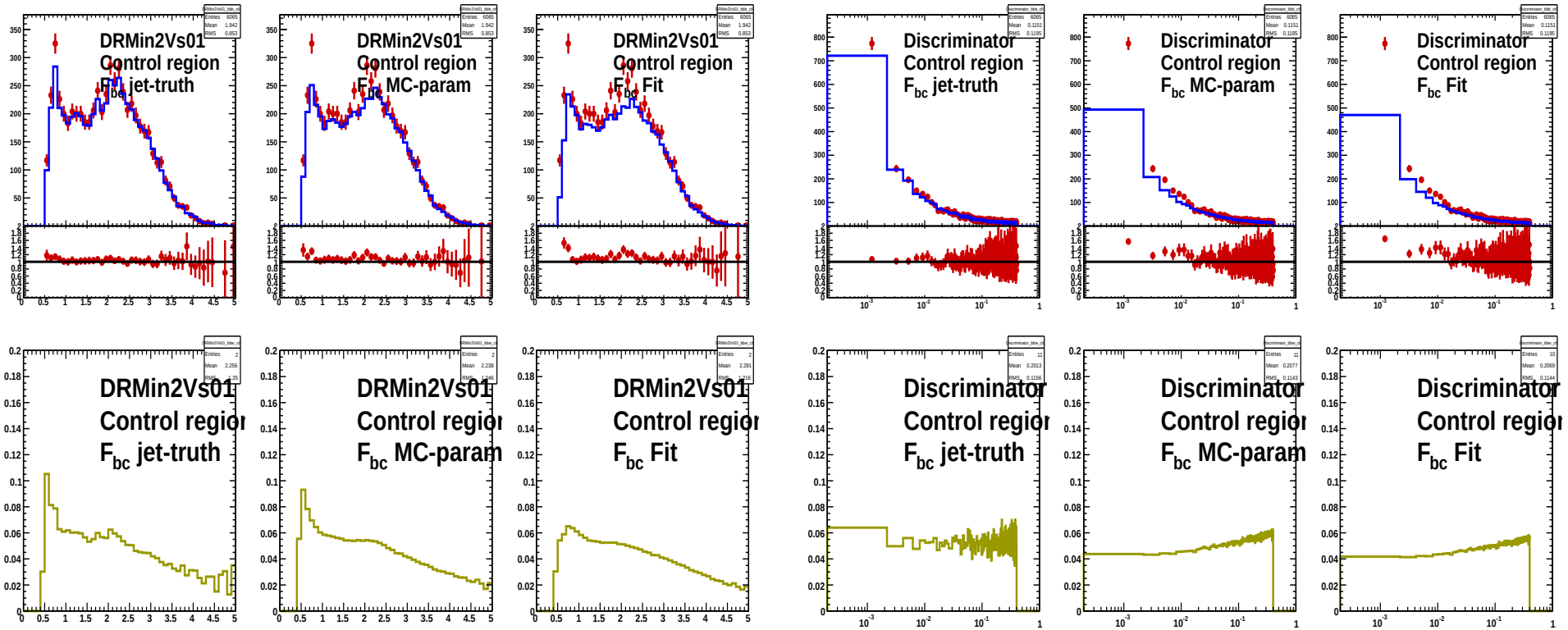
$5788/5405.13 = 1.07083$

Fit

$5788/5127.58 = 1.1288$

Better than before but still problems

Where it comes from?



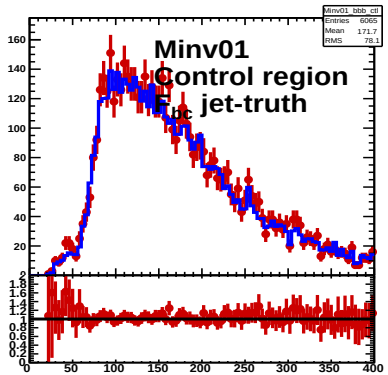
problem at low ΔR_{min} trying finer binning.

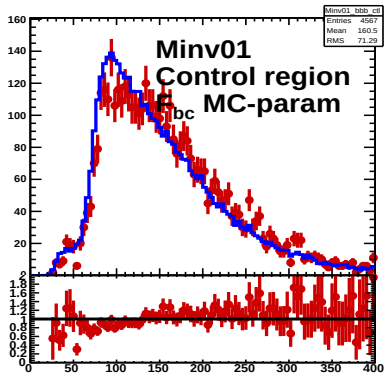
problem at low discriminatore value: try to esclude these events?

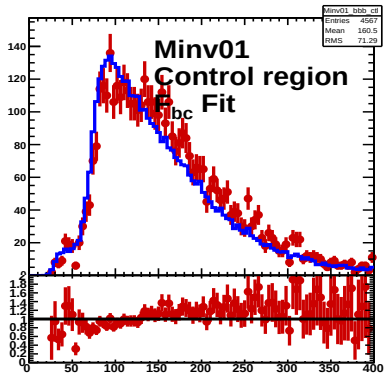
Exclude events with *Discriminator* < 0.01

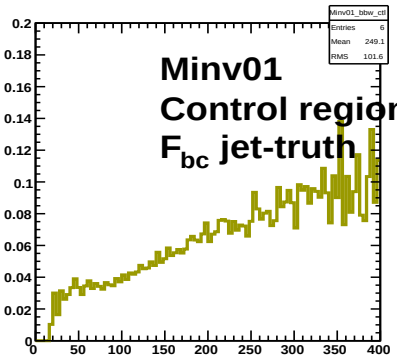
Look only at MC, control region

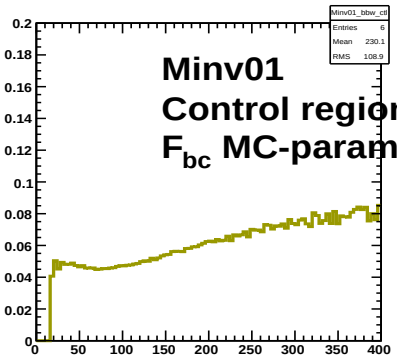
Top: $(bbj) \cdot P_b^{3rd-j}(j)$ vs (bbb). **bottom:** just $P_b^{3rd-j}(j)$

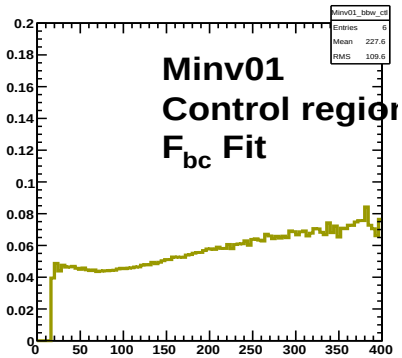












$$\frac{(bbb)}{(bbj) \cdot P_b^{3rd-j}(j)}$$

JetTruth (as before)

$5788/5581.15 = 1.03706$

MCTruth

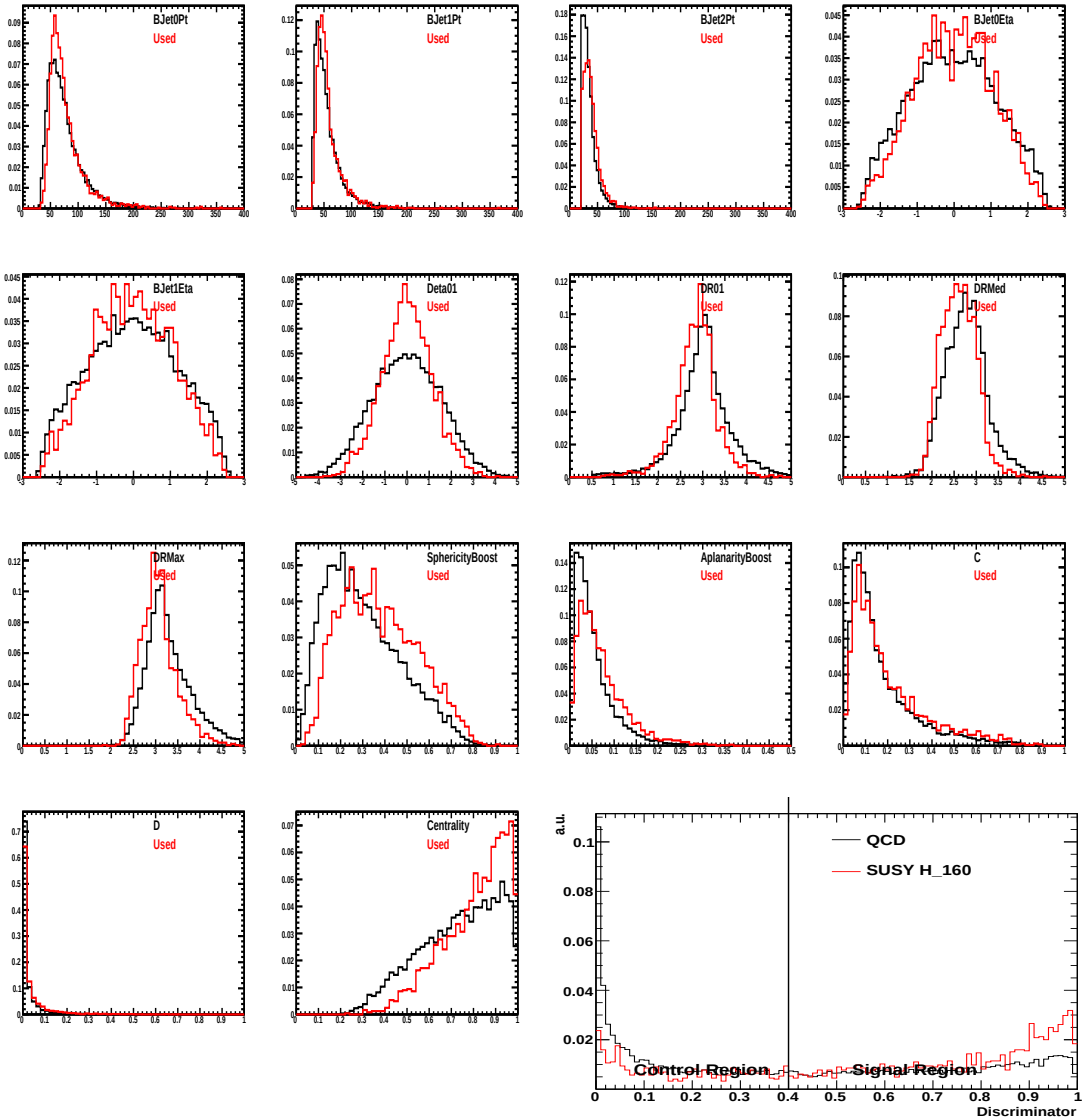
$4567/4460.24 = 1.00353$

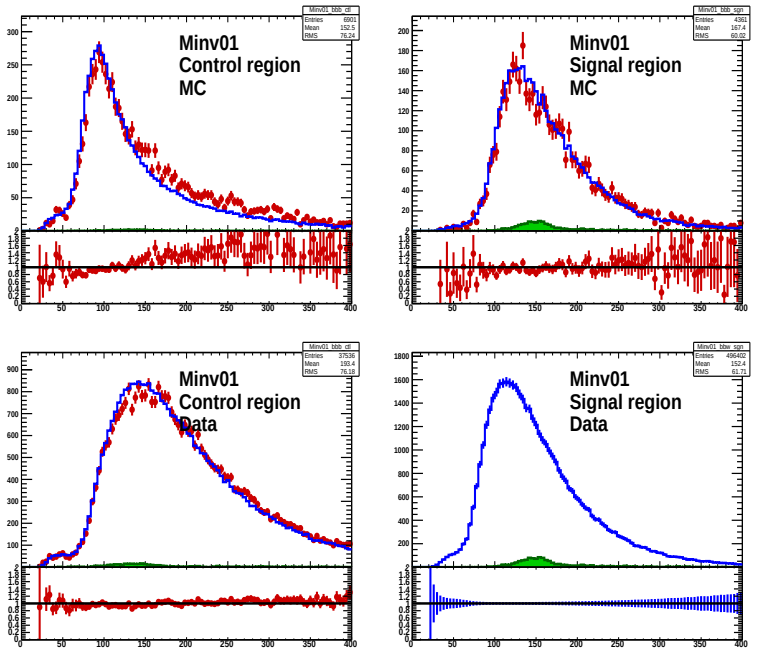
Fit

$4567/4215.31 = 1.06184$

Better than before, but trowing away 20% of events in control region...

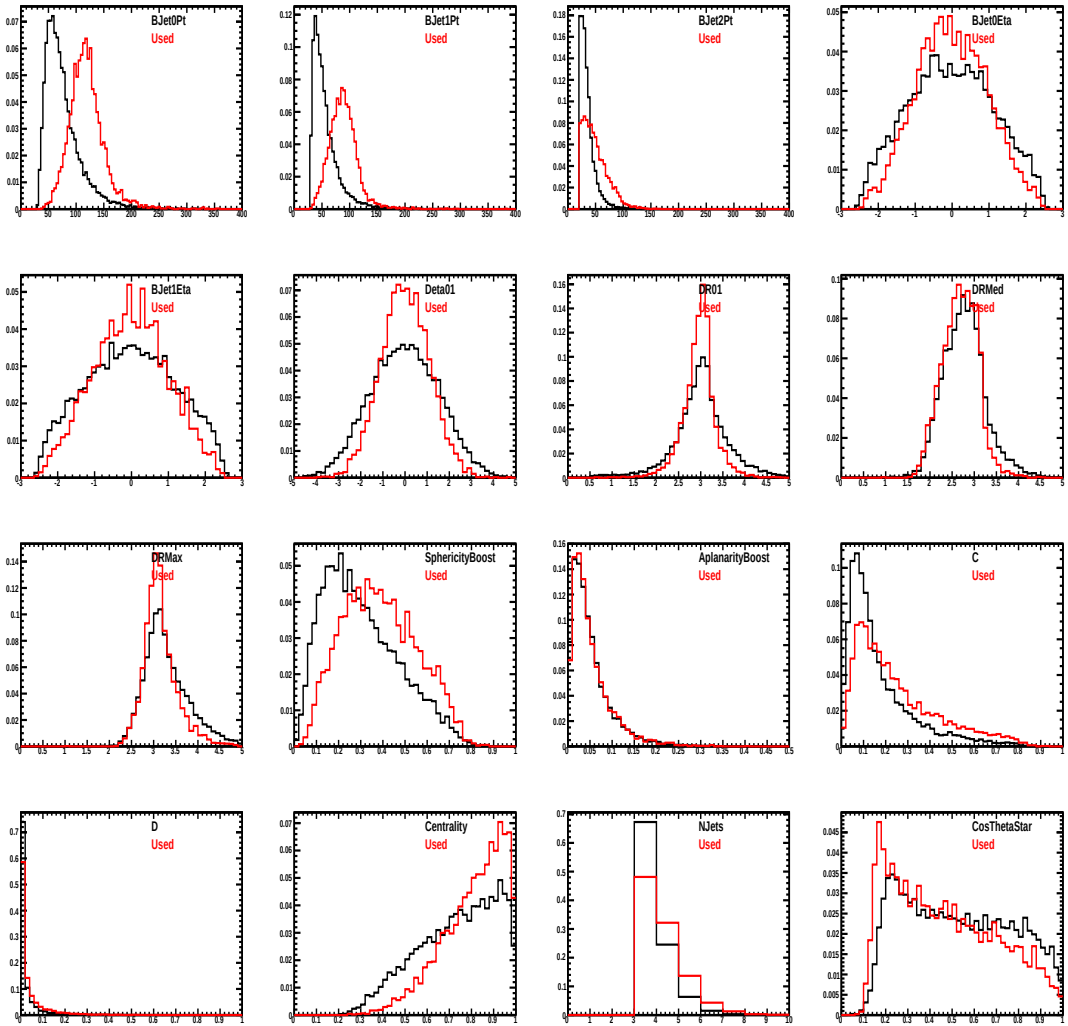
Try other mass points: $M_H = 160$

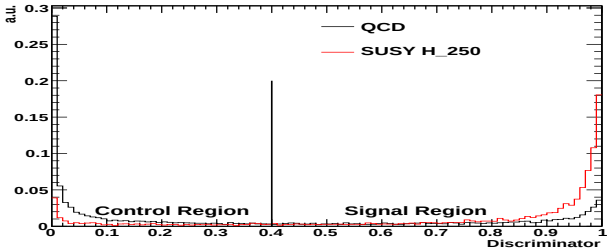


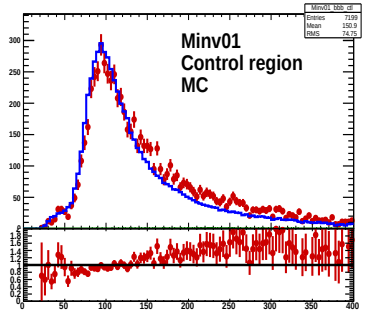


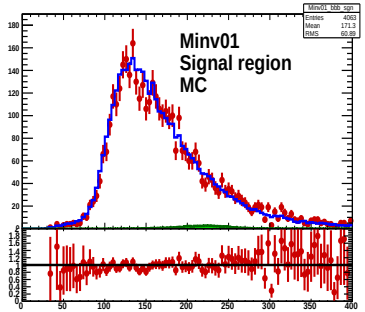
- Quite similar to $M_H = 120$ GeV

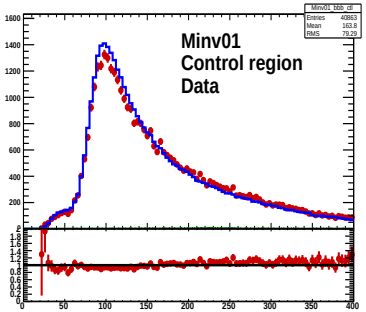
Try other mass points: $M_H = 250$

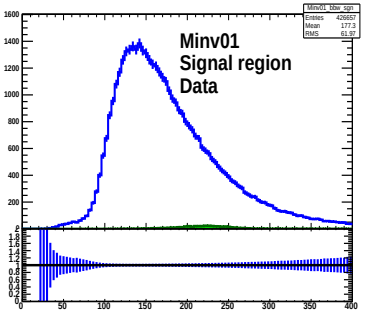






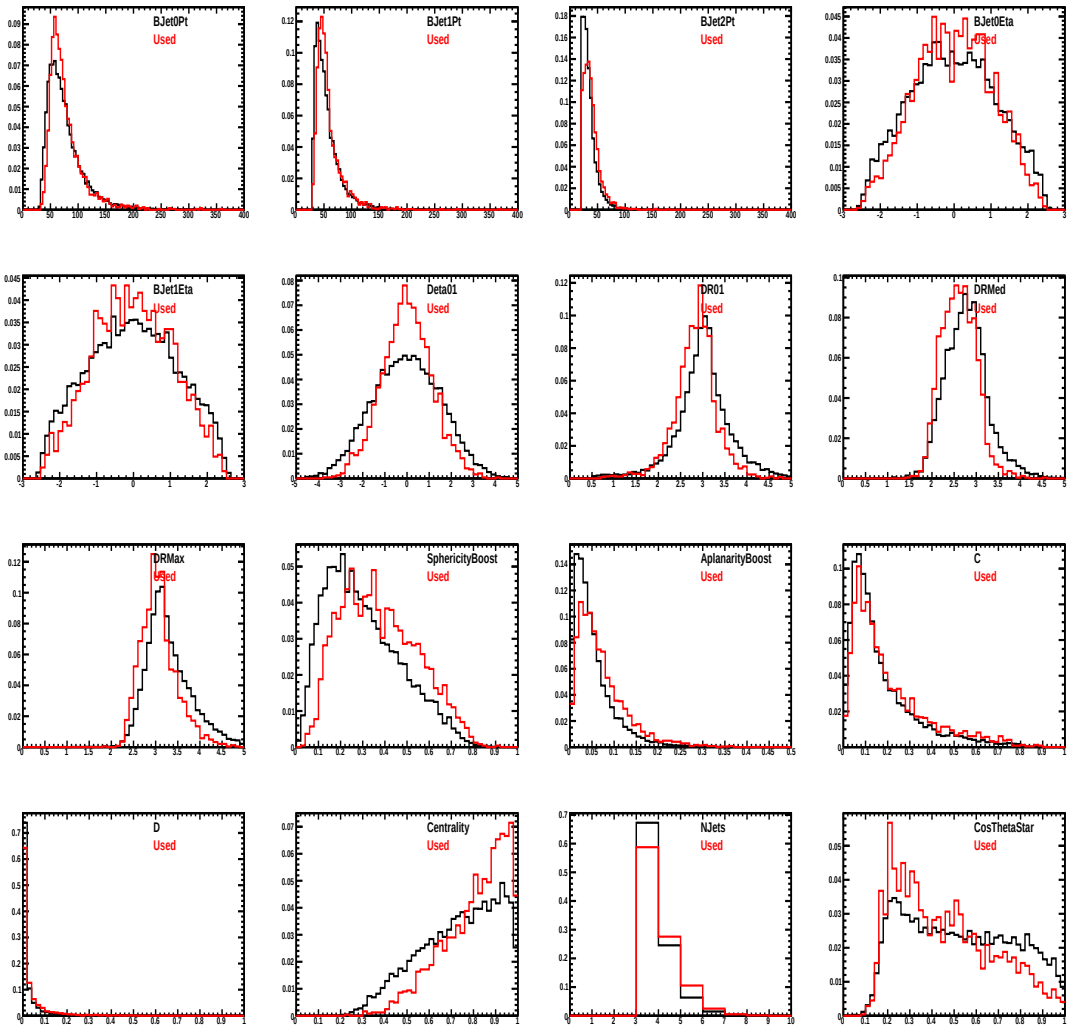


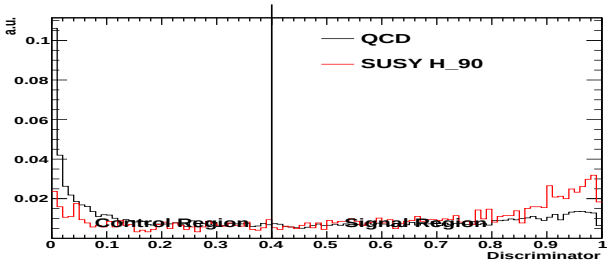


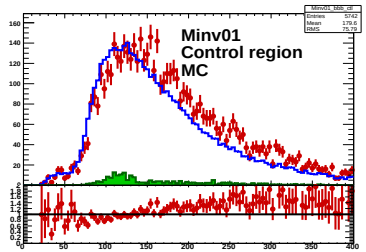


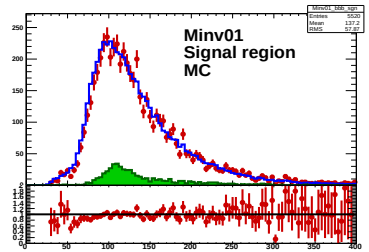
- Easier discrimiation;
- broader and lower signal;

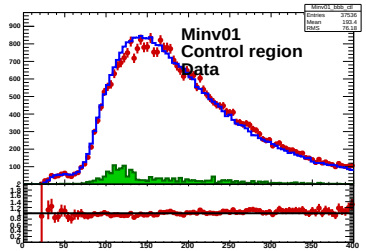
Try other mass points: $M_H = 90$

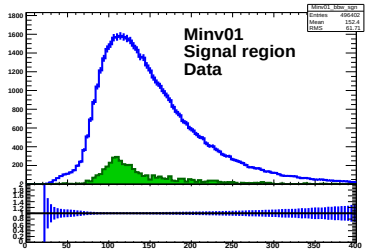












- harder discrimiation;
- leading and sub-leading $b-jets$ not anymore the best choice.

MVA and SUSY MC statistics

MVA

- have tried MVA using variables considered for discriminator
- training samples

background predicted (bbb) in signal region
signal SUSY MC

- problem: too few events in final sample with current SUSY MC sample
  - eg: $M_H = 120$ GeV: Produced 10^6 ev, $P_T^\mu > 15$ GeV $\approx 50 kev$, all selections (bbb) $\approx 2 kev$
  - similar for other mass points.
- Request MC production with filter at generator level on P_T^μ. Same number of events produced, $\times 20$ events passing all cuts

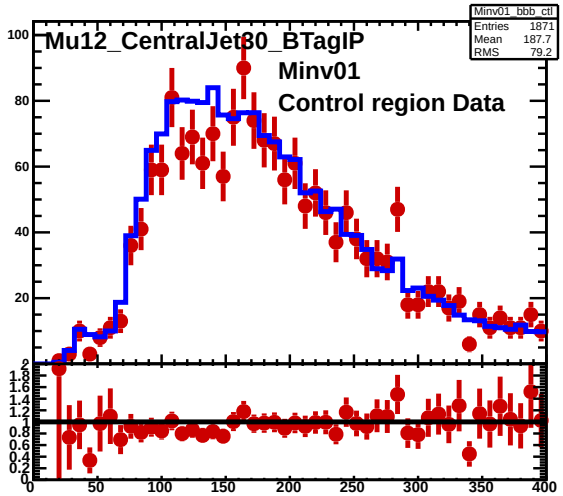
Summary

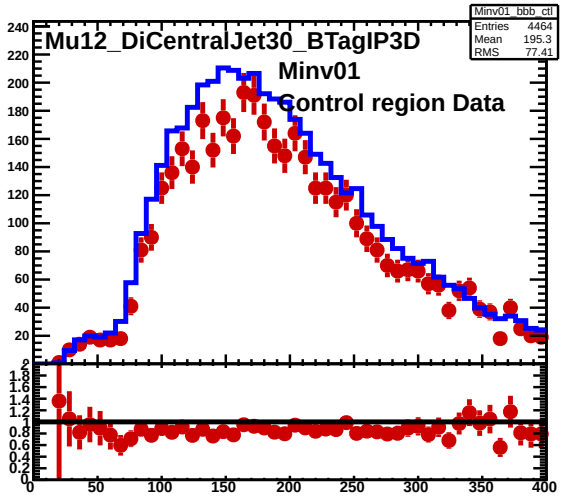
- Progress in MC closure test, closer but not there, yet.
  - need it to get systematics in extrapolation from control to signal region for data
  - other possibility: use $bjj \times P_b^{2nd-jet}(j)$ vs bbj in data
- Discriminators optimization in good shape;
- Results for different mass points
  - lower one requires additional studies;
- Need more SUSY MC for MVA;

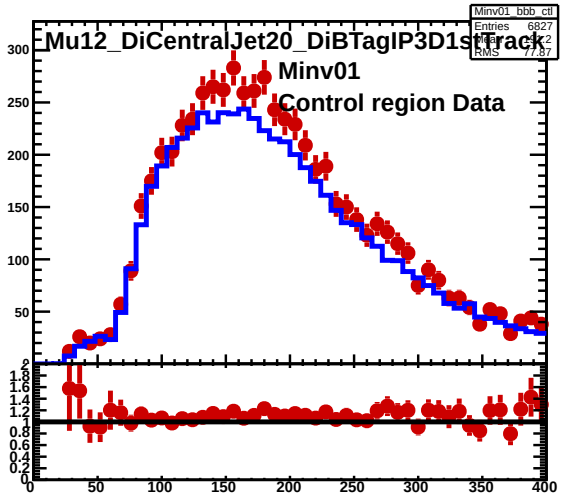
BACKUP

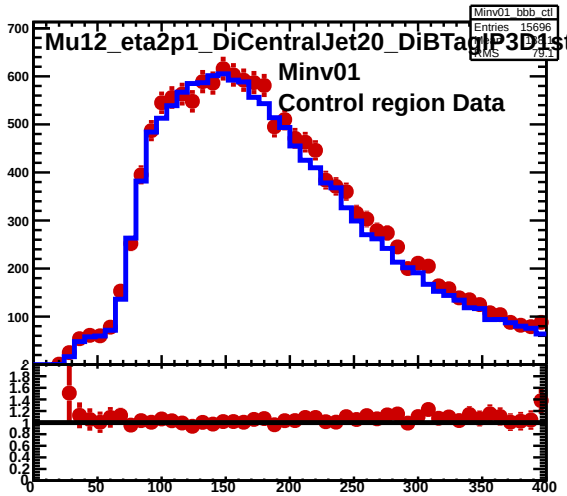
BACKUP

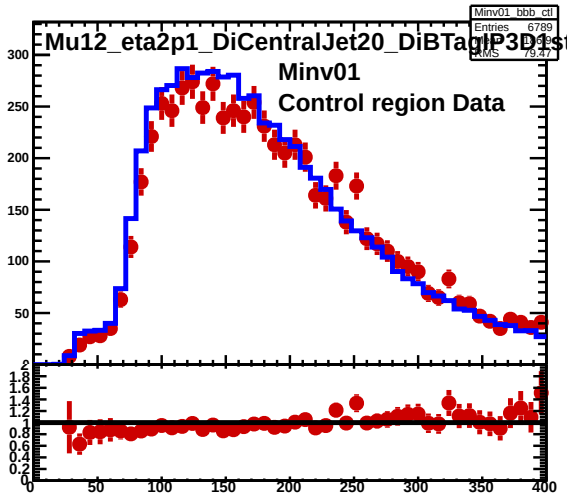
Data: HLT comparison



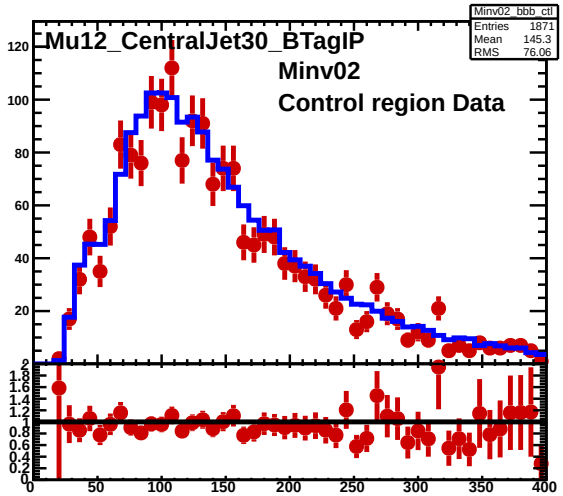


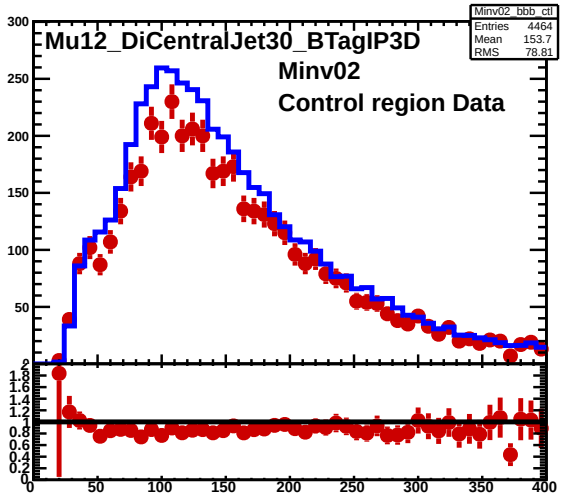


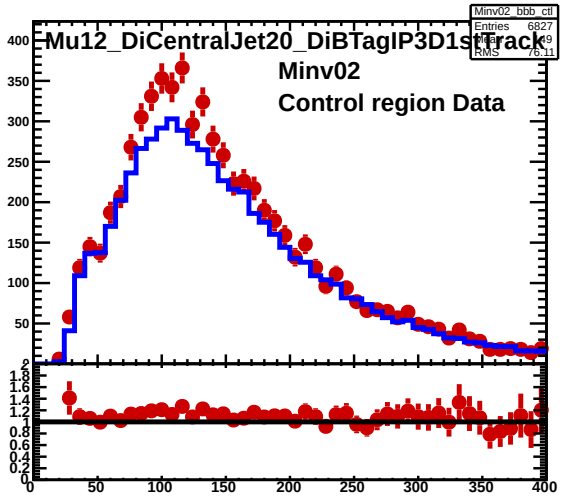


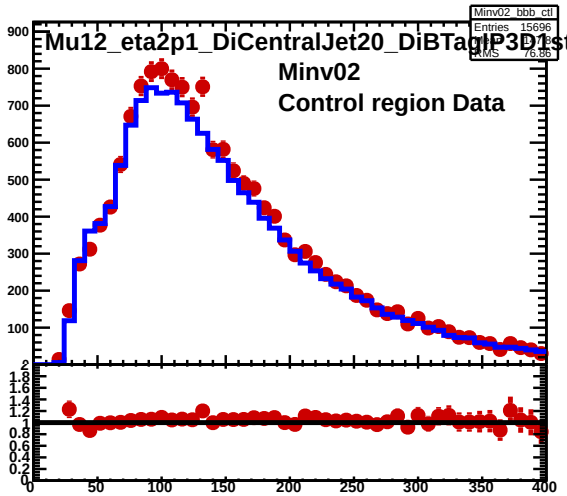


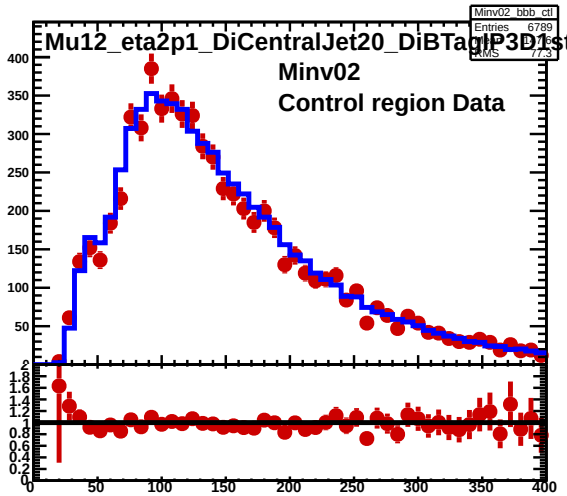
Data: HLT comparison











Data: HLT comparison