QCD Samples with VBF-like jets and MET - Feasibility Study for Run II

Joao Pela
Imperial College London
joao.pela@cern.ch

April 24, 2015

Abstract

Study on the feasibility of QCD samples with VBF-like jets and MET characteristics to be used in the CMS VBF Higgs to Invisible analsysis during the LHC Run II period. This study will cover cost (physics usability, computing and disk storage) of several possibilities to make such samples.

1 Introduction

Simulating and reconstructing quantities of QCD events in quantities comparable to the ones produced at the LHC experiments is impractical. At every second of LHC physics operation several million bunch crossings happen each one containing several collisions. With current hardware it takes in excess of one minute to fully simulate one of such bunch crossings. This constraints lead to QCD events being simulated in $p_\perp$ hats, where the first collision is generated in a specific $p_\perp$ range and then several other collision are added as pile-up. This extra collisions are generated without any constraints in $p_\perp$.

This bin method allows the user to have samples with increasing energy and study the influence of each of them in their own analysis. As a practical example a user does not need to look over millions of events to find high energy jets, he can just test the higher $p_\perp$ hats as needed.

On the other hand, there are some analysis that require jets and/or MET with very low values where this inclusive QCD samples will not have enough statistics to provide insight into the QCD behaviour. The CMS VBF Higgs to invisible is one of such analysis.

During the preparation the Run I VBF Higgs to invisible analysis a set of QCD samples with VBF like jets and real MET was generated. This samples allowed to understand the mechanisms that create real MET in QCD and how those could be mitigated.

The analysis is now preparing for Run II it was considered once again to be useful to have similar samples remade and possibly extended. It was identified that not only real MET is significant but also fake MET coming from detector miss measurement and as such a sample that could simulate such effects would be of great interest.

This study investigates the alternatives and attempts to quantify the costs of of producing such samples.

2 Objectives and methods

The objective of this study is to identify possible working points for the production of QCD samples with VBF-like jets and MET.

Since this type of events is very rare, we need to filter this events in order to have manageable sample sizes. There are also computing limitations, while running a Monte Carlo generator is not very costly in terms of CPU time, running the full CMS simulation is.

Steps where filters can be inserted:

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

To maximise physics usability of this samples the one should consider:

- Avoid any cut on generator level MET, in order to pick up fake MET events.
- Samples should be of similar size to the inclusive QCD samples
- $\Delta\phi(jet - jet)$ cuts should be avoided since the Run I analysis uses inverted cuts in this variable on a data-driven QCD estimation.

3 Cross section

The cross section for each QCD $p_\perp$ hat at

	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%

Table 1: Used datasets and their respective integrated luminosities after applying certified for physics filtering. Each dataset corresponds to a recording period/era of the 2012-13 data acquisition run.

As expected the cross section increase significantly from 8 TeV to 13 TeV and this effect becomes more and more significant as you go up in $p_\perp$ hat.

4 Run I VBF QCD samples characteristics

In the section we will summarize the characteristics of VBF QCD samples produced for the Run I analysis.

Two filters were applied to data, one over the generator level met and another over generator level jets. Here, generator level MET is defined as the vectorial sum of the $p_\perp$ of all neutrinos on the event. The implemented filters selected events according to the following characteristics:

MC Filter: Vectorial sum of neutrino E_T

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

$p_\perp$ Hat	Ev. Gen.	Filter Eff.	Events	XS $[pb]$	Eq. Lumi. $[fb^{-1}]$
80-120	39376000000	0.000049	1614416	1033680	38.09
120-170	7000000000	0.000283	2051000	156293.3	44.79
170-300	1375000000	0.000987	1391500	34138.15	40.28
300-470	80000000	0.002659	207840	1759.549	45.47
470-600	25000000	0.004127	104675	113.8791	219.53

Table 3: TODO

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

MC Filter: Dijet Filter

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

By applying this filters we were able to pass only a small fraction of the events to the detector simulation.

The resulting samples were publish in the CMS physics analysis database with url:

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

With the following database names:

$p_\perp$ hat	Identifier
80-120	/VBFQCD_Pt_80to120_MET40_step1_v1/pela-VBFQCD_Pt_80to120_MET40_step3_v1-3664d28163503ca8171ba37083c39fc9/USER
120-170	/VBFQCD_Pt_120to170_MET40_step1_v1/pela-VBFQCD_Pt_120to170_MET40_step3_v1-3664d28163503ca8171ba37083c39fc9/USER
170-300	/VBFQCD_Pt_170to300_MET40_step1_v1/pela-VBFQCD_Pt_170to300_MET40_step3_v1-3664d28163503ca8171ba37083c39fc9/USER
300-470	/VBFQCD_Pt_300to470_MET40_step1_v1/pela-VBFQCD_Pt_470to600_MET40_step3_v1-3664d28163503ca8171ba37083c39fc9/USER
470-600	/VBFQCD_Pt_470to600_MET40_step1_v1/pela-VBFQCD_Pt_470to600_MET40_step3_v2-3664d28163503ca8171ba37083c39fc9/USER

Table 4: TODO

5 Phys14 QCD samples

Before we start investigating possible production filters we need to define a baseline production. For this purpose, the Phys14 samples were chosen since this production was fairly recent and this samples were tested in a CMS wide exercise. Some details about this samples can be found in the table 5.

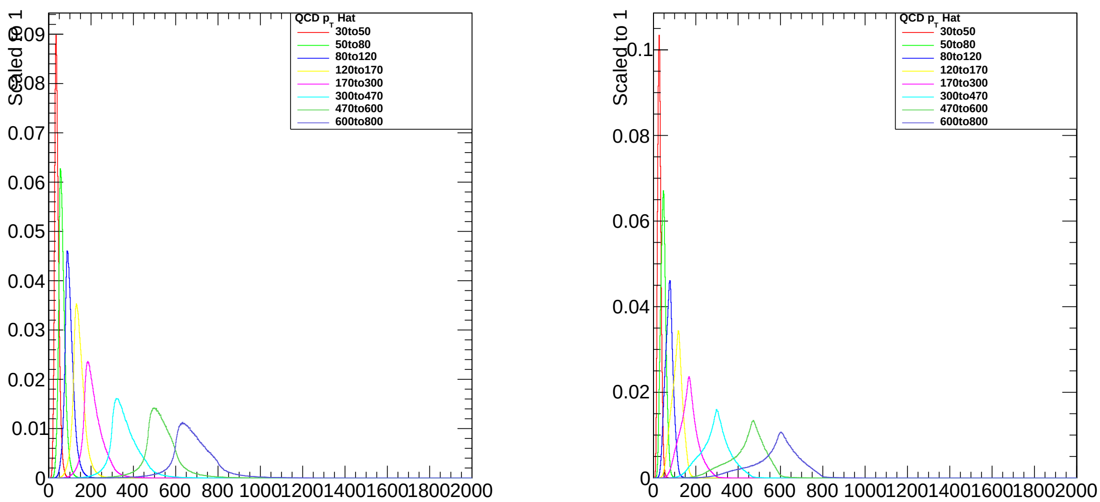
Dataset	Sample Size
/QCD_Pt-30to50_Tune4C_13TeV_pythia8/Phys14DR-AVE30BX50_tsg_castor_PHYS14_ST_V1-v2/AODSIM	5M
/QCD_Pt-50to80_Tune4C_13TeV_pythia8/Phys14DR-AVE30BX50_tsg_castor_PHYS14_ST_V1-v1/AODSIM	5M
/QCD_Pt-80to120_Tune4C_13TeV_pythia8/Phys14DR-AVE30BX50_tsg_castor_PHYS14_ST_V1-v1/AODSIM	5M
/QCD_Pt-120to170_Tune4C_13TeV_pythia8/Phys14DR-AVE30BX50_tsg_castor_PHYS14_ST_V1-v1/AODSIM	5M
/QCD_Pt-170to300_Tune4C_13TeV_pythia8/Phys14DR-AVE30BX50_tsg_castor_PHYS14_ST_V1-v1/AODSIM	3M
/QCD_Pt-300to470_Tune4C_13TeV_pythia8/Phys14DR-AVE30BX50_tsg_castor_PHYS14_ST_V1-v1/AODSIM	3M
/QCD_Pt-470to600_Tune4C_13TeV_pythia8/Phys14DR-AVE30BX50_tsg_castor_PHYS14_ST_V1-v1/AODSIM	3M
/QCD_Pt-600to800_Tune4C_13TeV_pythia8/Phys14DR-AVE30BX50_tsg_castor_PHYS14_ST_V1-v1/AODSIM	3M

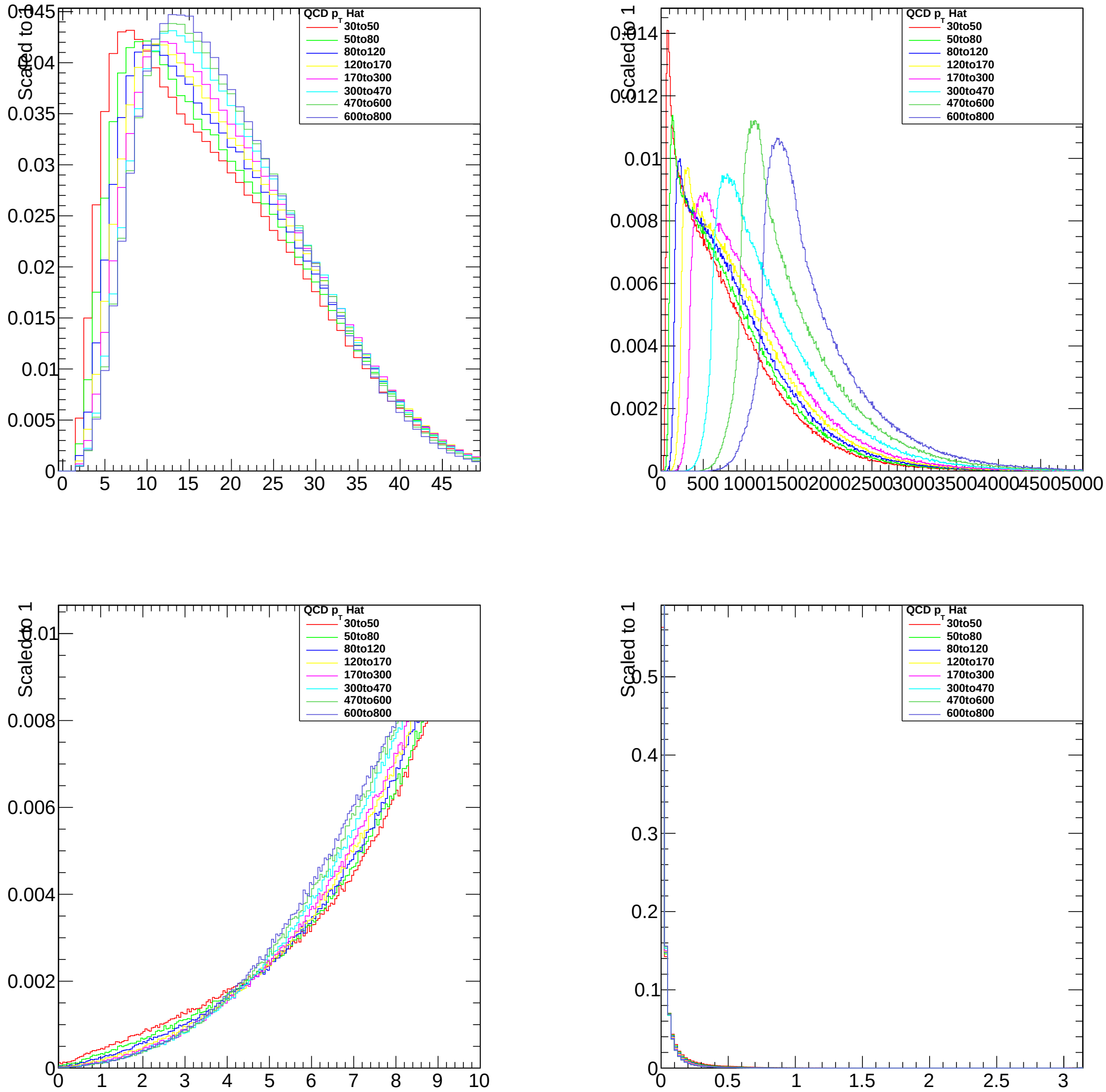
Table 5: Details on the inclusive QCD samples produces for the Phys14 exercise. Only the samples with average pile-up 30 and 50 ns separation are showed.

This samples were produced with CMSSW_7_2_0_patch1. We will be using the same version for this study since any produced test samples can be tested with the same receipts used to analyse Phys14 samples.

Unfortunately, it was found that this QCD samples were not created from scratch. They are simply a re-processing of previously produced samples produce in the Fall13 campaign with CMSSW_6_2_0_patch1. This means we cannot simply use the same configuration files used to produce Phys14 samples since this will not run any Monte Carlo generator but only re-process existing samples.

6 QCD pT hat characteristics





7 Producing unfiltered events

In order the understand the time cost of simulating events from scratch with CMSSW_7_2_0_patch1 configuration files were prepared using for each $p_{\perp}$ hat TODO l

It should be noted th

8 Generator level filters

Two generator level filters were developed to preform this studies/productions. They filtered event on the generator jets and MET.

$p_\perp$ 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

Table 6: Details on the hardware where the jobs for each $p_\perp$ got executed in the CERN lxplus batch system.

$p_\perp$ hat		Job Times		CPU Times				Size (MB)	
Min	Max	Total	Avg Event	Total	Total Event	Avg Event	Non-Event	Job	Avg Event
30	50	9985.93	99.8593	7520.65	7394.73	73.9473	125.92	128	1.28
50	80	10769.5	107.695	7462.92	7369.9	107.695	93.02	132	1.32
80	120	16576.5	165.765	12902	12739.3	127.393	162.7	144	1.44
120	170	10942.9	109.429	8302.49	8216.92	82.1692	85.57	152	1.52
170	300	18225.8	182.258	14636.6	14475.7	144.757	160.9	165	1.65
300	470	13346.1	133.461	11611.5	11523.6	115.236	87.9	177	1.77
470	600	15774.7	157.747	13511.4	13413.7	134.137	97.7	189	1.89
600	800	16422.1	164.221	15851.5	15726.5	157.265	125	192	1.92
Average		14005	140.05	11474.9	11357.5	117.8	117.3		

Table 7: Step 1

$p_\perp$ hat		Job Times		CPU Times				Size (MB)	
Min	Max	Total	Avg Event	Total	Total Event	Avg Event	Non-Event	Job	Avg Event
30	50	1134.6	11.346	1074.55	1023.63	10.2363	50.92	30	0.30
50	80	1348.35	13.4835	1163.04	1108.85	11.0885	54.19	32	0.32
80	120	2666.12	26.6612	2395.97	2308.39	23.0839	87.58	34	0.34
120	170	1452.54	14.5254	1276.21	1232.36	12.3236	43.85	35	0.35
170	300	2977.98	29.7798	2632.94	2548.5	25.485	84.44	37	0.37
300	470	3272.37	32.7237	1832.55	1776.45	17.7645	56.1	39	0.39
470	600	2499.6	24.996	2130.65	2055.35	20.5535	75.3	40	0.40
600	800	2800.28	28.0028	2726.66	2658.05	26.5805	68.61	41	0.41
Average		2269.0	22.7	1904.1	1838.9	18.4	65.1		

Table 8: Step 2

8.1 Generator Jets Filter

TODO: Description of the generator jet filter

8.2 Generator MET Filter

TODO: Description of the generator MET filter

A Code Setup

To reproduce this study you need to create a new CMSSW area, clone the git repository and get some configuration files from a remote server. To do this issue the following commands in a lxplus (SLC6) bash terminal:

```
cmsrel CMSSW_7_2_0_patch1
cd CMSSW_7_2_0_patch1/src/
cmsenv
git clone git@github.com:joaopela/VBFHiggsToInvisible.git
git checkout !!!TODO!!! <tag>
./VBFHiggsToInvisible/SampleProduction/scripts/getConfigFiles.sh
scramv1 b -j 5
```

The configuration files are located at:

```
VBFHiggsToInvisible/SampleProduction/test/13TeV/VBFQCD/
```

B Code Simulation

Below the cmsDriver.py code to generate the CMSSW configuration files to generate unfiltered QCD samples in the $p_{\perp}$ hat of 30 to 50 GeV. This generation/simulation has been made over two steps.

Step one comprises:

- GEN: Actual physics process simulation
- SIM: Geant simulation of the intereaction of particles with the detector
- DIGI: Simulation of the response of CMS electronics
- L1: Simulation of the Level 1 Trigger
- DIGI2RAW: Conversion of the Digital Signal back to "raw" signal
- HLT:GRun: High Level Trigger Simulation

```
cmsDriver.py Configuration/GenProduction/python/BTV-Fall13-00042-fragment.py
  --mc
  --fileout file:QCD_Pt-30to50_step1.root
  --eventcontent RAWSIM
  --customise SLHCUpgradeSimulations/Configuration/postLS1Customs.customisePostLS1,
      Configuration/StandardSequences/SimWithCastor_cff.customise
  --datatier GEN-SIM-RAW
  --pileup_input "dbs:/MinBias_TuneA2MB_13TeV-pythia8/Fall13-POSTLS162_V1-v1/GEN-SIM"
  --pileup 'AVE_35_BX_50ns,{"N":30}'
  --conditions PHYS14_ST_V1
  --step GEN,SIM,DIGI,L1,DIGI2RAW,HLT:GRun
  --magField 38T_PostLS1
  --geometry Extended2015
  --python_filename QCD_Pt-30to50_step1_cfg.py
  --no_exec
  -n 100
```

Step one comprises:

- RAW2DIGI: Simulation of RAW to DIGI process
- L1Reco: Reconstruction of Level 1 trigger quantities
- RECO: Reconstruction of the event

```
cmsDriver.py step2
  --mc
  --filein file:QCD_Pt-30to50_step1.root
  --fileout file:QCD_Pt-30to50_step2.root
  --eventcontent AODSIM
  --customise SLHCUpgradeSimulations/Configuration/postLS1Customs.customisePostLS1,
      Configuration/DataProcessing/Utils.addMonitoring
  --datatier AODSIM
  --conditions PHYS14_ST_V1
  --step RAW2DIGI,L1Reco,RECO
  --magField 38T_PostLS1
  --python_filename QCD_Pt-30to50_step2_cfg.py
  --no_exec
  -n 100
```

References

[1]