Imperial College London

## VBF Higgs to Invisible - Preapproval
HIG-14-038, AN-14-243

C. Asawatangtrakuldee, J. Brooke, D. Colling, G. Davies,
P. Dunne, A.M. Magnan, A. Nikitenko, Joao Pela

Why Higgs to Invisible?

Experimental motivation

- ▶ Current measurements of the 125 GeV Higgs boson are compatible with Standard Model (SM) expectations
- - large uncertainties can still accommodate significant beyond the SM (BSM) properties
- ▶ Additional Higgs bosons with exotic decays are not excluded
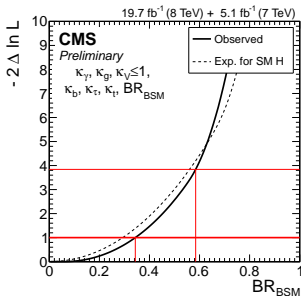

Theoretical motivation

- ▶ Many BSM theories predict Higgs boson decays to invisible final states:
- - e.g. SUSY, extra dimensions, fourth-generation neutrinos
- ▶ These final state particles are often dark matter candidates

Talk outline

- For run 1 we had two sets of triggers
  - Prompt trigger used for current published result: HIG-13-30
 - Parked triggers analysis progress presented today

Overview

- Reminder of prompt analysis
- Why we chose our new analysis strategy
- Details of parked analysis
  - Emphasis on changes from established prompt analysis
- Brief look at some other analysis techniques we are investigating
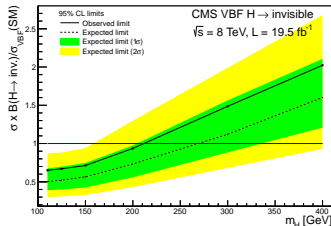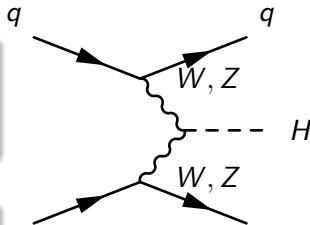
Prompt Analysis

- ▶ Single bin counting experiment
- - Signal region chosen to eliminate QCD and be above trigger turn ons
- ▶ Major backgrounds use data driven estimates:
- ▶ $Z \to \nu\nu, W \to \ell\nu$, QCD
- ▶ Minor backgrounds taken from MC:
- ▶ VV, $W\gamma$, $t\bar{t}$, single top
- ▶ Expected limit 49% at $m_H = 125$ GeV
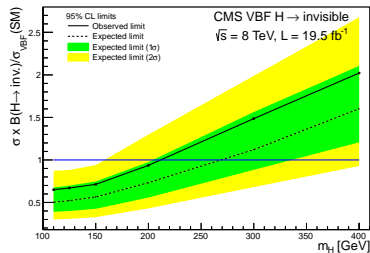
Data driven background estimation

$$W: N_S = N_S^{MC} \frac{N_C^{Data} - N_C^{Bkg}}{N_C^{MC}}$$

$$Z:$$
$$N_S^{Z \to \nu\nu} = \left(N_C^{Data} - N_C^{bkg} \right) \cdot \frac{\sigma(Z \to \nu\nu)}{\sigma(Z/\gamma^* \to \mu\mu)} \cdot \frac{\epsilon_S^{ZMC}}{\epsilon_C^{ZMC}}$$

Parked Triggers

- ▶ Use already analysed prompt trigger for run A
- ▶ One parked trigger for runs B and C, another for run D
- ▶ All parked and prompt triggers are seeded by L1ETM40
- ▶ Parked triggers have looser HLT thresholds
- ▶ This allows us to look at new regions of phase space and different analysis techniques

HLT

Run period	MET cut	dijet p_T cut	dijet mass cut
A	METnoMuons>65 GeV	DiPFJet40	MJJ800
B&C	N/A	DiJet35	MJJ700
D	N/A	DiJet30	MJJ700

Software framework strategy

Prompt analysis

- Two frameworks: Analyses A and B
- independent ntuples and analysis code

Parked analysis

- Insufficient manpower to maintain and develop two frameworks
- Moved to one fully developed framework
- New framework is development of analysis B and uses same ntuples
- Synchronised yields in signal and control regions between new framework and old analyses A and B
- Repeated expected limit calculation from HIG-13-030 analysis with the new framework and parked data
- Agrees with HIG-13-030 to within 2%, which is good given rereco, and change of global tag and triggers

Analysis strategy

Initial investigations

- Planned to define a loose pre-selection and model QCD shape
- Several options for analysis strategy:
  - Rectangular cuts and counting experiment
  - Rectangular cuts and shape experiment
  - MVA and counting experiment
  - MVA and shape experiment
- Due to trigger conditions no appropriate QCD control region found - details later

Final plan for parked data

- During QCD studies found new variables which remove QCD
  - Signal efficiency significantly higher than prompt selection
- Remaining backgrounds very signal like in variables studied so far
  - Opted for cut and count analysis

QCD options tried

Several methods tried to model QCD

Standard MC

- doesn't have enough events

Private VBF+MET enriched QCD MC sample

- Can only enrich in events with real met
- Can't model met from mismeasurement

Data-driven shape using different jet pairs in the event

- Jet kinematics are very biased
- Ordering in p_T and angle have been tried
- Reweighting individual distributions to fix others has been tried

Changes since prompt analysis

Trigger

- ▶ Parked trigger efficiency has been measured including variable correlation
  - This allows the trigger turn on region to be used

Signal region

- ▶ The signal region has been reoptimised for the looser parked triggers
  - New region uses new variable has higher signal efficiency with much less QCD

Background estimation

- ▶ A top control region has been added
- ▶ Minor modifications made to $W \to \tau\nu$ background estimation method
- ▶ QCD background estimation method changed
- ▶ $W\gamma$ contribution found to be modelled already by our $W \to \ell\nu$ Monte Carlo

Imperial College London

Trigger efficiency

- ▶ The variables used in prompt and parked triggers are highly correlated:
- - dijet mass, METnoMU, jet 2 p_T
- ▶ In the prompt analysis we neglected correlations and cut to ensure trigger was > 95% efficient
- ▶ For the parked analysis we use a 2D binning in dijet mass and jet 2 p_T
- - MJJ: 0,600,800,900,1000,5000
- - Jet 2 p_T: 30,40,50,60,1000
- ▶ In each bin we fit the METnoMU trigger turn on using an error function
- ▶ We then combine the turn ons from runs A, BC and D weighted by luminosity and apply this to MC events

METnoMU/GeV
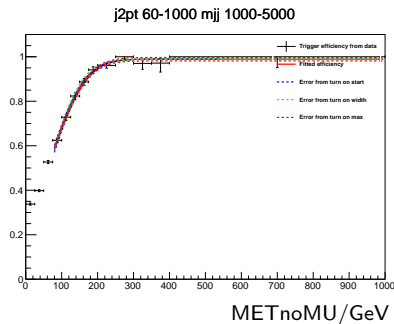
Preselection choice

- Trigger turn ons and detector acceptance impose the following cuts:
  - $\eta_{j1} \cdot \eta_{j2} < 0$, $|\eta_{j1,2}| < 4.7$, jet 1 $p_T > 50\,\text{GeV}$, $\Delta\eta_{jj} > 3.6$, jet 2 $p_T > 40\,\text{GeV}$, METnomu $> 90\,\text{GeV}$, $M_{jj} > 800\,\text{GeV}$

- As in the prompt analysis we also veto events with 'veto' electrons or muons

- QCD in plots is VBF enriched MC doesn't model all QCD

- Poor data-MC agreement from QCD contamination motivates the following additional cuts:

- $\frac{METnomu}{\sigma_{METnomu}} > 3.0$, $\text{Min}\Delta\phi(\text{all jets } p_T > 30\,\text{GeV}, METnomu) > 1.0$, $M_{jj} > 1000\,\text{GeV}$
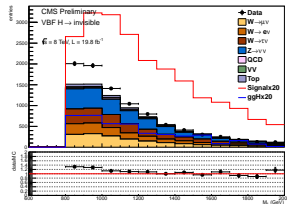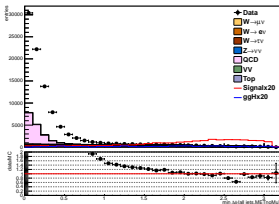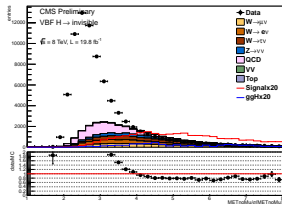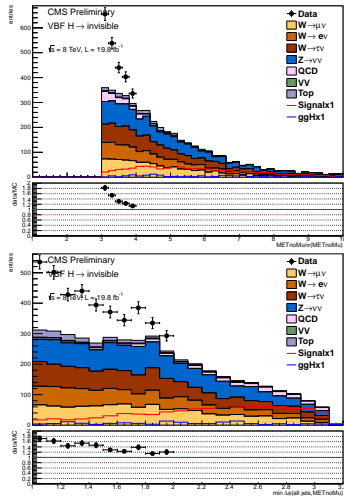
Signal region selection

- ▶ Can't model QCD shape so alter cuts to remove most QCD
- - Can then tolerate a larger uncertainty on QCD estimation
- ▶ Remaining QCD in region $\frac{METnoMU}{\sigma_{METnoMU}} < 4$ and $\text{Min}\Delta\phi(all\ jets, METnomu) < 2.0$
- ▶ Signal contribution also large for some regions of parameter space
- - We blind these regions and use as a basis for signal region optimisation
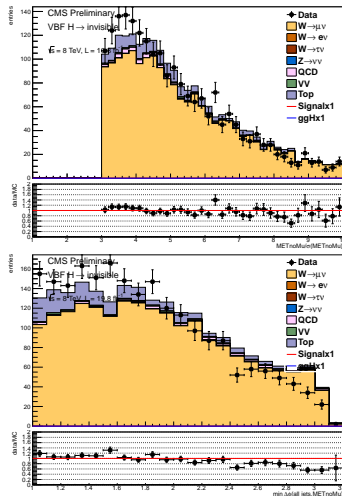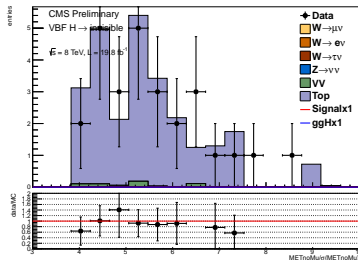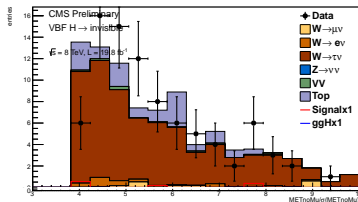
Signal region selection

- We optimise by choosing the cut values with the best 95% C.L. expected limit
  - Limit calculation details later
- We scanned through jet 2 p_T, dijet mass and $\text{Min}\Delta\phi(all\ jets, METnomu)$
- Best limit was found for:
  - No additional jet 2 p_T cut
  - $\frac{METnoMU}{\sigma_{METnoMU}} > 4$
  - $\text{Min}\Delta\phi(all\ jets, METnomu) > 2.5$
- This was used as our "signal region"
- Discrepancy outside signal region is from QCD

Top control region

- ▶ Top contribution to V+jets control regions is non-negligible
  - up to 16% in $W \to \tau\nu$
- ▶ Use method used for W backgrounds in prompt analysis
- ▶ Region: signal region with lepton veto replaced with requirement for 1 tight muon and 1 tight electron
  - Very few events in $e\mu$ region so also removed $\text{Min}\Delta\phi(all\ jets,\ METnomu)$ cut

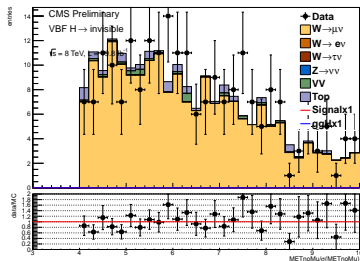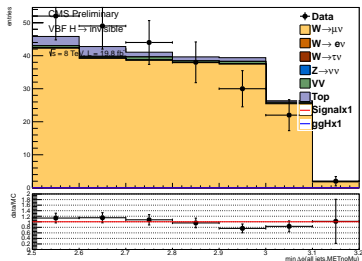|	
N_C^{data}	$28 \pm 5.3(\text{stat.})$
N_C^{bkg}	$0.6 \pm 0.2(\text{MC stat.})$
$N_S^{top\ MC}$	$9.6 \pm 1.8(\text{MC stat.})$
$N_C^{top\ MC}$	$42.6 \pm 5.2(\text{MC stat.})$
N_S^{top}	$6.1 \pm 1.2(\text{stat.}) \pm 1.4(\text{MC stat.})$

$$W \to e\nu$$

- ▶ Data-MC agreement good
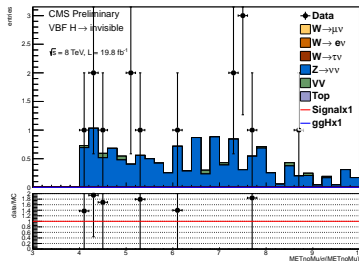- ▶ Same method used as for prompt analysis

Imperial College London
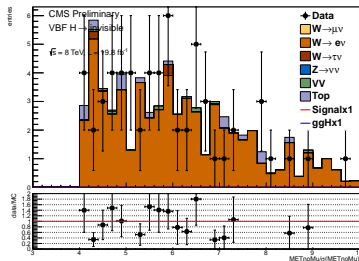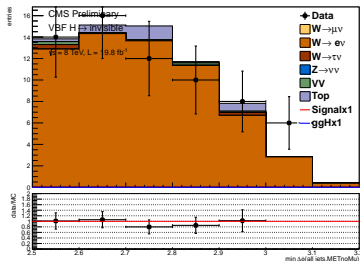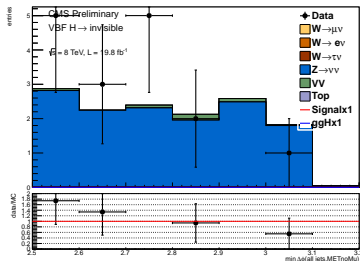
$W \to \mu\nu$

▶ Data-MC agreement good
▶ Same method used as for prompt analysis

$$Z \to \nu\nu$$

▶ Data-MC agreement good for limited statistics

▶ Same method used as for prompt analysis

$W \to \tau\nu$ control region

- ► For other W+jets backgrounds control region is:
  - signal region with lepton veto replaced with a requirement for a single lepton
- ► For $W \to \tau\nu$ there are not enough events in this region: 2 events for N_C^{Data}
  - In prompt analysis we removed the central jet veto (CJV)
- ► CJV no longer used, so we remove the $\text{Min}\Delta\phi(all\ jets, METnomu)$ cut
- ► This leads to QCD contamination so we require:
  - $\text{Min}\Delta\phi(leading\ 2\ jets, METnomu) > 1.0$
  - We also add an $m_T > 20$ GeV cut on the lepton-MET system to remove QCD contamination
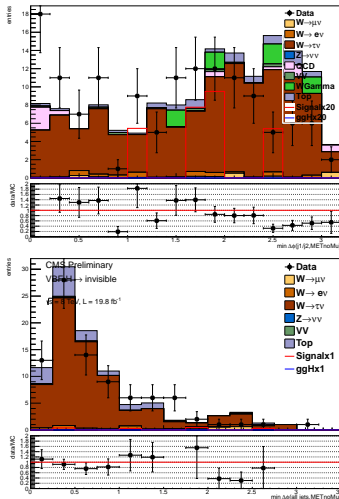
$W \to \tau\nu$ control region

- ▶ Estimate error from difference between control region and signal region cuts

- ▶ $W \to \mu\nu$ has enough events to see data driven weight variation with $\text{Min}\Delta\phi(\text{all jets}, \text{METnomu})$ cut

  - weight changes by 20% when loosening cut from 2.5 to 1.0.

  - We add a 20% systematic on the $W \to \tau\nu$ background

|	
N_C^{data}	$88 \pm 9.4(stat.)$
N_C^{bkg}	$15.2 \pm 4.8(MCstat.)$
N_S^{MC}	$176.1 \pm 10.5(MCstat.)$
N_C^{MC}	$133.9 \pm 8.0(MCstat.)$
N_S^{W}	$95.7 \pm 12.3(stat.) \pm 10.2(MCstat.)$
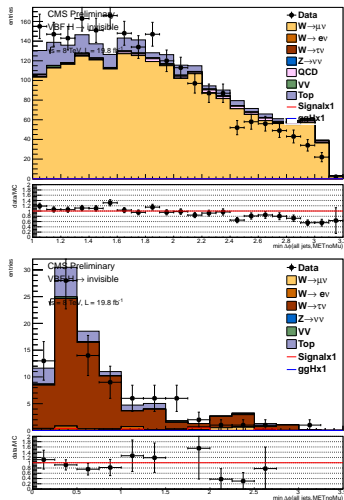
QCD background estimation: Shape region choice

- Try modelling QCD shape in preselection region
- QCD met must come from a jet, use region with low $\text{Min}\Delta\phi(all\ jets, METnomu)$
- Try inverted region with:
- $\text{Min}\Delta\phi(all\ jets, METnomu) < 1.0$
- $\text{Min}\Delta\phi(leading\ jets, METnomu) > 1.0$
- Has good shape agreement with enriched QCD MC
- Use shape taken from requiring:
- $\text{Min}\Delta\phi(all\ jets, METnomu) < 1.0$
- And replacing $\text{Min}\Delta\phi(all\ jets, METnomu)$ with $\text{Min}\Delta\phi(leading\ jets, METnomu)$
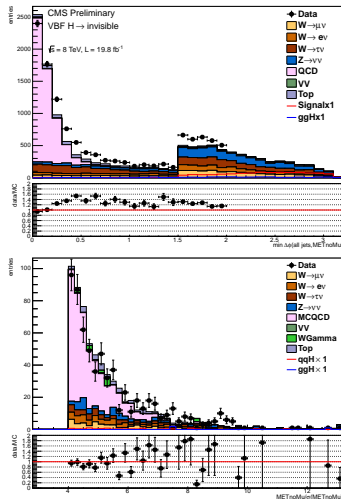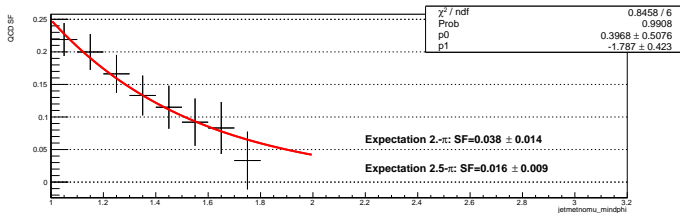
QCD background estimation: Scale factor

- ▶ Cannot use ABCD to normalise
- ▶ Scale factor shows strong dependence on cut variables
- ▶ Norm 2 and 3 have large signal contamination
- - Norm 3 also has low stats and odd because we forbid jets opposite the met in phi
- ▶ Fit scale factor variation in norm 1
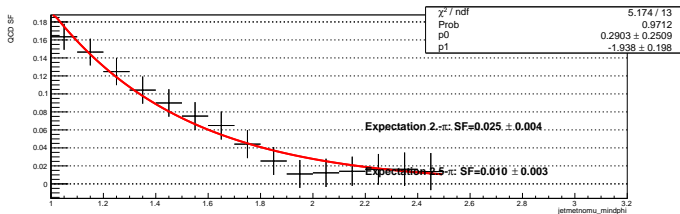- ▶ Check consistency in norm 2 and 3
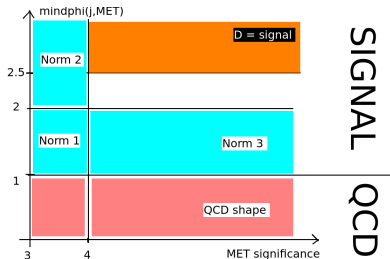
Scale factor variation

Norm 1

Norm 1+2

QCD background estimation: Result and systematics

Region	Factor	Extrapolation mindphi> 2.5	Extrapolation metsig> 4
Norm 1	0.22 ± 0.03	0.016 ± 0.009	0.04 ± 0.03
Norm 1+2	0.15 ± 0.02	0.010 ± 0.003	0.03 ± 0.02
Norm 1+3	0.41 ± 0.03	0.036 ± 0.062	1.10 ± 0.10
Norm 2	0.08 ± 0.02	-	0.05 ± 0.04
Norm 3	1.22 ± 0.15	0.60 ± 0.25	-

- ▶ Good agreement in all Min$\Delta\phi$(*all jets*, *METnomu*) extrapolations
- ▶ Norm 3 agreement in metsig is poor
- - As norm 3 has low statistics and is an odd region: drop
- ▶ Use envelope of norm 1 scale factors
- ▶ Final prediction: $N_S^{QCD} = 14 \pm 10$
- ▶ Expected limit 0.5% better with no QCD, 0.5% worse with double error

Results

Process	Number of events
$Z \to \nu\nu$	141.9 ± 36.4(stat.) ± 15.0(MC stat.)
$W \to e\nu$	59.7 ± 7.7(stat.) ± 5.2(MC stat.)
$W \to \mu\nu$	81.2 ± 5.6(stat.) ± 5.8(MC stat.)
$W \to \tau\nu$	95.7 ± 12.3(stat.) ± 10.2(MC stat.)
QCD	14 ± 10
Top	6.1 ± 1.2(stat.) ± 1.4(MC stat.)
VV	6.0 ± 0.6(MC stat.)
Total bkg.	405 ± 39.6(stat.) ± 19.8(MC stat.)
VBF signal	313.5 ± 9.4(MC stat.)
ggH signal	22.5 ± 6.0(MC stat.)
Total signal	336 ± 11.1(MC stat.)

Expected limits

- Used Higgs combine package with Asymptotic CLs method
- Performed a single bin counting experiment
- Analysis blind so have expected limits only
- 95% C.L. Median limit on B(H→inv.) for $m_H = 125$ GeV is: **31%**
  - 1σ band is 23-42%
  - 2σ band is 17-57%
- Prompt analysis expected limit was 49%
- We intend to run other mass points:
- 110, 150, 200, 300 and 400 GeV

Uncertainties by decreasing impact
Control region statistics
$Z \to \nu\nu\text{-}Z/\gamma^* \to \mu\mu$ extrapolation
nb above to be revisited
JES
$W \to \tau\nu$ extrapolation
MC statistics
QCD systematics
lepton ID efficiency
JER
UES
luminosity
PU weighting
theory uncertainties

Uncertainty impact table - impacts larger than 0%

Median expected limit with:		
All Nuisances: 30.8%	No Nuisances: 12.2%	
Nuisance	Removal impact/%	Addition impact/%
CMS_VBFHinv_zvv_stat:	-12.7%	64.6%
CMS_VBFHinv_zvv_extrapfacunc:	-7.6%	47.4%
CMS_scale_j:	-6.3%	19.2%
CMS_VBFHinv_tau_extrapfacunc:	-4.7%	24.9%
CMS_VBFHinv_zvv_norm:	-1.9%	18.4%
CMS_VBFHinv_wtau_stat:	-1.9%	12.0%
CMS_VBFHinv_wtau_norm:	-1.3%	8.8%
CMS_VBFHinv_wel_stat:	-1.3%	4.8%
CMS_VBFHinv_qcd_norm:	-1.3%	4.0%
CMS_VBFHinv_tau_eff:	-0.6%	5.6%
CMS_VBFHinv_wmu_norm:	-0.6%	3.2%
CMS_eff_m:	-0.6%	2.4%
CMS_VBFHinv_wmu_stat:	-0.6%	2.4%
CMS_VBFHinv_wel_norm:	-0.6%	2.4%
CMS_res_j:	-0.6%	0.8%
CMS_scale_met:	0.0%	0.8%

BDT Study

- ▶ Had a quick look at MVA analysis
- ▶ Started from cut based signal region
- - Only region with negligible QCD
- ▶ Best expected limit obtained 30%
- - Does not take into account any increased systematic
- - Therefore unlikely to be worthwhile
- ▶ New variables could make MVA worthwile
- ▶ Ability to model QCD would enable looser starting selection which may make MVA worthwhile
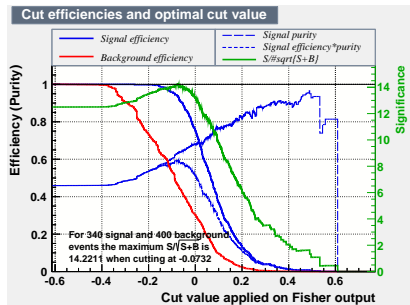
Summary

Parked analysis

- ▶ New cut based analysis found with negligible QCD and higher signal efficiency
  - Can then accept remaining QCD estimate with large error
  - Made possible by looser parked trigger thresholds
- ▶ MVA investigated
  - Not sufficient benefit without new variables or looser preselection
- ▶ Full cut based analysis presented
  - Expected limit 31%
  - Improved from 49% for prompt analysis

Run 2 prospects

- ▶ Prescaled looser control triggers planned to enable better QCD control region
- ▶ Will reinvistigate shape analysis and MVA

Backup

VBF enriched QCD MC

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:
 ▶ $m_{jj} > 700 \ GeV$
  ▶ $\Delta\eta > 3.2$

QCD Inc 80–600 GeV
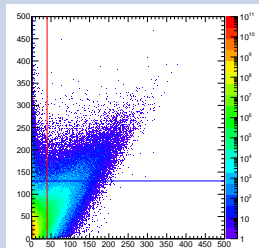
There 2 distict populations of events: real and fake met.

| 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

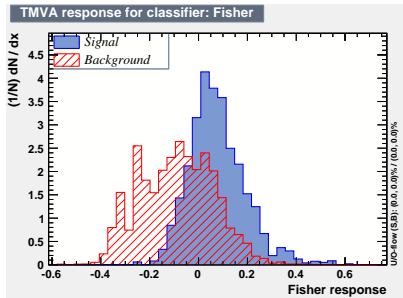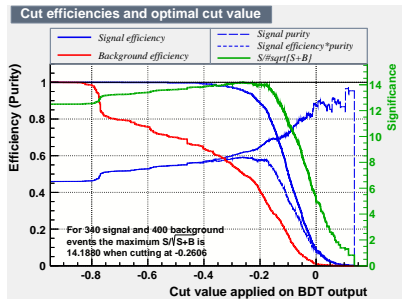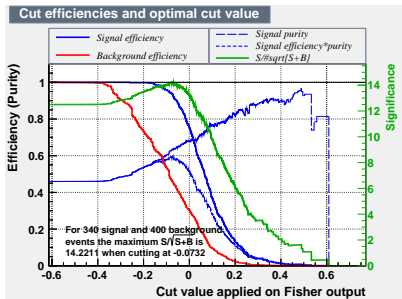## QCD Data-Driven: j1j3, j2j3 shape
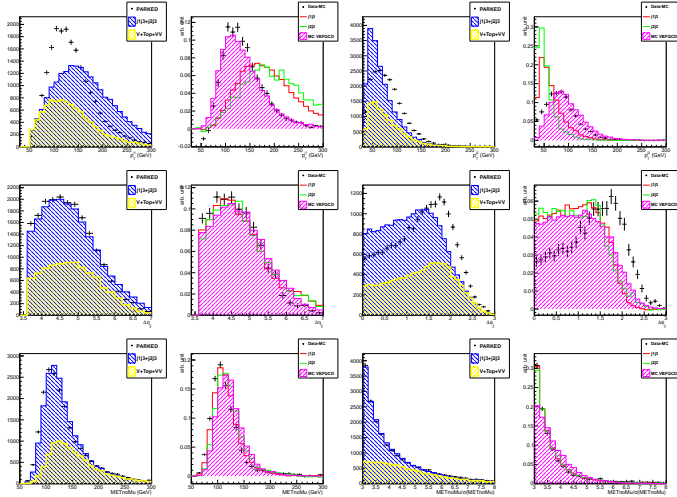
MVA Details
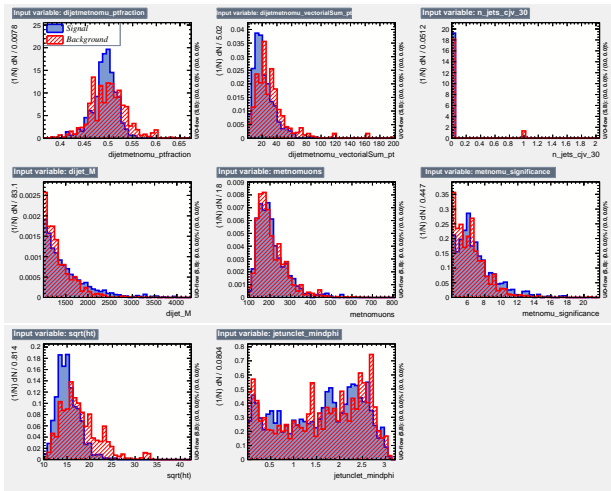
MVA Details
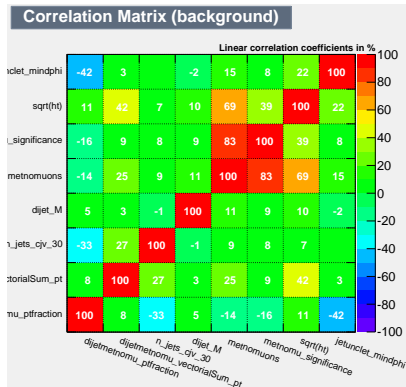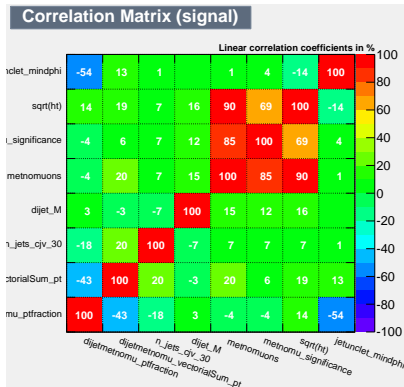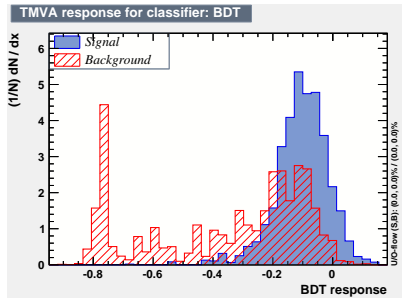
MVA Details

MVA Details

No systematics limits

- Prompt limit with no systematics 16.6%