QCD VBF+MET samples for Run2 Update

João Pela

Imperial College London

2015-04-28

Introduction

During Run I a set of QCD samples with VBF like jets and real MET was generated, which allowed:

- Understand real MET in QCD
- Indirectly "determine the importance" of fake MET events (which we were not modelling)

For Run II making such samples once again to be useful, but:

- Is it possible (cross section increase significantly)?
- What are the costs (CPU time and storage)?
- Is it possible to have samples what include fake MET?
- What cuts to apply? And what is the physics toll?

Run I samples

As a reminder here are the generator cuts used to generate the run I QCD samples.

MC Filter: Vectorial sum of neutrino E_T

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

MC Filter: Dijet Filter (AK5 GenJet no μ)

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

Filters and goals

Steps where filters can (easily...) be inserted:

- Monte Carlo Generation
- Level 1 Trigger
- High Level Trigger
- Offline

Goals for this samples:

- Avoid: Generator MET cut, (also pick fake MET events).
- Samples should be of similar size to the inclusive QCD samples.
- $\Delta\phi(jet - jet)$ cuts should be avoided if possible (Run I analysis uses $\Delta\phi(jet - jet)$ cut for data-driven QCD estimation).

Cross Sections

Cross Section for some QCD $p_\perp$ hats

	Cross Section [pb]		
$p_\perp$ hat [GeV]	8 TeV	13 TeV	Change
30-50	66285328	161500000.	+243.6%
50-80	8148778.0	22110000.	+271.3%
80-120	1033680.0	3000114.3	+290.2%
120-170	156293.3	493200.	+315.6%
170-300	34138.15	120300.	+352.4%
300-470	1759.549	7475.	+424.8%
470-600	113.8791	587.1	+515.5%
600-800	26.99	167.	+618.7%

As expected cross section for the this QCD $p_\perp$ hats increase significantly from 8 to 13 TeV.

Events for 10 and 30 fb^-1

pT Hat	X-Section (pb)	10 fb-1	30 fb-1
30-50	1.62×10^8	1.62×10^{12}	4.85×10^{12}
50-80	2.21×10^7	2.21×10^{11}	6.63×10^{11}
80-120	3.00×10^6	3.00×10^{10}	9.00×10^{10}
120-170	4.93×10^5	4.93×10^9	1.48×10^{10}
170-300	1.20×10^5	1.20×10^9	3.61×10^9
300-470	7.48×10^3	7.48×10^7	2.24×10^8
470-600	587.1	5.87×10^6	1.76×10^7
600-800	167	1.67×10^6	5.01×10^6

Table : Quantity of event for each of the studied QCD $p_\perp$ hats for 10 and 30 fb^-1 of integrated luminosity.

Knowing that the current QCD samples sizes (470-600: 2.9M and 600-800: 2.8M) we can conclude:

- For 10 fb^-1 we need to simulate up to bin 470-600
- For 30 fb^-1 we need to simulate up to bin 600-800

Unfiltered production: Hardware

A single job for the whole simulation chain for 100 events was submitted to CERN lxbatch.

- GEN, SIM, DIGI, L1, DIGI2RAW, HLT:GRun
- RAW2DIGI, L1Reco, RECO

Hardware characteristics

$p_{\perp}$ hat		System Characteristics		
Min	Max	CPU Model	Core	RAM (kB)
30	50	Intel Core i7 9xx	16	24023052
50	80	Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz	32	62533844
80	120	AMD Opteron(TM) Processor 6276	32	62533828
120	170	Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz	8	62533828
170	300	AMD Opteron(TM) Processor 6276	32	62533828
300	470	Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz	16	31225964
470	600	Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz	32	62533844
600	800	Intel Core i7 9xx	16	24023052

To note that at lxplus (which I will assume is representative of the grid resources) machines can be very different in terms of CPU and number of cores. Differences were observed in CPU time between machines executing exactly the same code of +50% (sometimes as high as +100%).

Imperial College London

Unfiltered production: Step 1

Here are the statistics for the step 1 computing

Step 1 statistics

	CPU Times					
$p_\perp$ hat	Total	Total Event	Avg Event	Non-Event	Ev. Size (MB)	Total Time (s)
30-50	7520.65	7394.73	73.9473	125.92	1.28	1.19×10^{14}
50-80	7462.92	7369.9	107.695	93.02	1.32	2.38×10^{13}
80-120	12902	12739.3	127.393	162.7	1.44	3.82×10^{12}
120-170	8302.49	8216.92	82.1692	85.57	1.52	4.05×10^{11}
170-300	14636.6	14475.7	144.757	160.9	1.65	1.74×10^{11}
300-470	11611.5	11523.6	115.236	87.9	1.77	8.61×10^{9}
470-600	13511.4	13413.7	134.137	97.7	1.89	7.88×10^{8}
600-800	15851.5	15726.5	157.265	125	1.92	2.63×10^{8}
Average	11474.9	11357.5	117.8	117.3	1.6	

Some conclusions

- It would be impossible to process every single event, it would take several millennia on a single CPU. We need some kind of gen filter.
- On average events take between 1 and 3 minutes to go over the whole step.
- Event size is under 2 MB (this is normal) and increases with $p_\perp$ hat.

NOTE: I am including PU at average 30 interactions.

Unfiltered production: Step 2

Here are the statistics for the step 2 computing

Step 2 statistics

	CPU Times				
$p_\perp$ hat	Total	Total Event	Avg Event	Non-Event	Ev. Size (MB)
30-50	1074.55	1023.63	10.2363	50.92	0.30
50-80	1163.04	1108.85	11.0885	54.19	0.32
80-120	2395.97	2308.39	23.0839	87.58	0.34
120-170	1276.21	1232.36	12.3236	43.85	0.35
170-300	2632.94	2548.5	25.485	84.44	0.37
300-470	1832.55	1776.45	17.7645	56.1	0.39
470-600	2130.65	2055.35	20.5535	75.3	0.40
600-800	2726.66	2658.05	26.5805	68.61	0.41
Average	2269.0	1838.9	18.4	65.1	0.36

Some conclusions

- This step will not be a problem since it is after the selection
- Event size is under 0.5 MB (this is normal) and increases with $p_\perp$ hat.

First working point

GenJets only ($p_\perp > 50$, $|eta| < 4.75$, $\Delta_\eta > 3.5$, $\Delta_\phi < 1.5$, $m_{jj} > 1000$)

Step 1 statistics: 1M events

	CPU Times					
$p_\perp$ hat	Passed	Filter Eff Event	Total Filter	Ev Total S1 CPU (s)	S1 5k CPU (h)	S1 5k CPU (d)
30-50	6	6.00×10^{-6}	9.69×10^6	7.17×10^8	39.81	1.66
50-80	417	4.17×10^{-4}	9.22×10^7	9.93×10^9	551.63	22.98
80-120	3204	3.20×10^{-3}	9.61×10^7	1.22×10^{10}	680.30	28.35
120-170	8857	8.86×10^{-3}	4.37×10^7	3.59×10^9	199.41	8.31
170-300	18828	1.88×10^{-2}	2.27×10^7	3.28×10^9	182.15	7.59
300-470	36953	3.70×10^{-2}	2.76×10^6	3.18×10^8	17.68	0.74
470-600	46455	4.65×10^{-2}	2.73×10^5	3.66×10^7	2.03	0.08
600-800	47430	4.74×10^{-2}	7.92×10^4	1.25×10^7	0.69	0.03
Total				3.01×10^{10}	1673.71	69.74

- NOTE 1: 10x the statistics of last week!
- NOTE 2: Time values do not include generator level times just reco times (should be dominant for higher $p_\perp$ hats.

We could make this samples with 5k CPU in about 2 months, can be done privately with some work, but can be done easily by central production!

Other working points

Filter statistics

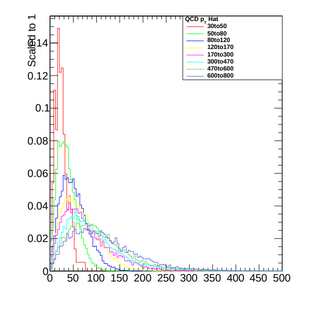
	CPU Times			
Filter	Ev Total S1 CPU (s)	S1 5k CPU (h)	S1 5k CPU (d)	Note
Pt20_Eta5p0_DEta3p2_Mjj700	4.03E+12	223652.83	9318.87	Same as Run I without Generator MET
Pt50_Eta4p75_DEta3p5_Dphi1p5_Mjj1000	3.01E+10	1673.71	69.74	Current working point
Pt40_Eta4p75_DEta3p5_Mjj600	1.00E+12	55531.17	2313.80	Same as new HLT path
Pt40_Eta4p75_DEta3p5_Dphi1p5_Mjj600	1.90E+11	10539.90	439.16	Same as new HLT path + $\Delta\phi$
Pt40_Eta4p75_DEta3p5_Dphi1p5_Mjj1000	6.87E+10	3816.26	159.01	Lower 10 GeV Dijet $p_\perp$
Pt50_Eta4p75_DEta3p5_Dphi2p0_Mjj1000	4.10E+10	2280.27	95.01	+0.5 $\Delta\phi$ cut
Pt50_Eta4p75_DEta3p5_Dphi2p5_Mjj1000	6.24E+10	3467.57	144.48	+1.0 $\Delta\phi$ cut
Pt50_Eta4p75_DEta3p5_Mjj1000	2.34E+11	13025.79	542.74	No $\Delta\phi$ cut

- Changing any studied parameter increases greatly the time to be spent generating
- Touching $\Delta\phi$ increases processing time about 1 month per 0.5 relaxation
- Currently proposed working point same as last week

Offline Filter Efficiency

For working point $p_\perp > 50$, $|eta| < 4.75$, $\Delta_\eta > 3.5$, $\Delta_\phi < 1.5$, $m_{jj} > 1000$

Level 1 MET



- Here L1 MET is the one the comes from this CMSSW version, its NOT the latest greatest.
- We can see that many events have high L1 MET (much more than just real MET)

Level 1 MET

- L1T MET $>$ 40: 1.48E+08 (biggest sample 80-120 with 6.01E+07)
- L1T MET $>$ 70: 6.13E+07 (biggest sample 80-120 with 2.39E+07)

60M for 8 samples is not unreasonable!

HLT PFMET170

Efficiency

$p_\perp$ hat	Generated	Passed	Filter Eff
30-50	80000000	0	0
50-80	20000000	0	0
80-120	3800000	0	0
120-170	2400000	0	0
170-300	700000	8	0.000011
300-470	460000	32	0.000070
470-600	320000	112	0.000350
600-800	270000	194	0.000719

- Most events are killed and we now have dedicated path for this

Summary

Summary:

- Found a working point with is feasible with no MET cut at generator level with the caveat that it has a delta phi cut.
- A document including all the information is being written and will be sent around soon.

Next steps:

- Study decrease of m_{jj} and $\Delta\eta$ (done not digested)
- Finish offline efficiency study