Status update on semileptonic MSSM $bb\Phi(bb)$ analysis

Yes, we are alive!

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

INFN and University Padova

Hbb meeting,
CERN, 7 October 2011

Outline

Outline

Trigger summary so far

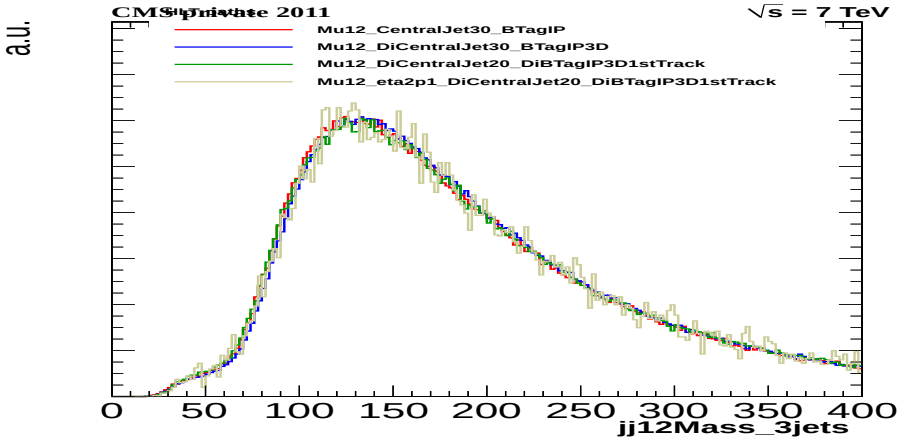
Run Range	start date	HLT Path	Int Lumi /pb	# triggers
160065 - 161176	2011.03.09	HLT_Mu17_CentralJet30_v1	10.86	420,150
161216 - 163261	2011.03.22	HLT_Mu17_CentralJet30_v2	50.33	2,291,565
160065 - 163261	2011.03.09	HLT_Mu17_CentralJet30	61.19	2,711,715
163269 - 163869	2011.04.21	HLT_Mu17_CentralJet30_BTagIP_v2	178.94	1,501,522
163269 - 163869	2011.04.21	HLT_Mu17_CentralJet30_BTagIP	178.94	1,501,522
163738 - 163869	2011.04.30	HLT_Mu12_CentralJet30_BTagIP_v2	49.36	941,949
165088 - 165633	2011.05.15	HLT_Mu12_CentralJet30_BTagIP_v4	146.05	2,864,344
163738 - 165633	2011.04.30	HLT_Mu12_CentralJet30_BTagIP	195.41	3,806,293
165970 - 167043	2011.05.27	HLT_Mu12_DiCentralJet30_BTagIP3D_v1	551.94	4,943,604
167078 - 167913	2011.06.16	HLT_Mu12_DiCentralJet30_BTagIP3D_v3	240.01	2,157,993
165970 - 170901	2011.05.27	HLT_Mu12_DiCentralJet30_BTagIP3D	791.95	7,101,597
167039 - 167913	2011.06.16	HLT_Mu12_DiCentralJet20_DiBTagIP3D1stTrack_v1	262.12	634,201
170249 - 173198	2011.07.16	HLT_Mu12_DiCentralJet20_DiBTagIP3D1stTrack_v3	935.8	2,260,949
167039 - 173198	2011.06.16	HLT_Mu12_DiCentralJet20_DiBTagIP3D1stTrack	1197.92	2,895,150
173236 - 177878	2011.08.12	HLT_Mu12_eta2p1_DiCentralJet20_DiBTagIP3D1stTrack_v1	1837.33	4,353,016
173236 - 177878	2011.08.12	HLT_Mu12_eta2p1_DiCentralJet20_DiBTagIP3D1stTrack_v1	1837.33	4,353,016
		HLT_Mu17_Anything	240.13	4,213,237
		HLT_Mu12_Anything	4022.61	18,622,056
		All path	4262.74	22,368,293

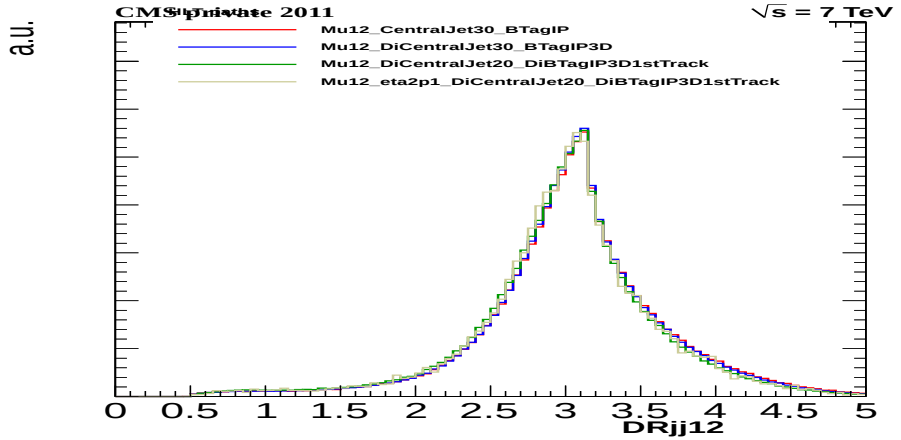
3 HLT paths covers $\sim 3.8\ fb^{-1}$ (from $\sim 4.2\ fb^{-1}$ total: $\sim 90\%$)

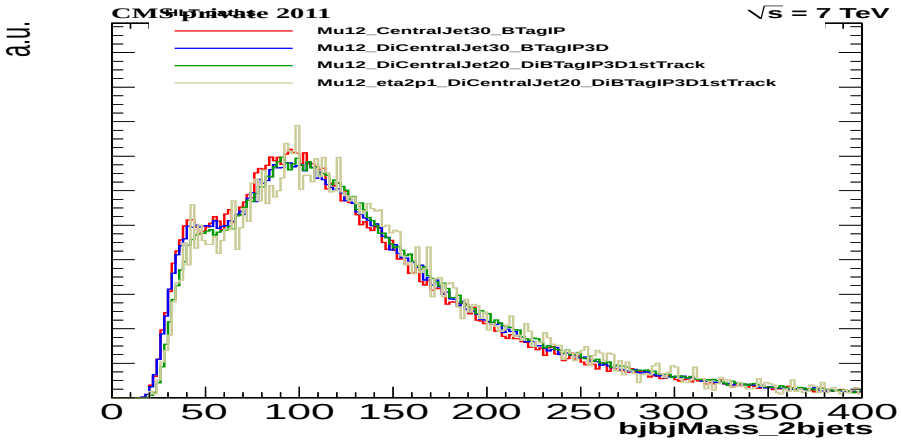
Trying to keep trigger table up-to-date on HiggsBBbb twiki page

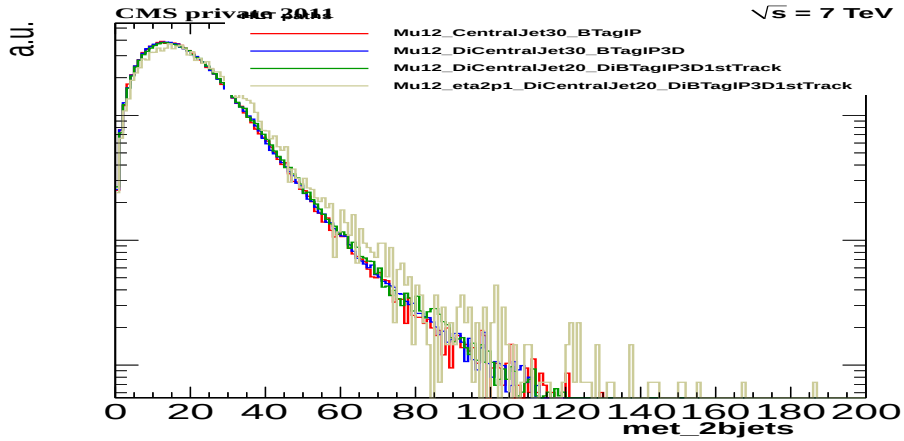
Comparing HLT paths

Pre-selection: HLT, 3 jets $E_t > 25$ $|\eta| < 2.6$, 1 *mu* $p_t > 15$









Most probably we can use all HLT paths at once.

NB. Yellow line is for **Mu12_eta2p1_DiCentralJet20_DiBTagIP3D1stTrack**, small sample analyzed

Outline

Get background from "data"

- $bb\Phi bb$ analysis is based on kinematic distributions of triple b-jets events (such as – but not limited to – M_{bb}), to be compared with expected irreducible background;
- dominant background is from QCD (multijet)
  - `pythia` multijet description is poor: better for upcoming sample with `alpgen`
  - contribution from double gluon splitting and from single gluon splitting plus mistagged b
- how to get a data-driven estimate of background?
- General idea: start from a sample with less than three b-tag and extrapolate to three b-tags.

General idea

In formulas

Let $F(bbb)$ be a generic distribution for three b-tag jets, and $F(bbj)$ the same distribution for two b-tag jets plus a third w/o b-tag requirements.

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

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

- ϵ_x b-tag efficiency for x=b,c,light jets;
- $f_x^{(3)}$ fraction of x=b,c,light jets in third jet of the event.

Not interested in separating c from *light*, so will work with average $\epsilon_q = \langle \epsilon_c, \epsilon_l \rangle$

$$P_b^{3rd-j}(j) = \epsilon_b \cdot f_b^{(3)} + \epsilon_q \cdot (1 - f_b^{(3)})$$

Let's go in small step.

Start from b+2jet data-sample, try to estimate *bbj* and compare with real yield/distribution;

Test in *bjj bbj*

As before

To check if the system works, try it on *bbj*/*bjj* sample (pure MC)

$$F(bbj) = F(bjj) \times P_b^{2nd-j}(j)$$

$$P_b^{2nd-j}(j) = \epsilon_b \cdot f_b^{(2)} + \epsilon_q \cdot (1 - f_b^{(2)})$$

μ (from trigger) in the b-tagged jet (first one in this case)

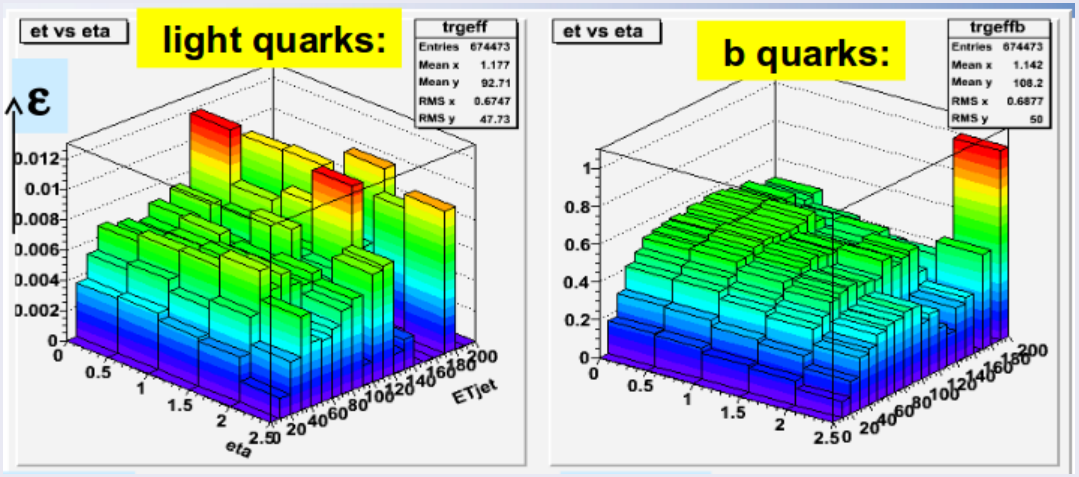
goals:

- compare $F(bbj)$ with $F(bjj) \times P_b^{2nd-j}$
- check the validity of the method ;
- get systematic uncertainties;

Step one: get bb from bj. ϵ_b, ϵ_q

ϵ_b, ϵ_q

- get these probability from MC



- check mistag probability on W hadronic decay selected in semileptonic $t\bar{t}$ sample.
- Estimated sample size $\mathcal{O}(1000)$ events, topology (multijet, multibtag) similar to that of bbHbb
- check b-tag probability of true B by looking at the second b from top, same sample.

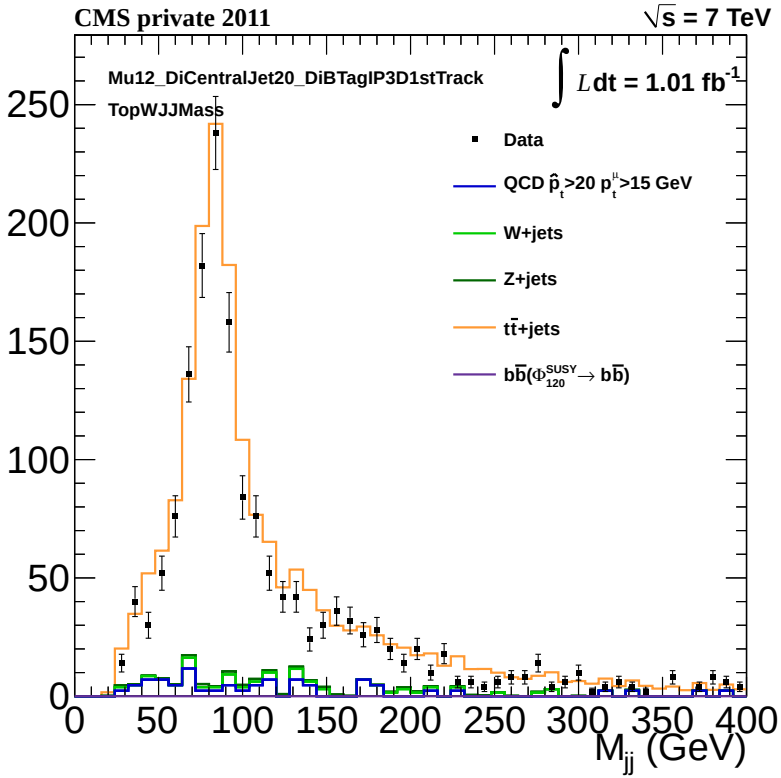
Semileptonic $t\bar{t}$ analysis

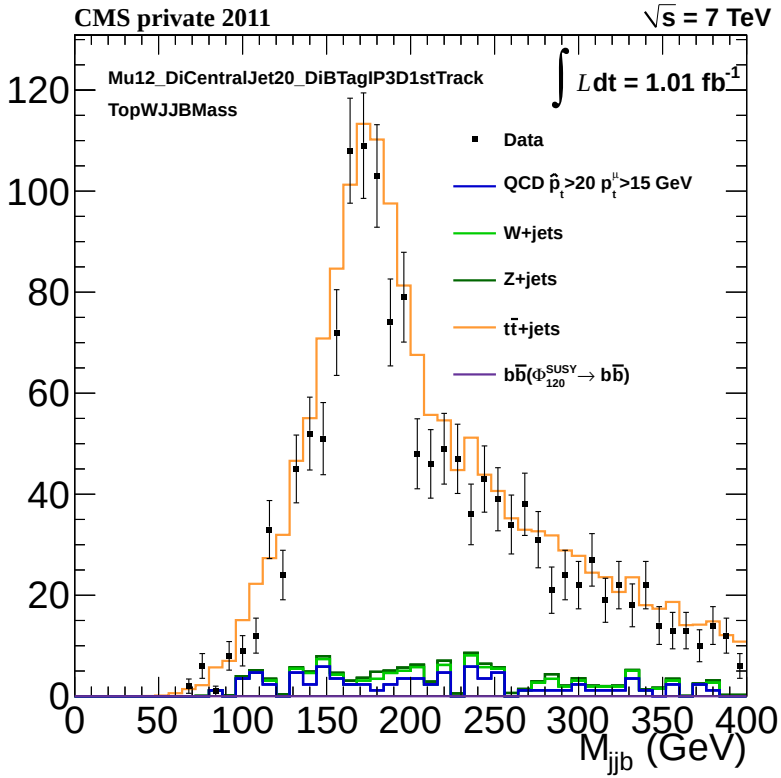
Simple $t\bar{t} \to Wb + Wb \to \mu\nu b + jjb$ selection:

high $p_t(> 25)$ isolated muon, MET> 20, 4 jets of which 2 b-jets.

Starting from `MuHad Trigger` (electron channel also possible to double statistics)

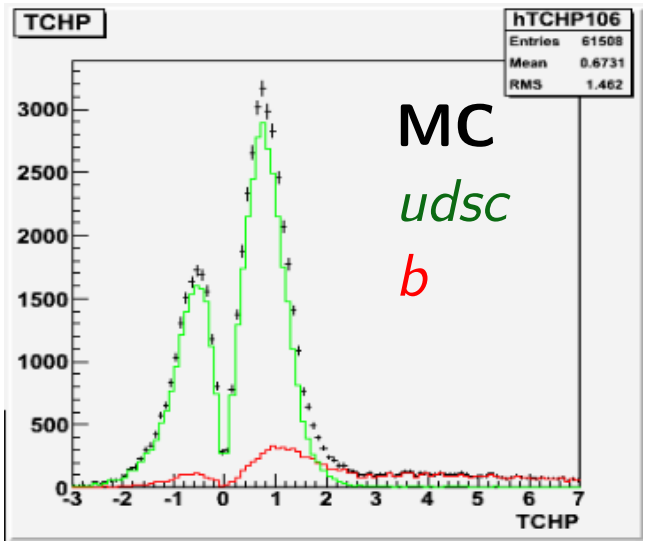
Got a very clean $W \to jj$ peak: $\mathcal{O}(1000\ events/fb^{-1})$

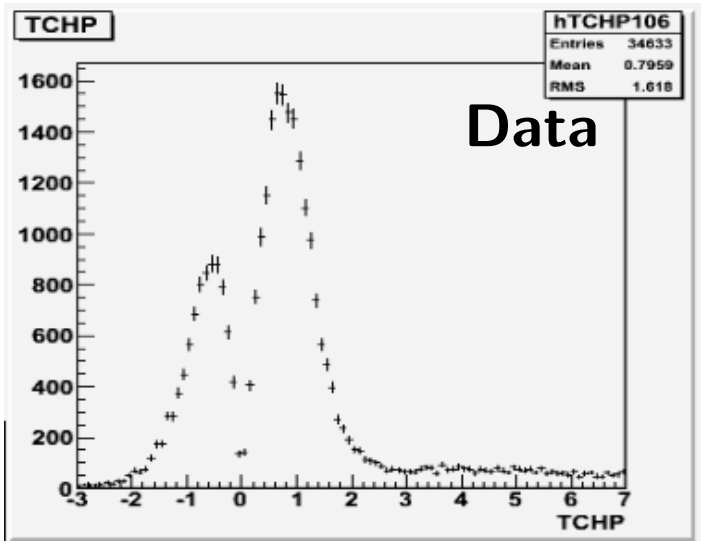




How to get f_b

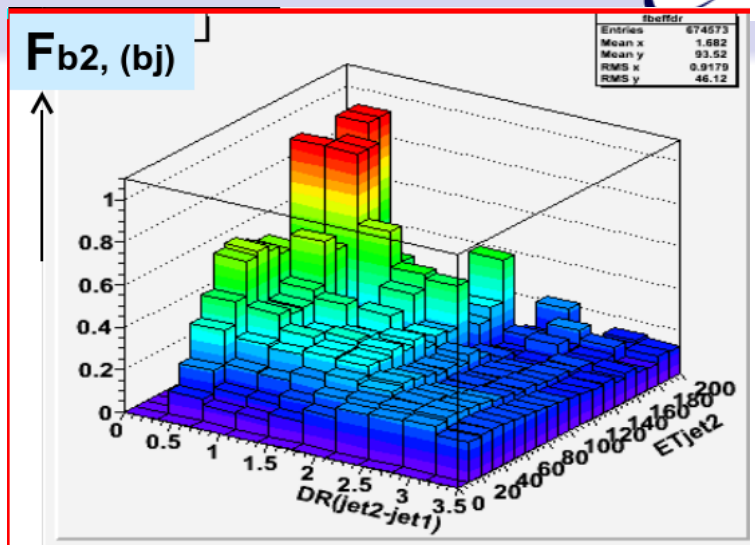
- surely cannot rely on MC for this
- Get it from data by looking at b related variable for all jets in a given bin of E_t $|\eta|$ (and possibly other variables); (see example below)
- could be directly a b-tag variable (such as TCHP discriminator)
- use these distribution in MC to get a template for b-trueB and b-fakeB and fit the data with them, leaving as a free parameters f_x

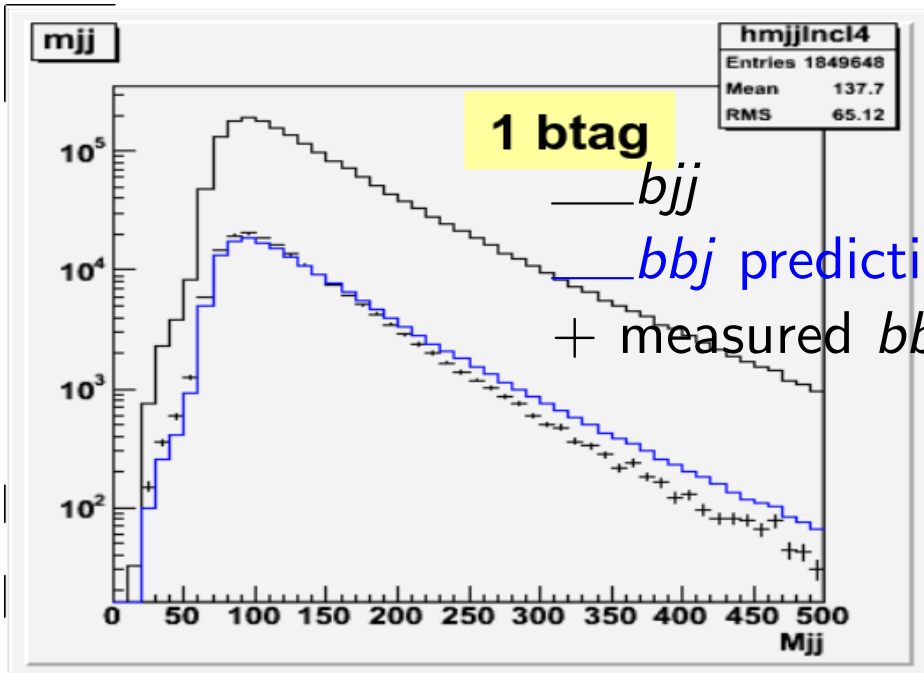


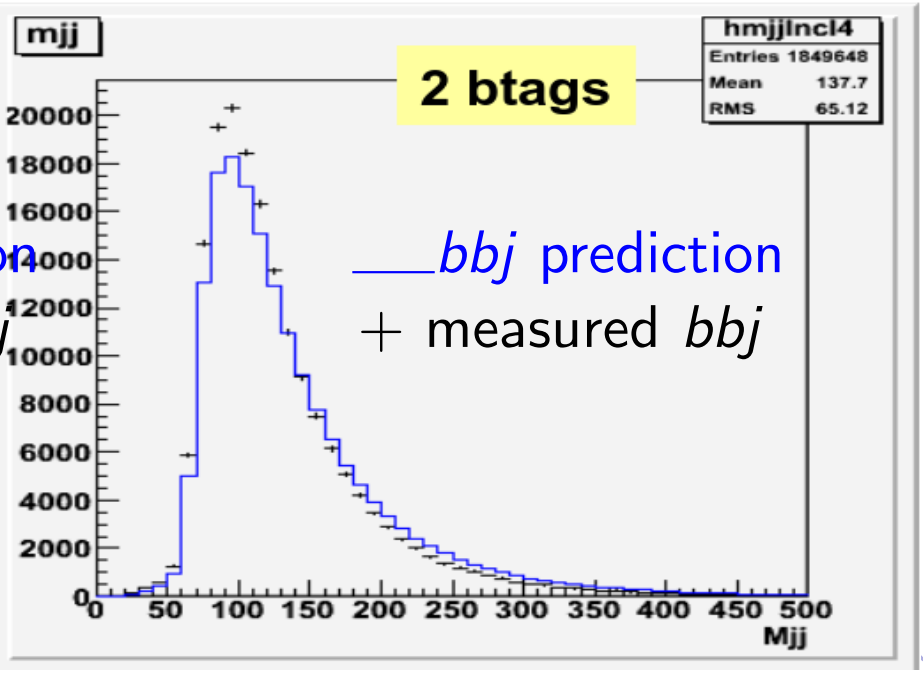


How it works for *bjj* *bbj* (MC)

- Pure MC vs MC study;
- Playing a trick, $f_b^{(2)}(\Delta R, E_t^{(2)})$ is MC truth $\Rightarrow$
- results are encouraging: possibly a better parametrization is needed (N_{trk}?)

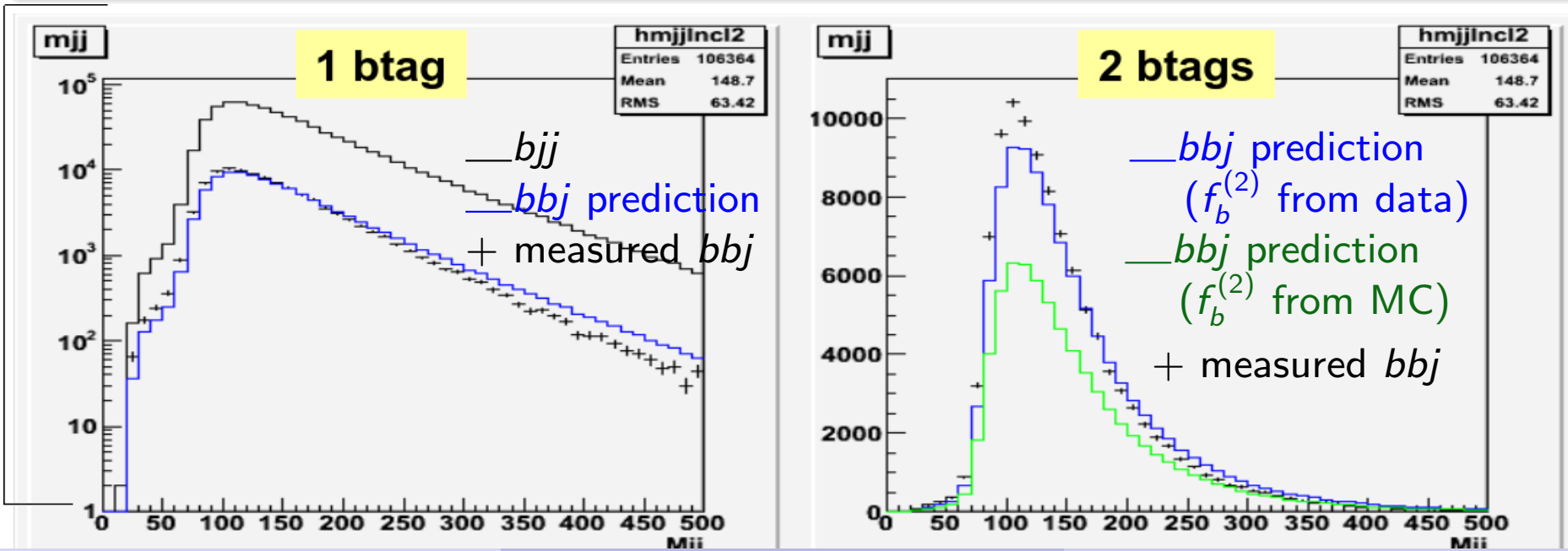






What about real data?

- HLT path: `HLT_Mu12_DiCentralJet30_BTagIP3D`;
- $\int \mathcal{L} dt = 552\ pb^{-1}$;
- with $f_b^{(2)}$ from MC normalization is poor (Green line);
- use Data TCHP fit to get normalization of $f_b^{(2)}$, dependency on $(\Delta R, E_t^{(2)})$ still from MC (Blue line): better agreement



Ok, what about three b prediction? (still pure MC)

Probably is not supposed to work, but ...

Start from bjj events:

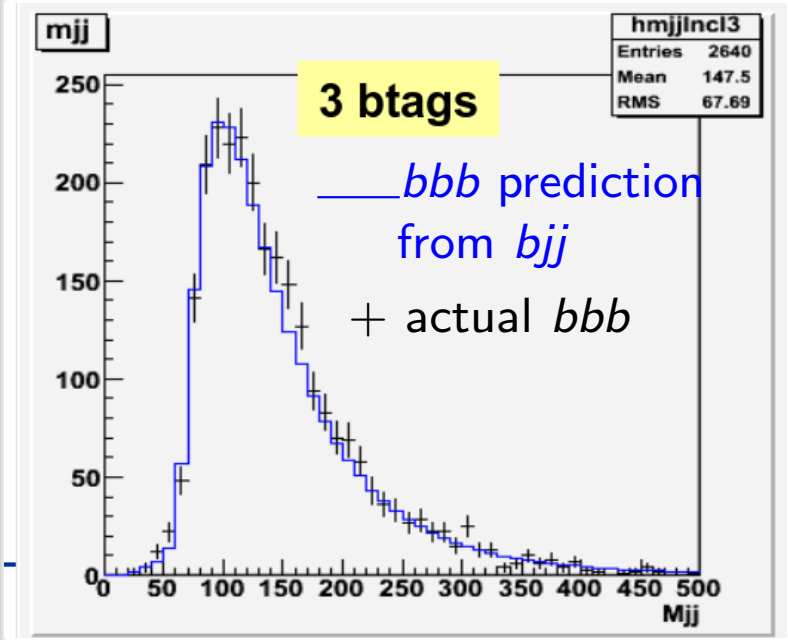
$$F(bbb) = F(bjj) \times P_b^{2nd-j}(j) \times P_b^{3rd-j}(j)$$

$$P_b^{2rd-j}(j) = \epsilon_b \cdot f_b^{(2)} + \epsilon_q \cdot (1 - f_l^{(2)})$$

$$P_b^{3rd-j}(j) = \epsilon_b \cdot f_b^{(3)} + \epsilon_q \cdot (1 - f_b^{(3)})$$

$f_b^{(2,3)}$ is fraction of b in second and third jet, obtained as before comparing bjj with bbj and bjb, respectively.

problem: are we taking into account the double gluon splitting mechanism in bjj? Namely, how reliable is $f_b^{(2,3)}$ (especially $f_b^{(3)}$)?



NB: Normalization is forced

Shape is remakably well predicted

MC:Pythia QCD μ-enriched $\hat{p}_t > 20, p_t^{\mu} > 15$

~ 30 *Mevents* (expect $3\times$ in Fall11)

Can we play the same trick with *bbj* vs *bbb*?

Short answer: NO

- Just using $F(bbb) = F(bbj) \times P_b^{3rd-j}(j)$ is inherently biased
- getting P_b (f_b) from any distribution with 3 b-tagged jets would include also possible signal;
- so f_b would include also a possible contribution from signal, thus the prediction would hide a possible signal as background.
- In short, measuring f_b is measuring background+signal.

Yes, but . . .

If we have a control region in three *b* jets sample, where we know that only background is present (or where signal is significantly lower) we can use it to get P_b (f_b).

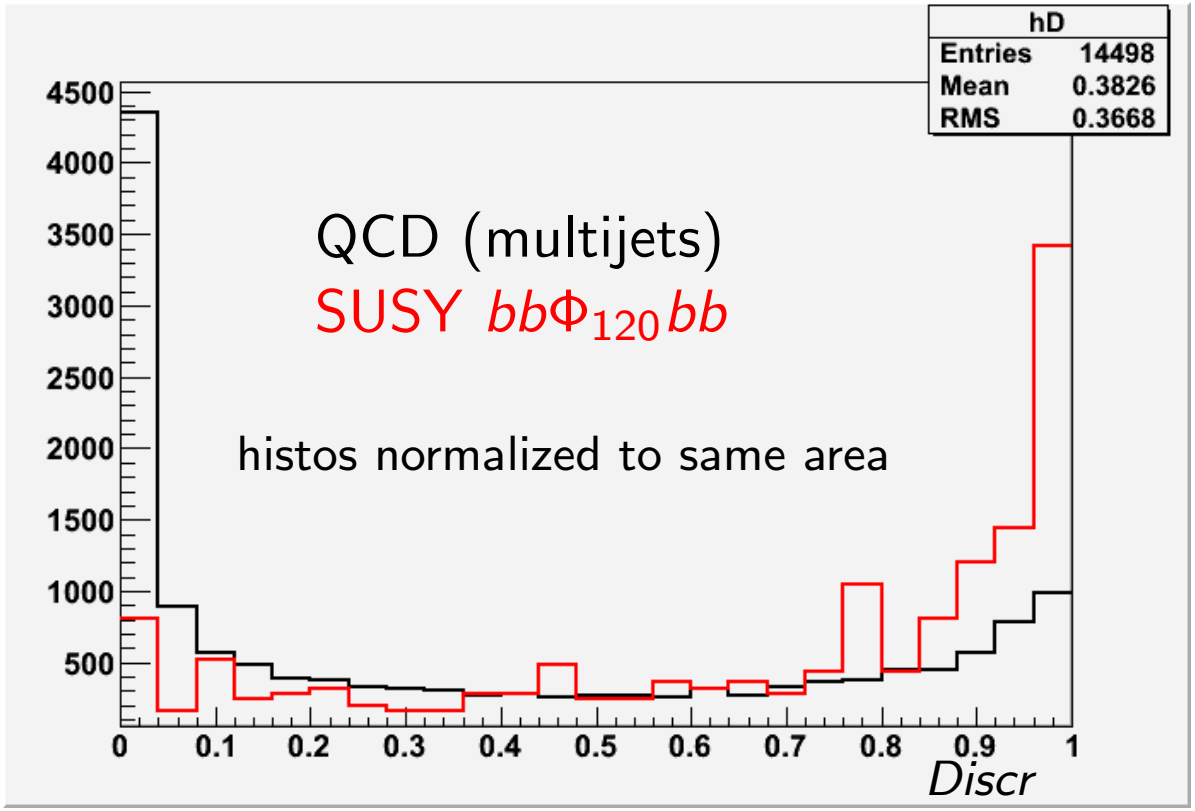
Do we have such a control region? Yes, we think so.

Use MC QCD (multijet) and Susy signal to build a probability function.

Control region (from MC)

Pre-selection 3 bjets ($E_t > 30, 30, 25\ GeV$, $CSV > 0.8, 0.8, 0.6$)

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

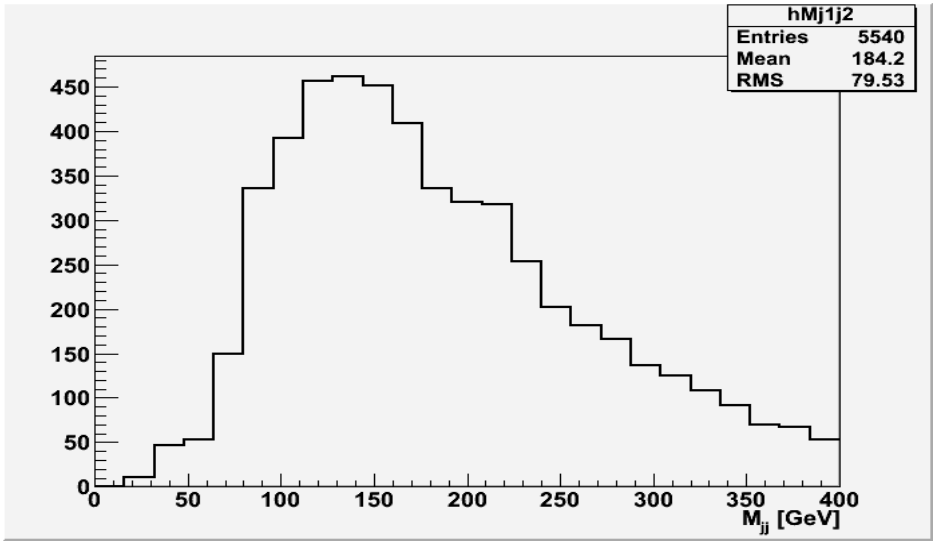


variables used (plots in backup): still studying

- $\Delta\eta(j^{(1)}, j^{(2)})$
- $\Delta\phi(j^{(1)}, j^{(2)})$
- $\Delta R(j^{(1)}, j^{(2)})$
- $\Delta R(j^{(1)}, j^{(3)})$
- $\Delta R(j^{(2)}, j^{(3)})$
- $\Delta R_{min,med,max}(jj)$
- $p_t^{j_{(1)}}$
- $\eta^{j^{(1)}+j^{(2)}}$
- Sphericity (good discrimination)
- Acoplanarity
- $\theta(j_{(1)}, j^{(1)} + j^{(2)})$

Control region (II)

MC QCD (multijets), same preselection as previous slide and $Discr < 0.1$



- Events in control region fill nicely the whole phase space;
- can be used to get $P_b^{3rd-j}(j)$ from data (bbb vs bbj);
- or to normalize/correct $P_b^{2nd-j}(j) \times P_b^{3rd-j}(j)$ (bjj vs bbj, bjb);
- will try both

Outline

Conclusion

- We have data;
- We might have all the ingredients to get background expectation
- Some even shown to work;
- Still a lot of work ahead;

Backup

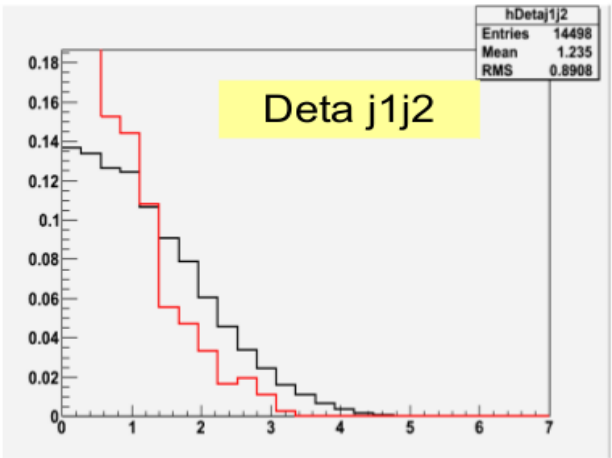
BACKUP slides

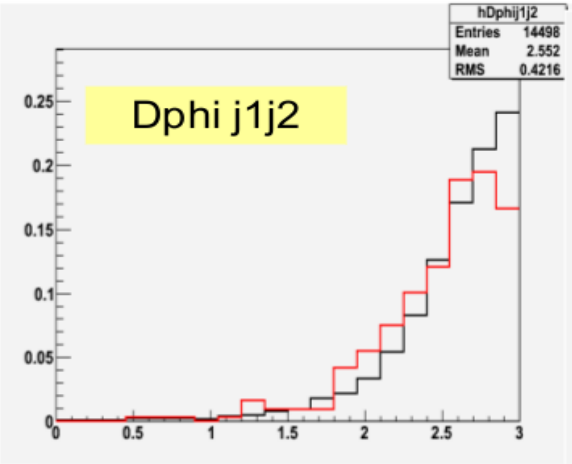
Event Selection:

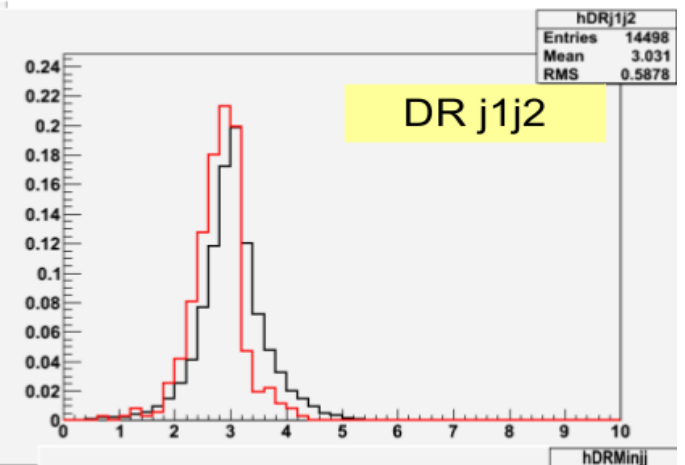
1) 2 leading jet pt > 30 GeV & CSV > 0.8. At least one matched to the trigger muon
2) 3rd jet pt > 25 & CSV > 0.6

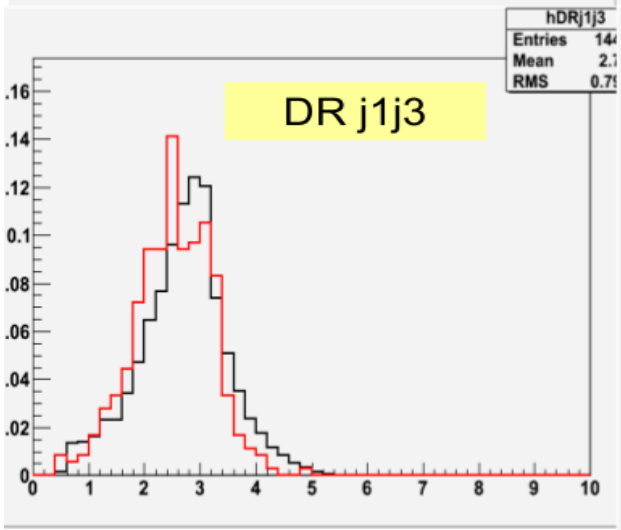
QCD

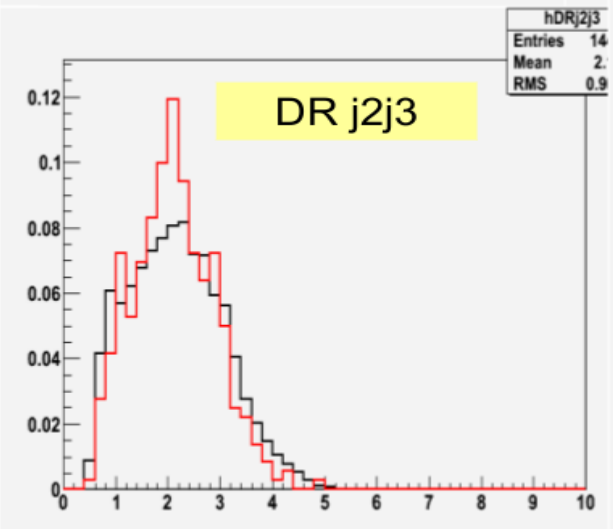
BBHBB M120GeV

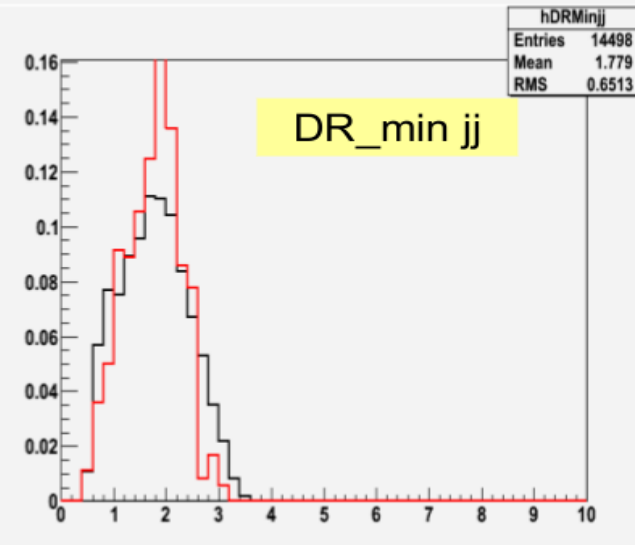


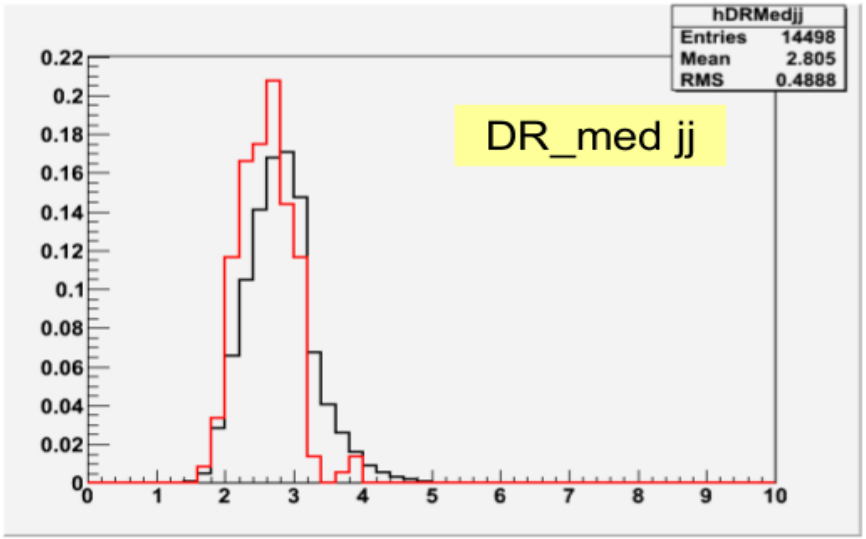


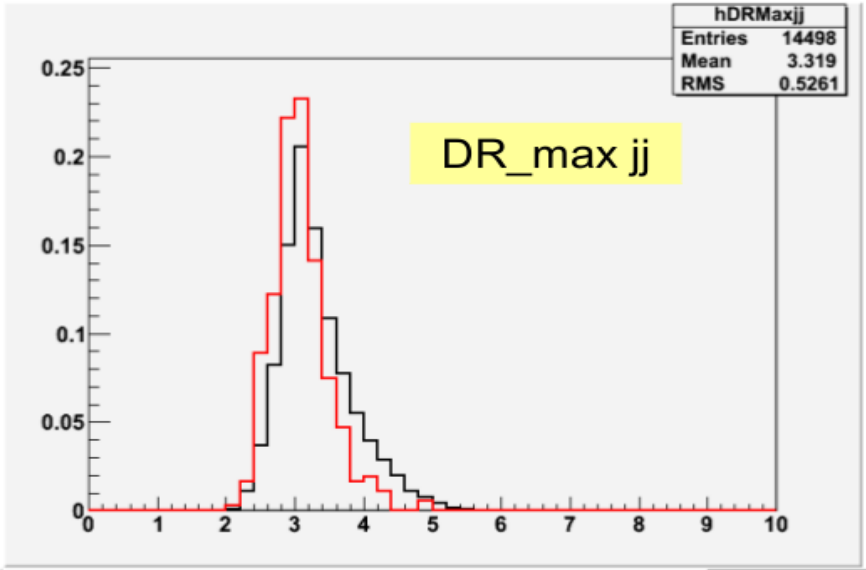


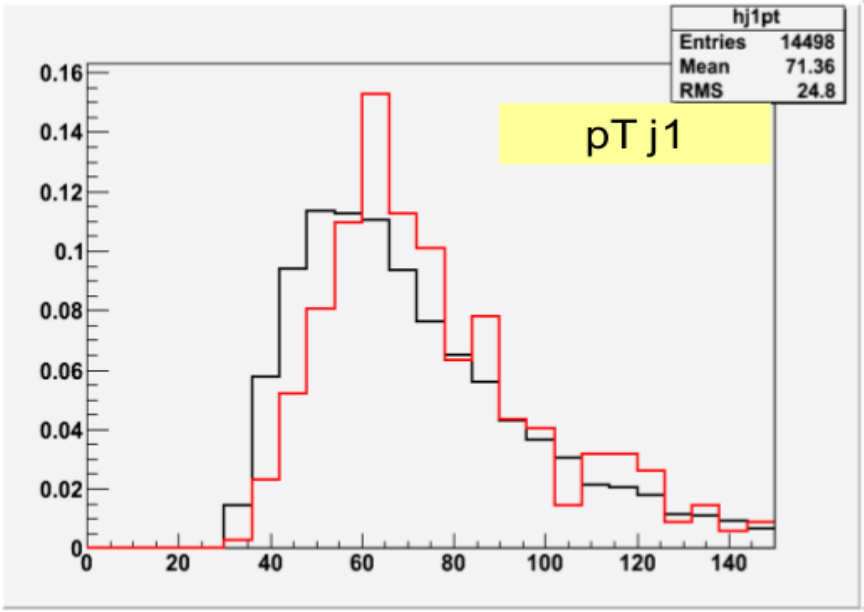


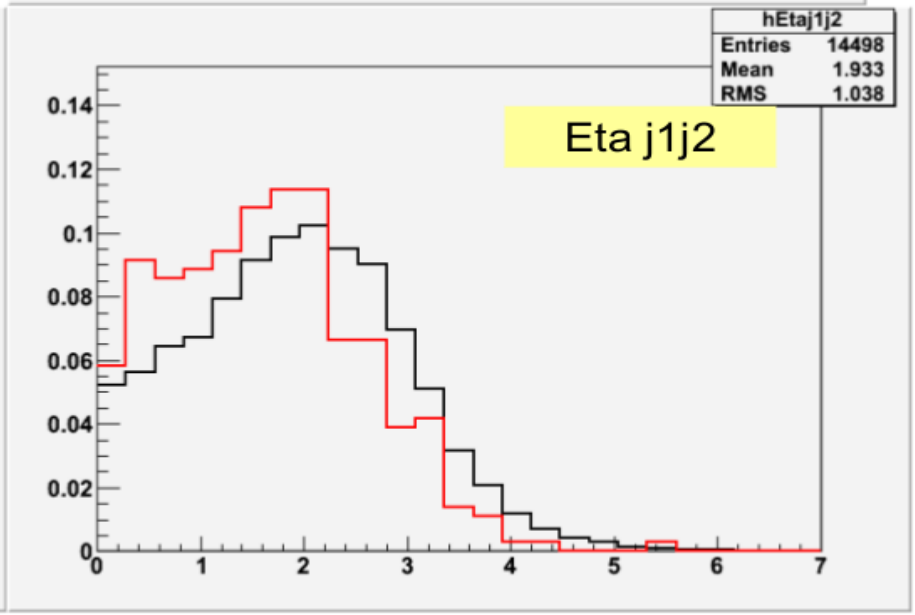


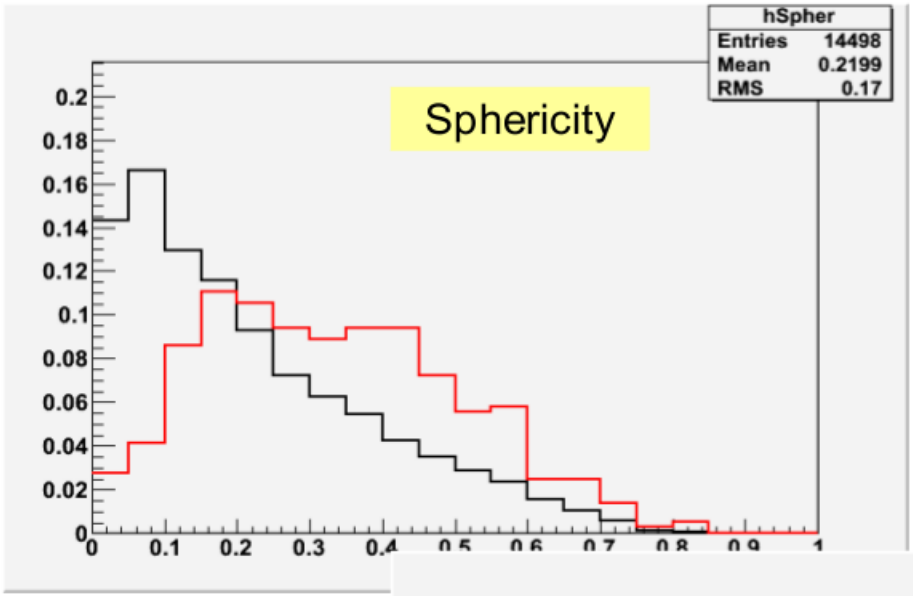


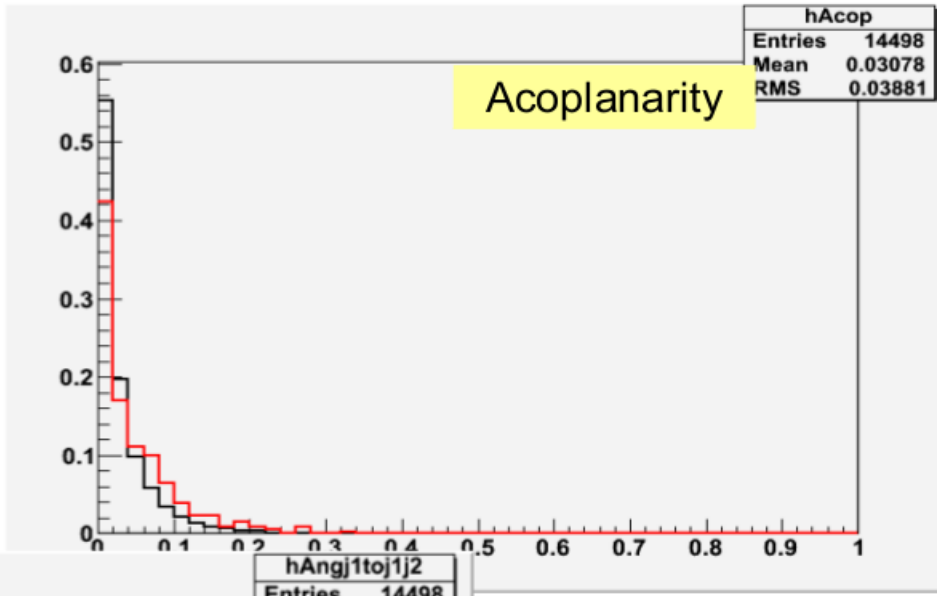

hDRMedjj
Entries 14498
Mean 2.805
RMS 0.4888
DR_med jj

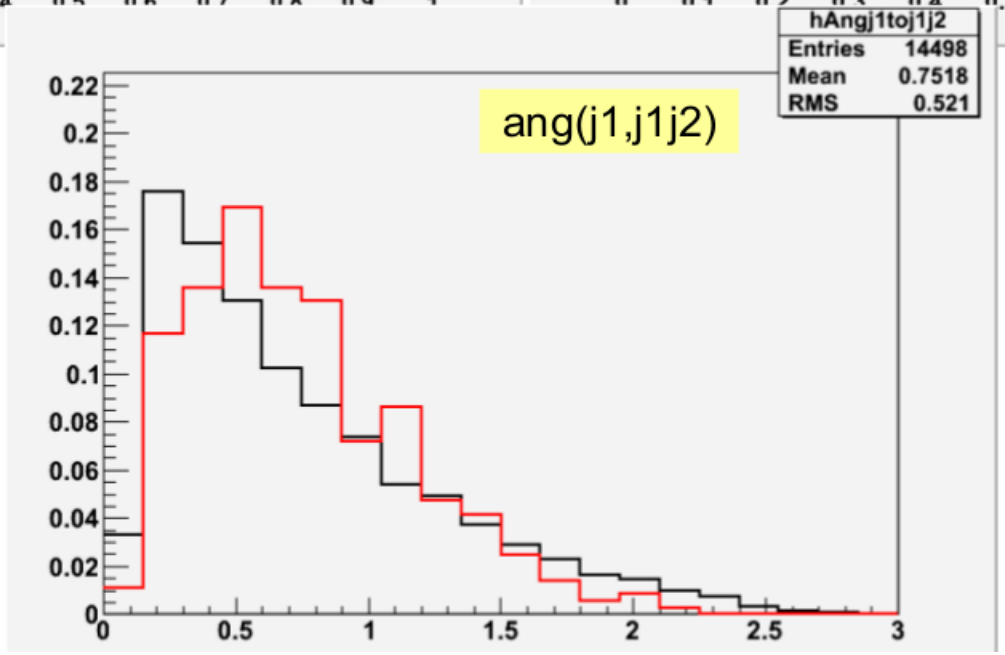

hDRMaxjj
Entries 14498
Mean 3.319
RMS 0.5261
DR_max jj


hj1pt
Entries 14498
Mean 71.36
RMS 24.8
pT j1


hEtaj1j2
Entries 14498
Mean 1.933
RMS 1.038
Eta j1j2

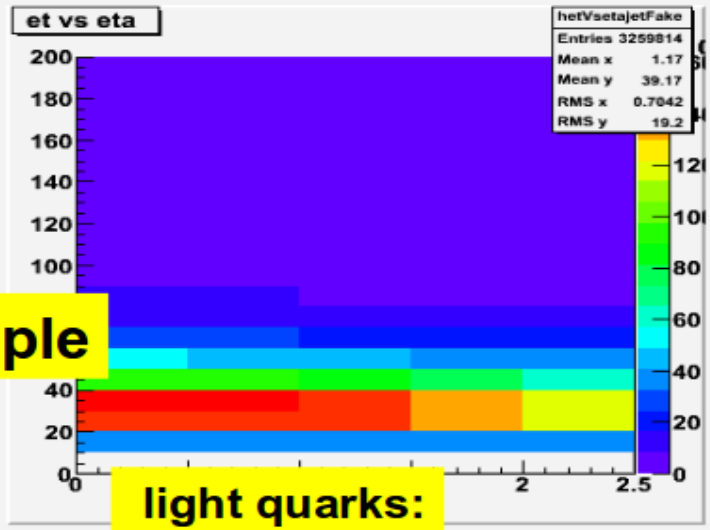

hSpher
Entries 14498
Mean 0.2199
RMS 0.17
Sphericity


hAcop
Entries 14498
Mean 0.03078
RMS 0.03881
Acoplanarity

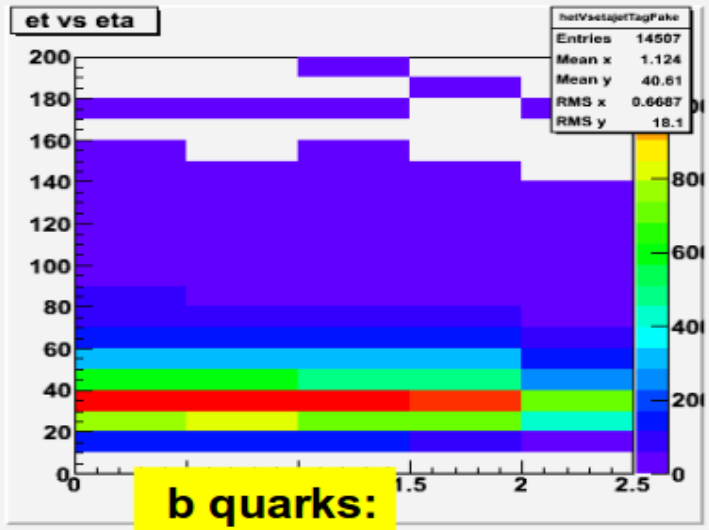

hAngj1toj1j2
Entries 14498
Mean 0.7518
RMS 0.521
ang(j1,j1j2)

Tag matrices (MC)

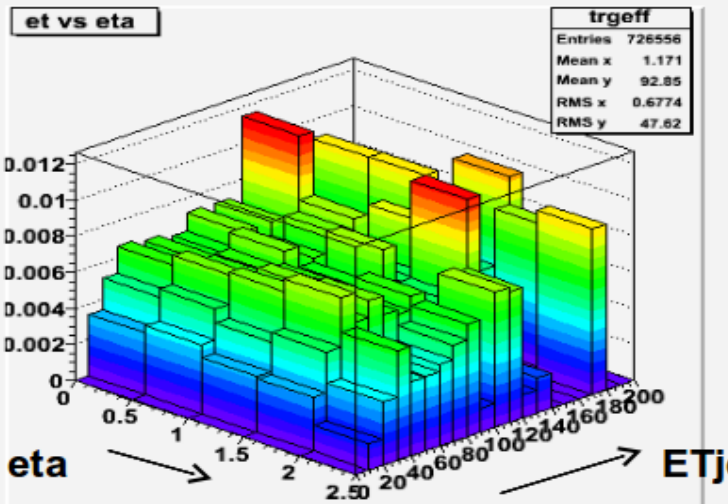
bj sample

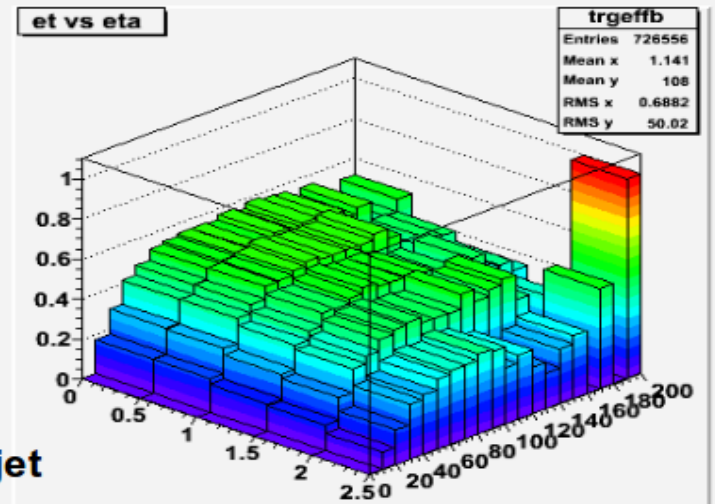


light quarks:

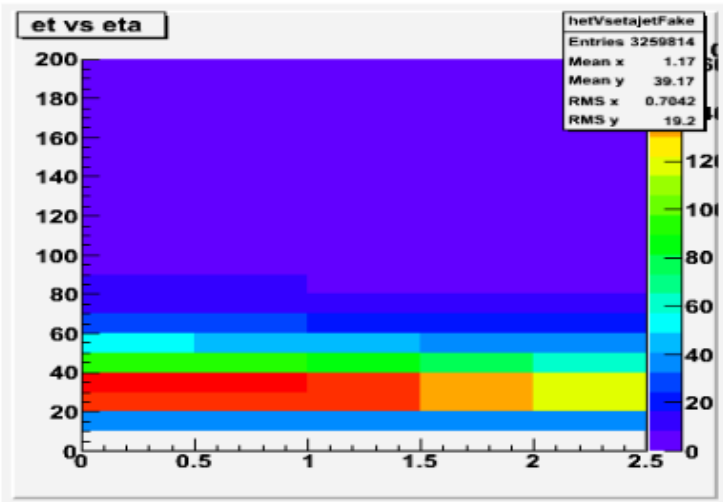


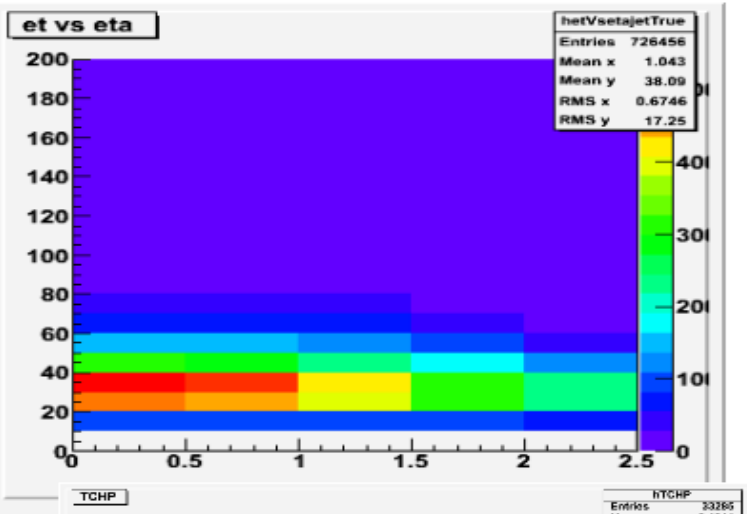
b quarks:

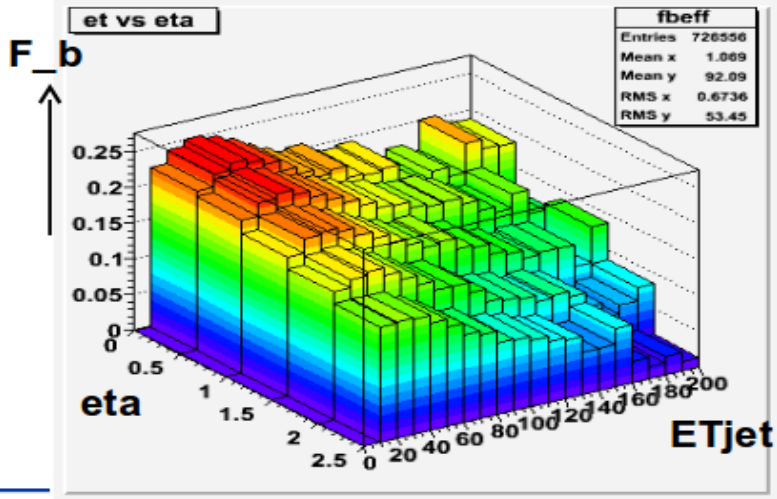


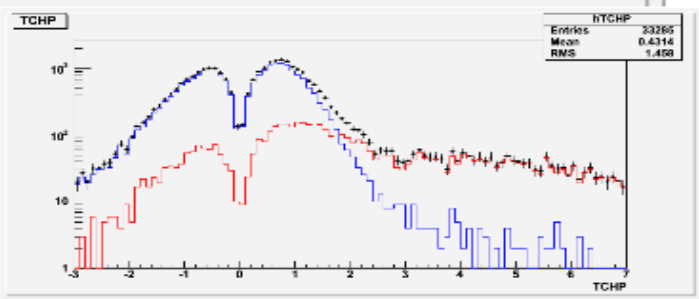


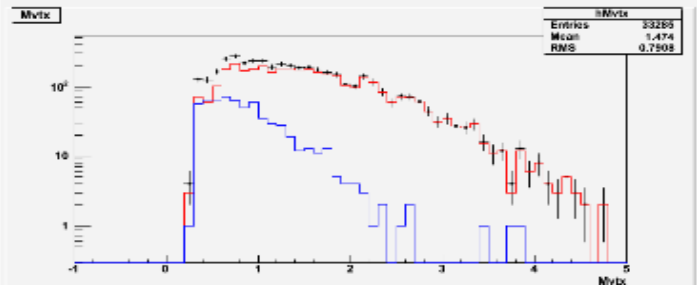
f_b(ET,η) matrix







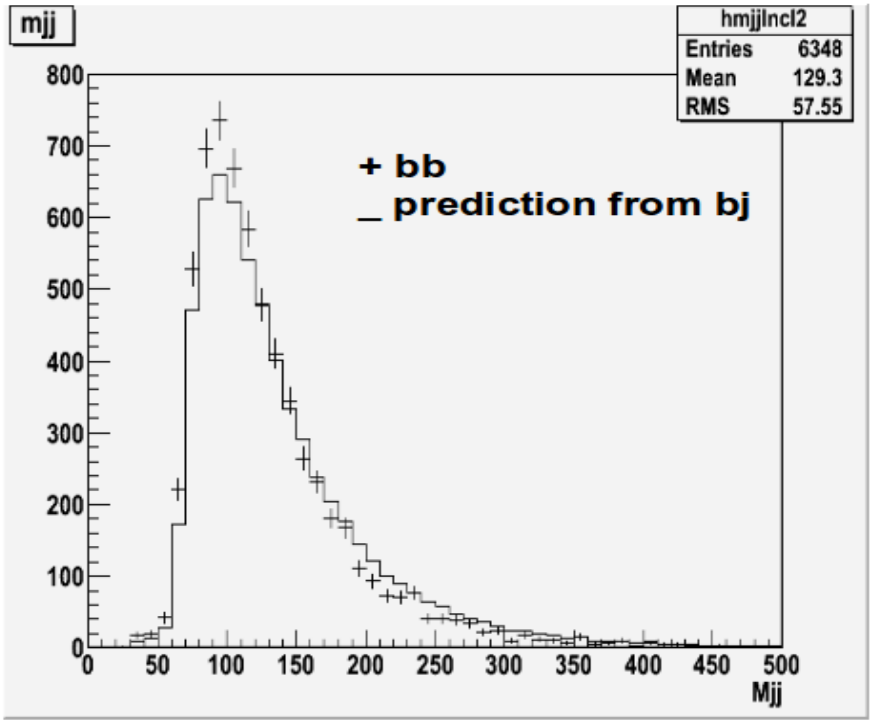




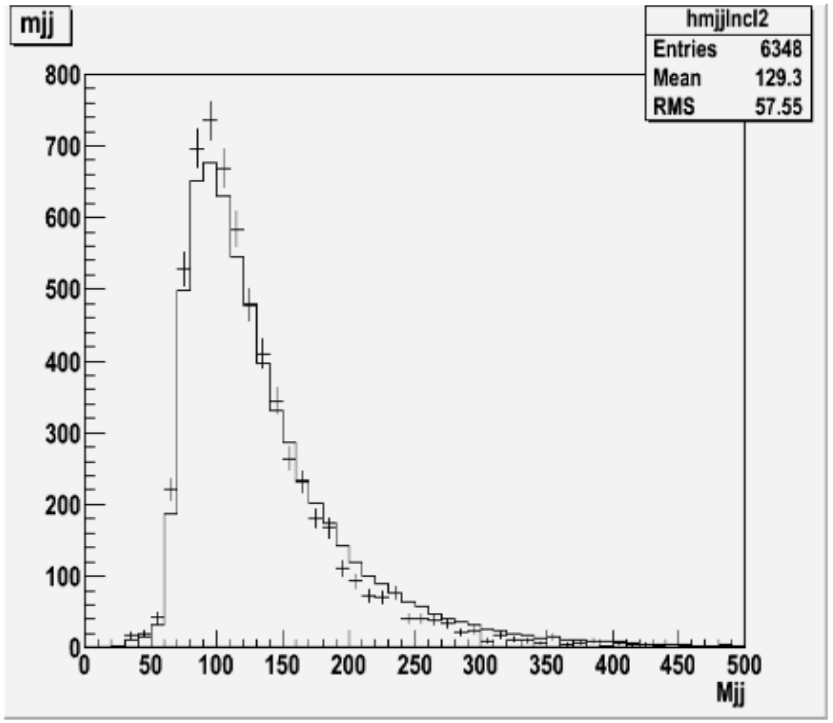
10/4/11

bj sample, Mjj in MC: "data" vs prediction

(Et,eta) parametrization



(Et,DR) parametrization



(MC vs MC: should work, not so easy (possible poor ϵ and f_b parametrization)