MC VBF+MET QCD Samples Studies

João Pela

Imperial College London

2014-02-09

Today's presentation

Topics

- QCD VBF samples details
- Kinematics distributions
- Yields Comparison QCD Inclusive and VBF-like
- GenMET vs RecoMET Study

Introduction

QCD are by far the most frequent processes in collisions at CMS. The elevated cross sections of such processes mean it is normally impossible to generate samples big enough to simulate significant amounts of of equivalent luminosity so they can be used in data analysis.

Methodology

In order to overcome this problem we generated MC QCD samples with MET plus VBF-like jets.

- Real MET (vectorial sum of generator level neutrino p_T)
- VBF-like jets (AK5 generator level jets)

This type of event have a significantly smaller cross section and so to simulate high integrated luminosity samples.

MC Filter Details

MC Filter: Vectorial sum of neutrino E_T

- $\sum E_{\perp}(\vec{\nu}) > 40\ GeV$

MC Filter: Dijet Filter

- Select jets with:
  - $p_{\perp} > 20\ GeV$
  - $|\eta| < 5.0$
- From selected jets at least one pair with:
  - $m_{jj} > 700\ GeV$
  - $\Delta\eta > 3.2$

Sample Details

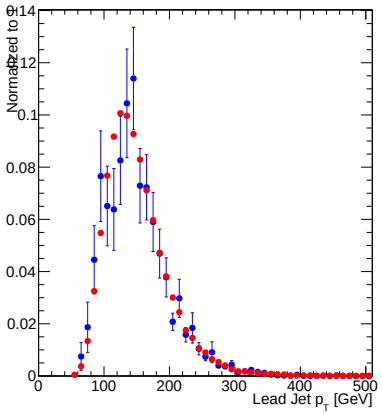
Database URL:
https://cmsdbsprod.cern.ch:8443/cms_dbs_ph_analysis_01_writer/servlet/DBSServlet

Sample	Identifier
QCD-Pt-80to120	/VBFQCD_Pt_80to120_MET40_step1_v1/pela-VBFQCD_Pt_80to120_MET40_step3_v1-3664d28163503ca8171ba37083c39fc9/USER
QCD-Pt-120to170	/VBFQCD_Pt_120to170_MET40_step1_v1/pela-VBFQCD_Pt_120to170_MET40_step3_v1-3664d28163503ca8171ba37083c39fc9/USER
QCD-Pt-170to300	/VBFQCD_Pt_170to300_MET40_step1_v1/pela-VBFQCD_Pt_170to300_MET40_step3_v1-3664d28163503ca8171ba37083c39fc9/USER
QCD-Pt-300to470	/VBFQCD_Pt_300to470_MET40_step1_v1/pela-VBFQCD_Pt_470to600_MET40_step3_v1-3664d28163503ca8171ba37083c39fc9/USER
QCD-Pt-470to600	/VBFQCD_Pt_470to600_MET40_step1_v1/pela-VBFQCD_Pt_470to600_MET40_step3_v2-3664d28163503ca8171ba37083c39fc9/USER

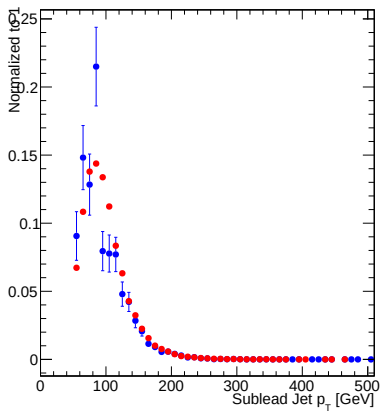
Sample	Ev. Gen.	Filter Eff.	Events	XS [pb]	Eq. Lumi. [fb^{-1}]
QCD-Pt-80to120	39376000000	0.000049	1614416	1033680	38.09
QCD-Pt-120to170	7000000000	0.000283	2051000	156293.3	44.79
QCD-Pt-170to300	1375000000	0.000987	1391500	34138.15	40.28
QCD-Pt-300to470	80000000	0.002659	207840	1759.549	45.47
QCD-Pt-470to600	25000000	0.004127	104675	113.8791	219.53

Kinematics Distributions

$\Delta\eta$ cut: Jet_1 p_T

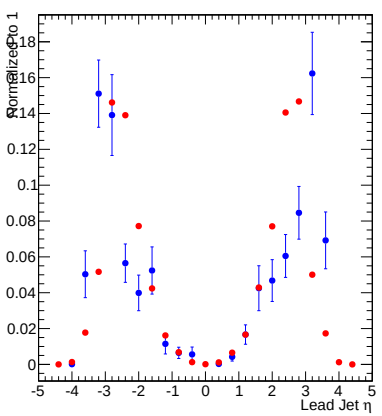


$\Delta\eta$ cut: Jet_2 p_T

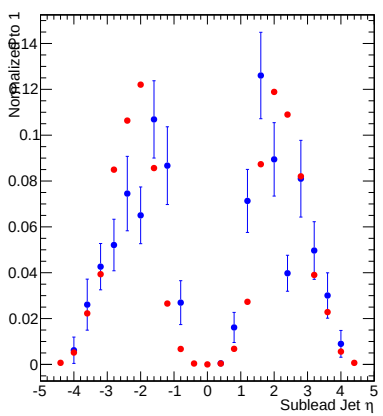


- This plots are produced after jets selection (p_T and η) and $\Delta\eta$ cuts over leading jets which is the last cut where we have significant statistics to compare Inclusive and VBF-like samples
- Legend:
  - BLUE: QCD Inclusive
  - RED: QCD VBF-like
- Jet p_T, Jet η variables seem o match to a reasonable level.

$\Delta\eta$ cut: Jet_1 η

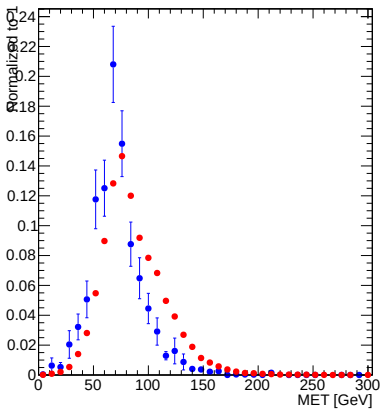


$\Delta\eta$ cut: Jet_2 η

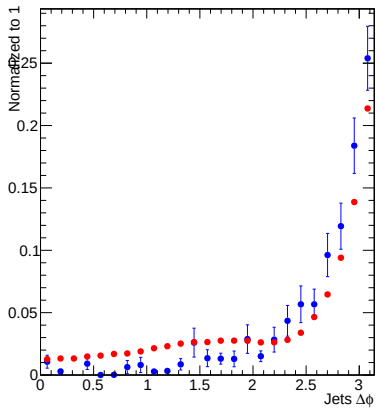


Kinematics Distributions

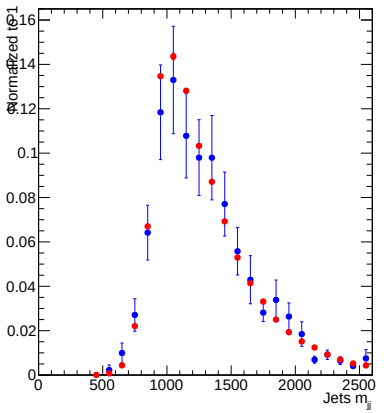
$\Delta\eta$ cut: MET



$\Delta\eta$ cut: $\Delta\Phi_{jj}$



$\Delta\eta$ cut: M_{jj}



- This plots are produced after jets selection (p_T and η) and $\Delta\eta$ cuts over leading jets which is the last cut where we have significant statistics to compare Inclusive and VBF-like samples
- Legend:
  - BLUE: QCD Inclusive
  - RED: QCD VBF-like
- m_{jj} variables seem o match to a reasonable level
- MET does not match well since we have not applied at least the MET cut yet.
- $\Delta\phi_{jj}$ seems to start to match but we can only compare with more cuts applied and at this level we will not have statistics anymore.

Imperial College London

MC Yields Comparison

Absolute Yields

	80-120	80-120	120-170	120-170	170-300	170-300	300-470	300-470	470-600	470-600
Sample	Inc	VBF	Inc	VBF	Inc	VBF	Inc	VBF	Inc	VBF
HLTMetClean	1509	118990	4595	257091	11866	271997	39157	56799	50413	28823
JetPair	587	71790	2337	163602	6405	168792	21505	32431	27991	16280
DEta	189	31982	543	43187	647	16988	354	539	35	38
MET	1	852	11	1948	39	1543	48	108	8	11
TightMjj	0	372	7	1231	35	1380	48	108	8	11
CJVpass	0	39	4	171	19	282	18	21	2	2
DPhi	0	12	0	18	0	6	0	0	0	0

Number of entries for each QCD p_T hat for after several cuts in current cut flow

Weighted Yields

	80-120	80-120	120-170	120-170	170-300	170-300	300-470	300-470	470-600	470-600
Sample	Inc	VBF	Inc	VBF	Inc	VBF	Inc	VBF	Inc	VBF
HLTMetClean	2436603.32	55846.48	1446548.90	104815.88	957742.98	124768.50	166055.20	23030.77	21139.22	2394.68
JetPair	469225.26	12722.61	364877.70	27234.70	260740.43	35168.11	50960.28	6958.12	7709.01	934.73
DEta	276326.77	7371.62	142118.25	9995.54	42229.51	4987.06	1233.85	153.28	12.45	2.26
MET	1.50	300.35	4672.18	682.16	3577.84	661.70	232.67	43.28	4.06	0.82
TightMjj	0.00	154.99	3625.30	464.06	3309.76	597.43	232.67	43.28	4.06	0.82
CJVpass	0.00	15.17	1858.14	63.84	1773.69	118.74	81.75	9.61	0.92	0.19
DPhi	0.00	5.10	0.00	8.23	0.00	2.53	0.00	0.00	0.00	0.00

Weighted (trigger, PU, ID, cross section) number of events for each QCD p_T hat for after several cuts in current cut flow.

Imperial College London

Some Conclusions

Conclusions

- Kinematics shapes seem to tend to match to inclusive samples after a few cuts are applied.
- After all selection cuts have been applied (even after MET cut) it is difficult to compare samples because QCD Inclusive sample become heavily suppressed.
- How ever there are significant yield differences specially evident in the QCD 470-600 *GeV* where QCD Inc samples have statistics comparable with 2012 data.

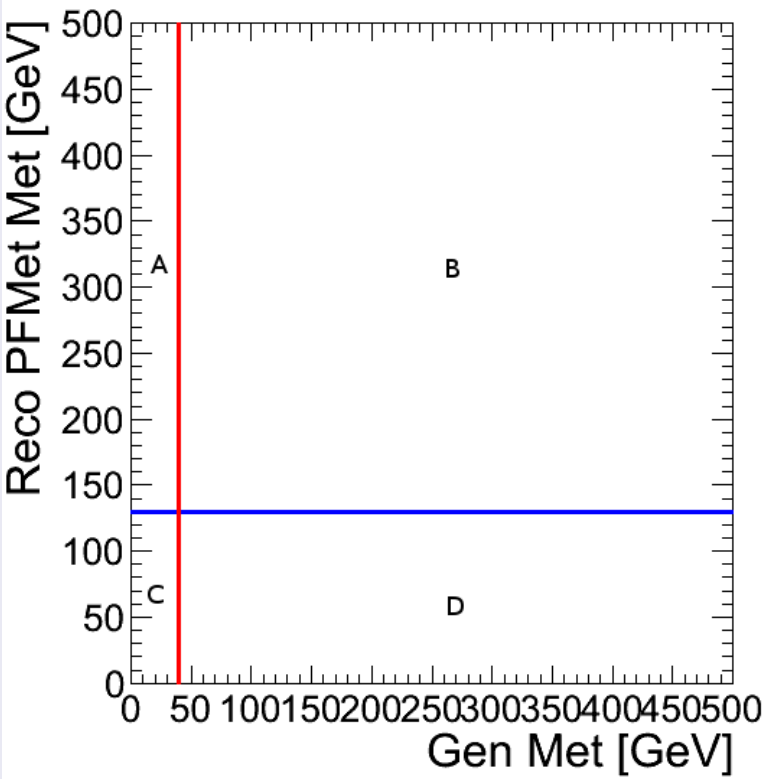
Plans

- It was suggested that this discrepancies may be cause by fake MET events. So I started a GenMET vs RecoMET study.
- Because QCD Inclusive samples are not enough, we need to compare QCD VBF-like samples with data, for this we need to choose a QCD dominated region.

Fake met contribution study

To evaluate how much events pass out analysis cut of $MET > 130\ GeV$ that have a significant contribution of fake MET which the new QCD VBF-like samples will not be able to simulate we need to look at the inclusive samples.

Area Definition



We can define 4 areas:

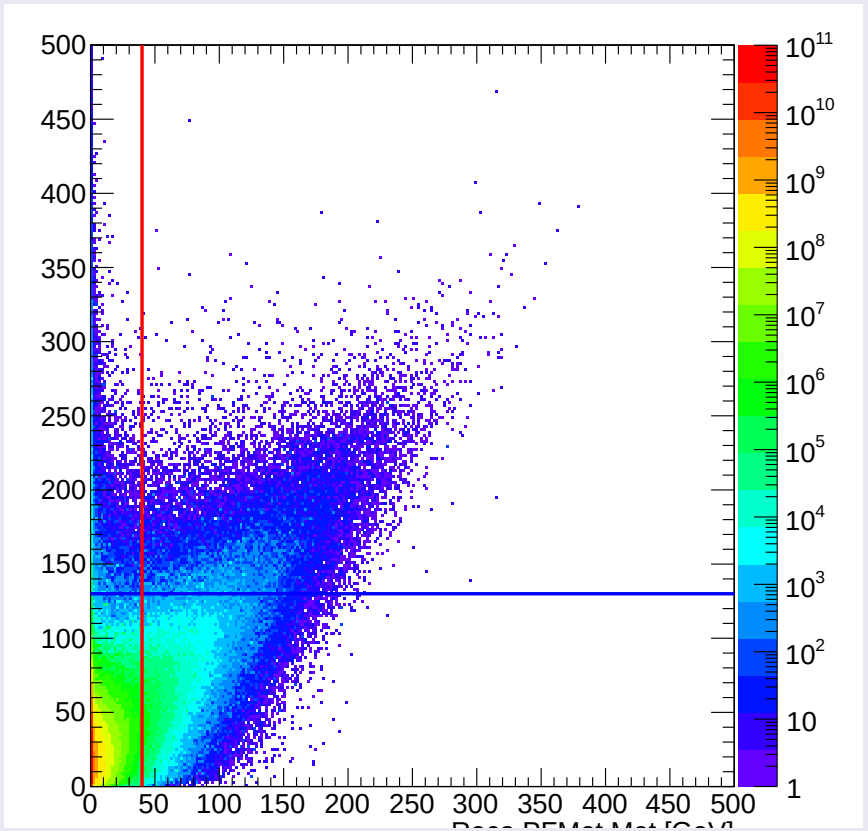
- A: Accepted by Analysis MET cut but rejected by GenMET Filter
- B: Accepted by Analysis MET cut and accepted by GenMET Filter
- C: Rejected by Analysis MET cut and rejected by GenMET Filter
- D: Rejected by Analysis MET cut but accepted by GenMET Filter

From this we can define the B area normalization factor as $\frac{A+B}{B}$

Gen Met Vs Reco MET I

Plots here do now have any weighting but cross section since filters will operate over genEvents with no weigting this (so this is just a scaling). Plot on the right Adds all the QCD inclusive p_T hats taking into account the relative cross section.

QCD Inc 80-600 *GeV*



- There are clearly 2 population
  - Events with real met along the diagonal of the plot
  - Events with fake met along 0 gen MET vertical line
- VBF samples only have events present on B area

We can see in both plots that there is a significant population in both areas A and B.

Gen Met Vs Reco MET - Into Numbers

Now we can calculate what is the percentage of events in each area and the normalization factor for B are.

Areas and normalization factor

Sample	A	B	C	D	Factor
MC_QCD-Pt-30to50-pythia6	0.000000	0.000000	0.999997	0.000003	inf
MC_QCD-Pt-50to80-pythia6	0.000000	0.000000	0.999884	0.000116	nan
MC_QCD-Pt-80to120-pythia6	0.000006	0.000001	0.998800	0.001193	5.625
MC_QCD-Pt-120to170-pythia6	0.000065	0.000024	0.995715	0.004195	3.676
MC_QCD-Pt-170to300-pythia6	0.000684	0.000281	0.989966	0.009069	3.432
MC_QCD-Pt-300to470-pythia6	0.005185	0.002331	0.976764	0.015721	3.224
MC_QCD-Pt-470to600-pythia6	0.016652	0.005900	0.959474	0.017973	3.823
MC_QCD-Pt-600to800-pythia6	0.034093	0.008591	0.939409	0.017906	4.969
MC_QCD-Pt-800to1000-pythia6	0.068863	0.011177	0.903115	0.016845	7.161
MC_QCD-Pt-1000to1400-pythia6	0.117719	0.012717	0.854500	0.015063	10.257
MC_QCD-Pt-1400to1800-pythia6	0.202556	0.014259	0.770444	0.012741	15.206
MC_QCD-Pt-1800-pythia6	0.285060	0.015070	0.688829	0.011041	19.916
Total	0.000001	0.000000	0.999954	0.000044	4.761343

Table : Relative are for A, B, C and D areas and factor to normalize B area to A+B (normalize QCD VBF samples to inclusive at $MET > 130\ GeV$ cut.

The normalization factors seem to be approximatly of the order of the discrepancy in yields seen last week.

Aplying Factors to last week tables

Considerations

- Factors are calculated for uncorrected (PU, Trigger and ID) events and applied to tables from last week which is to some level wrong but at first approximation gives an idea of the effect of this normalization factors.
- Factors only acount for events lost due to fake MET which is just one of the 2 filters applied, the VBF QCD jets filters while have its own losses which need to be accounted in parallel.

Correcting a MET cut level

	80-120	80-120	120-170	120-170	170-300	170-300	300-470	300-470	470-600	470-600
Cut	Inc	VBF	Inc	VBF	Inc	VBF	Inc	VBF	Inc	VBF
MET (Last Week)	1.50	300.35	4672.18	682.16	3577.84	661.70	232.67	43.28	4.06	0.82
MET (Apply factor)	1.50	1689.46	4672.18	2507.62	3577.84	2270.95	232.67	139.53	4.06	3.13

Number are now much closer to match, execept on the 80-120 GeV bin where QCD inclusive only has one event highly suppressed but non xsec weights. Note bin 470-600 GeV where both QCD Inc and QCD VBF have enough statistics to simulate 20 fb^{-1} now are much closer.

Summary and next steps

Summary:

- QCD VBF samples show some promise in describing the QCD contribution to our analysis (some event on the sample pass the whole selection)
- There is some discrepancy on the comparison between QCD Inclusive and VBF-like samples but that seams to be due to fake met contribution.
- the GenMET vs RecoMET study clearly showed that we have 2 population of events real and fake met which are clearly separated which gives us hope that we may able to suppress the fake met events.

Next Steps:

- Add genJets filter study in order to fully understand VBF-like sample
- Compare QCD VBF-like sample with data in QCD dominated regions
- Determine sample normalization from data
- Attempt to suppress fake met population (in QCD Inclusive sample) using MVA or variables (MET significance, MET-Dijet angle, etc) to suppress regions sample cannot model.

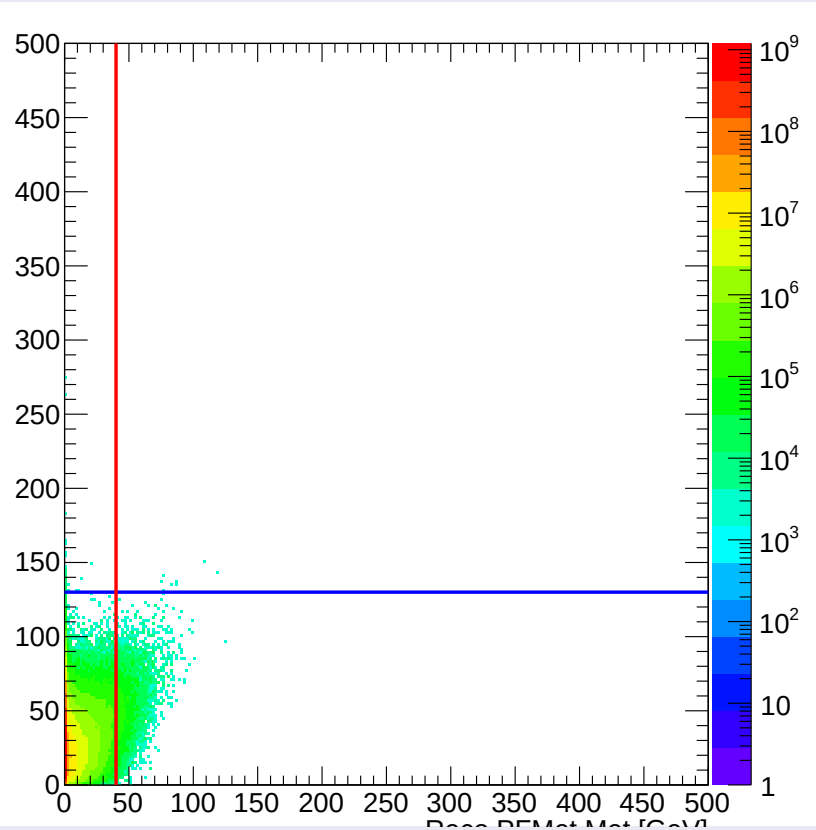
Imperial College London

Backup Slides

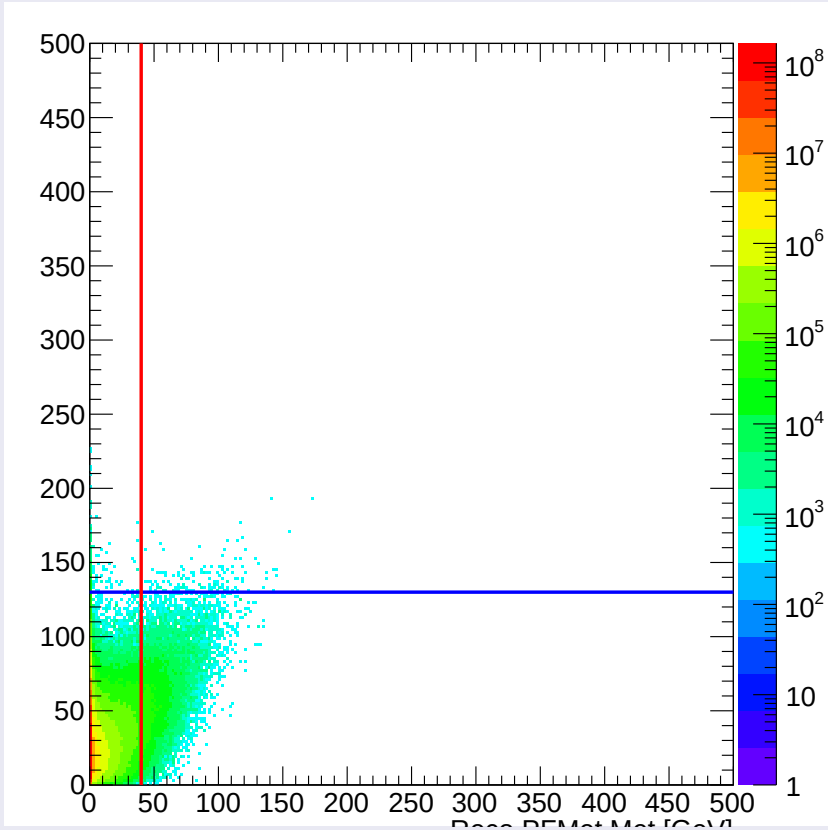
Gen Met Vs Reco MET I

Plots here do now have any weighting but cross section since filters will operate over genEvents with no weigting this (so this is just a scaling).

80to120



120to170

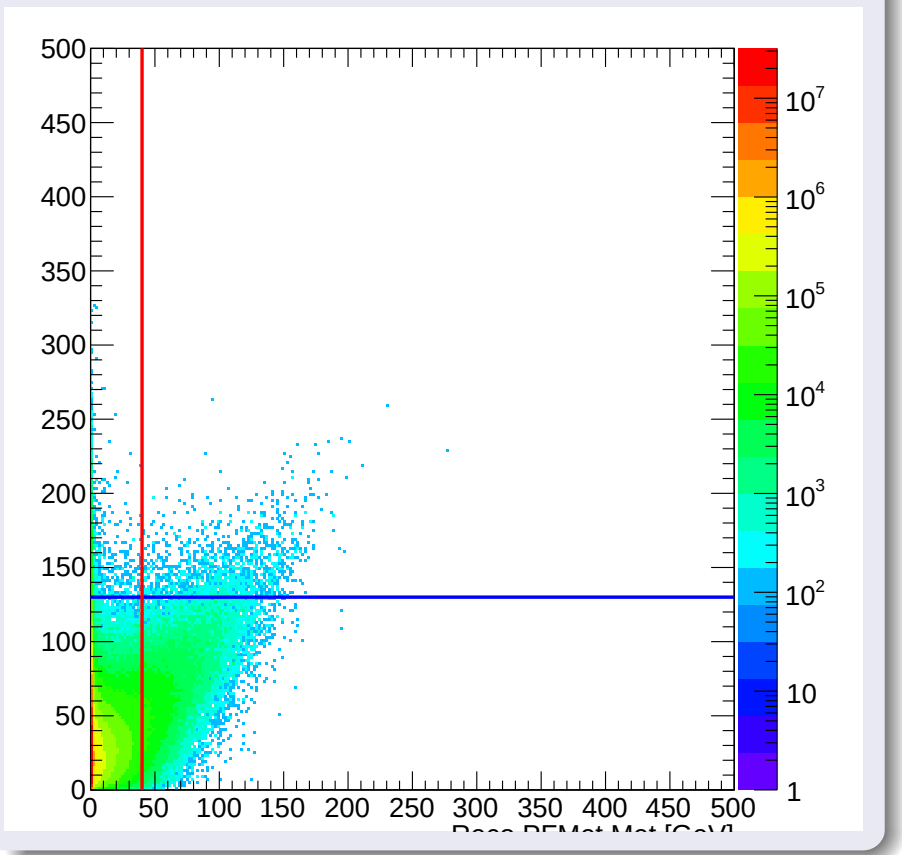


For both this p_T hats most event fall on zone C or D and are rejected analysis cut.

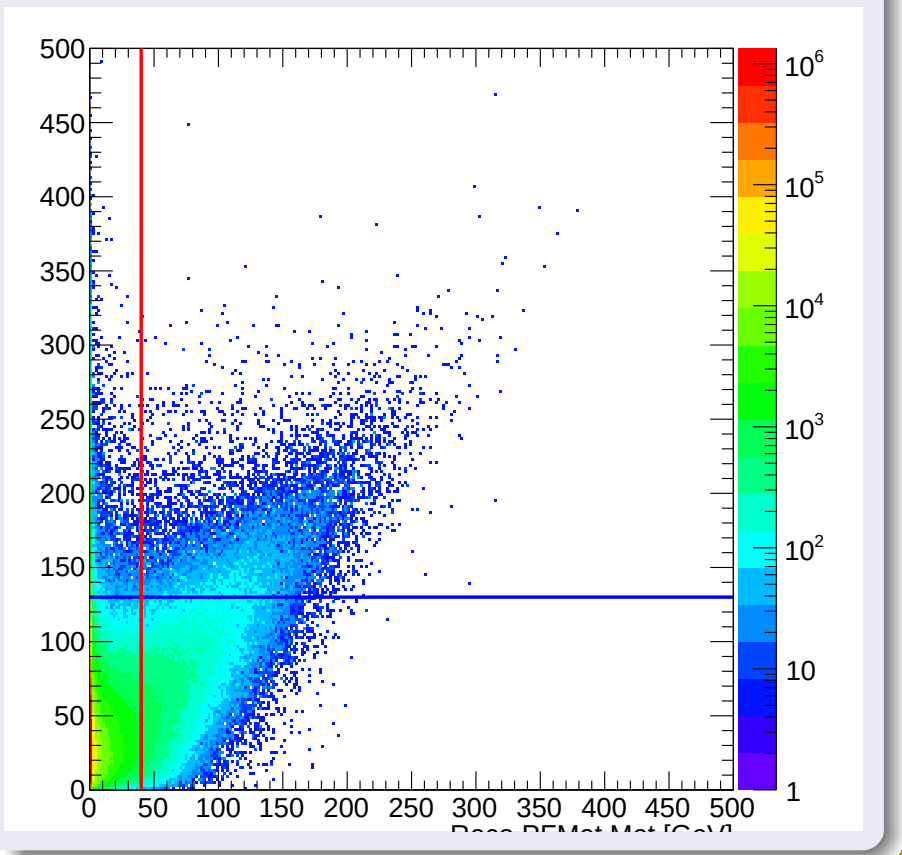
Gen Met Vs Reco MET II

Plots here do now have any weighting but cross section since filters will operate over genEvents with no weigting this (so this is just a scaling).

170to300



300to470

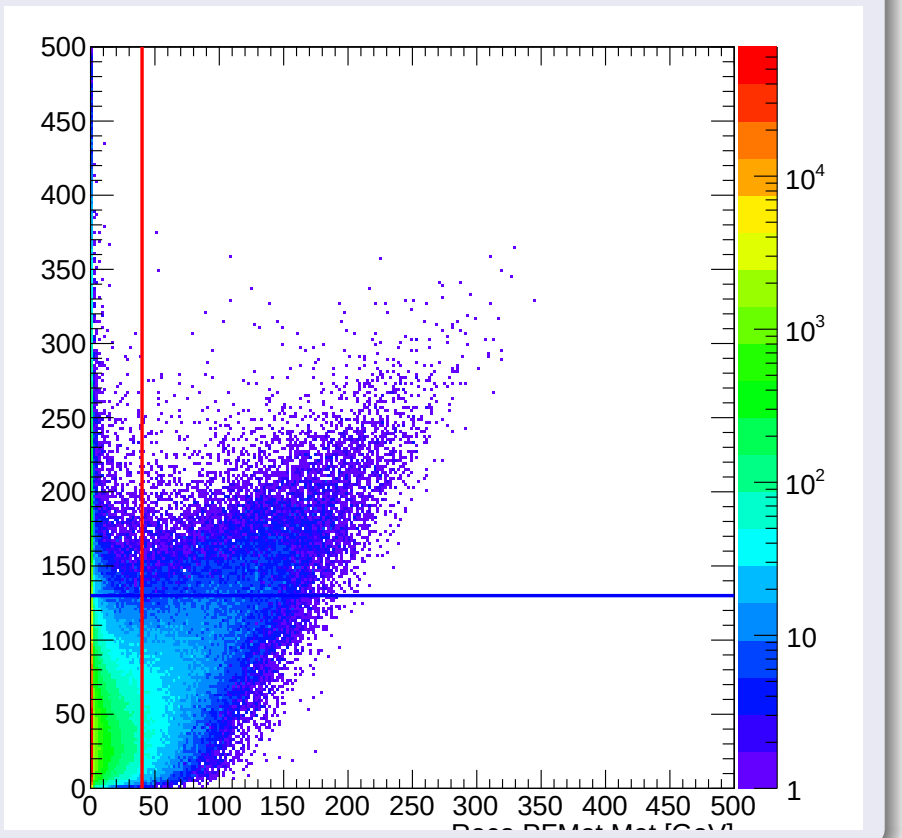


We can se for this p_T hats a significant population of events now in areas A and B.

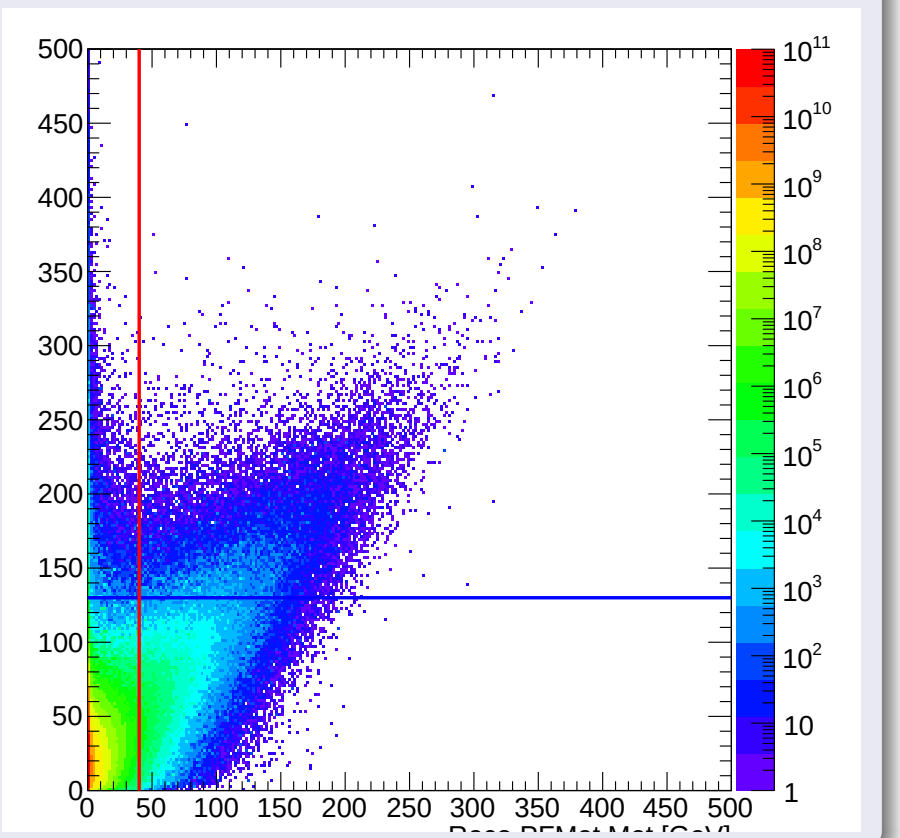
Gen Met Vs Reco MET III

Plots here do now have any weighting but cross section since filters will operate over genEvents with no weigting this (so this is just a scaling). Plot on the right Adds all the QCD inclusive p_T hats taking into account the relative cross section.

470to600



Total



We can see in both plots that there is a significant population in both areas A and B.