Introduction for Long Exercise on Higgs Invisible

A.-M. Magnan, IC London
P. Meridiani, Roma

20/01/2015, CMSDAS@Bari

References to analyses done/ongoing with run I data

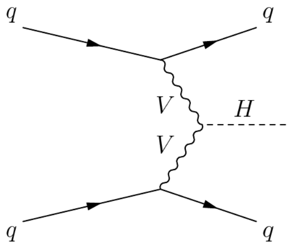
- Published analysis: `http://cms.cern.ch/iCMS/analysisadmin/cadilines?id=1355&ancode=HIG-13-030&line=HIG-13-030`
- Pre-approved with parked data:`http://cms.cern.ch/iCMS/analysisadmin/cadilines?id=1355&ancode=HIG-14-038&line=HIG-14-038`
- and analysis notes connected to these CADI entries.
- In this exercise: will concentrate on the VBF production only.

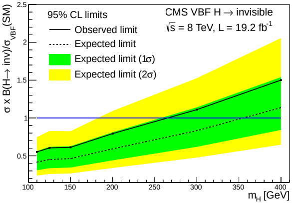
Many thanks to ...

- Chayanit Asawatangtrakuldee: the main analyser for published analysis.
- Patrick Dunne: the main analyser for parked data analysis.
- Renjie Wang for sharing his code to convert the limit to Dark Matter models.
- ... and the rest of the Higgs invisible team.

Higgs to invisible production in Vector Boson Fusion mode

- Higher cross-section than VH associated production mechanisms.
- Final state: 2 jets + missing energy (MET).
- Main background: QCD !! But, fortunately:
- no lepton,
- very characteristic kinematics for the dijet pair: forward-backward with high dijet mass,
- no hard hadronic activity expected in the rapidity gap between the two jets,
- large and significant MET.
- From measured yields: extract 95% CL limits on the branching ratio to invisible.



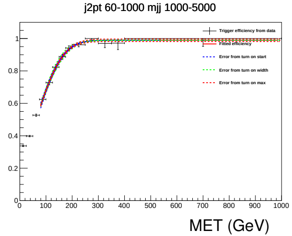


Triggers

Run I

- L1 seed: ETM40.

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



Proposed for Run II

- L1 seed: ETM70.
- L1 seed for control trigger: ETM50.

Trigger	MET cut	dijet pT cut	dijet mass cut	eff	rate
Baseline	PFMET170	-	-	9%	-
VBF	PFMETNoMu140	DiPFJetVBF40	MJJ600	10.5%	4.7 Hz
Control	PFMETNoMu80	DiPFJetVBF40	MJJ600	10.5%	0.5 Hz

Strategy (1/3)

Fiducial and trigger-driven selections for Run I triggers

- $|\eta^{j1,j2}| < 4.7$, $\eta^{j1} \times \eta^{j2} < 0$, $\Delta\eta_{jj} > 3.6$
- $p_T^{j1} > 50$ GeV, $p_T^{j2} > 40$ GeV, $M_{jj} > 1000$ GeV.
- PFMETNoMu$>$ 90 GeV.

Fiducial and trigger-driven selections for Run II triggers

- $|\eta^{j1,j2}| < 4.7$, $\eta^{j1} \times \eta^{j2} < 0$, $\Delta\eta_{jj} > 3.6$

Guesses for Baseline

- $p_T^{j1} > 30$ GeV, $p_T^{j2} > 30$ GeV, $M_{jj} > 400$ GeV.
- PFMETNoMu$>$ 250 GeV.

Guesses for VBF+Control

- $p_T^{j1} > 50$ GeV, $p_T^{j2} > 50$ GeV, $M_{jj} > 900$ GeV.
- PFMETNoMu$>$ 200 GeV.

After trigger sel: what is our sample composition ?

- 99.999% QCD multijets: mainly fake MET.
- W+jets, top, VV: real MET from leptonic decays of Ws, lepton(s) not reconstructed (e.g. too soft, outside e/μ fiducial regions).
- W+jets, W$\rightarrow$ $\tau_{had} + \nu$: hadronic tau jet either one of the tagging jet or extra jet.
- Z+jets, Z$\rightarrow$ $\nu\nu$: for this, Z$\rightarrow$ $\mu\mu$ is a great handle. Muons: little calorimetric energy, p_T can be added vectorially to MET to mimic invisible decay.
- EWK W/Z+2j production: same VBF jet topology.

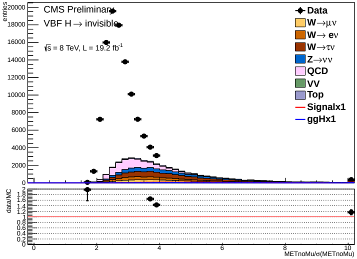
The METnoMuons subtlety

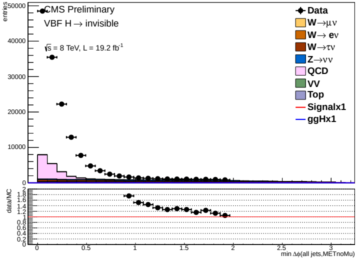
- Will use Z$\rightarrow$ $\mu\mu$ data to estimate Z$\rightarrow$ $\nu\nu$.
- Need to have stat after trigger: ignore muons when calculating MET.
- Note: L1MET is also only made with calorimeter towers.
- Baseline trigger for RunII is not so good for us: will cut a lot of Z$\rightarrow$ $\mu\mu$ statistics.

Strategy (2/3)

GET RID OF QCD MULTIJET BACKGROUND

- MET significance > 4 (calculated without tight muons).
- MET isolated from hard jets in ϕ: min$\Delta\phi$(all jets $p_T > 30$ GeV, MET)> 2.0 (again ignoring tight muons in the MET).





Strategy (3/3)

Optimise selection on 95% CL limit

- Blind the analysis.
- Signal region: apply electron and muon vetos (loose ID, medium isolation).
- Control regions (see next slides): extract SF to apply on MC with same selection as signal except lepton requirements.
- Optimise cut-based selection in signal region using rescaled MC: find best cut values for MET significance, min$\Delta\phi$(all jets $p_T > 30$ GeV, MET), p_T^{jets}, M_{jj}.
- Final figure of merit is the 95% CL limit.
- Neglect QCD for the optimisation.

jetmetmindphi	cut	2.0	2.2	2.3	2.4
	limit	51.4	37.1	41.6	42.3
mjj	cut	1000	1100	1200	1300
	limit	41.6	38.8	37.8	40.0
j2pt	cut	40	45	50	55
	limit	37.8	37.0	37.4	38.8
metsig	cut	4.0	4.2	4.5	
	limit	37.0	37.6	39.1	

Extraction of limits

- Simple cut& count approach: after killing QCD, other backgrounds are too signal like $\Rightarrow$ no improvement from using multivariate techniques.
- Background estimates: as much as possible rely on data.
- Use a control region per background:
  - W+jets: exactly one lepton (e,μ,τ).
  - Z+jets: exactly two muons.
  - top: one electron + one muon.
- Normalise MC to data in control regions.
- Check data-MC agreement in shapes in control regions.
- Extract yields in signal region from normalised MC.
- Based on agreement in control regions: unblind.

Normalisation to control regions

Generic equation

$$N_S^A = N_S^{A\,MC} \frac{N_C^{Data} - N_C^{bkg}}{N_C^{A\,MC}}, \qquad (1)$$

where

- N_S^A: number of events from process A in the signal region.
- $N_{S/C}^{A\,MC}$: number of events predicted by process A Monte Carlo in the signal/control region.
- N_C^{Data}: number of data events in the control region.
- N_C^{Bkg}: number of events from other background processes in the control region.
- A = top, W$\rightarrow$e, W$\rightarrow \mu$,W$\rightarrow \tau_{had}$

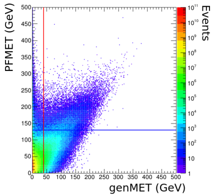
Complication for $\mu\mu$ region

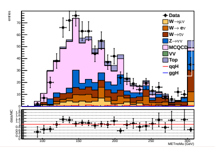
$$N_S^{Z\rightarrow\nu\nu} = \left(N_C^{Data} - N_C^{bkg}\right) \cdot \frac{\sigma\left(Z \rightarrow \nu\nu\right)}{\sigma\left(Z \rightarrow \mu\mu\right)} \cdot \frac{\epsilon_S^{ZMC}}{\epsilon_C^{ZMC}}, \qquad (2)$$

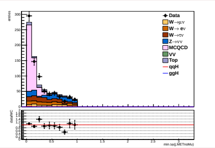
A word on QCD multijet

MC - not enough stat

- VBF-enriched: genMET>40 GeV $\Rightarrow$ missing fake MET contribution.
- Always 3rd jet $p_T > 30$ GeV aligned with MET: jet with high neutrino component.
- Good agreement with data in shape for this population.

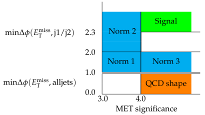


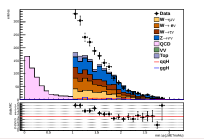


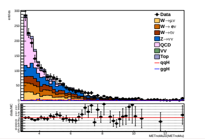


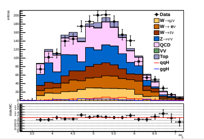
Data-driven - tricky

- Use events with non-isolated MET to predict isolated component.
- ABCD doesn't work: signal contamination too large in 3/4 regions ! Use extrapolations.
- In the end: small number with large uncertainty $\Rightarrow$ small impact on limit.



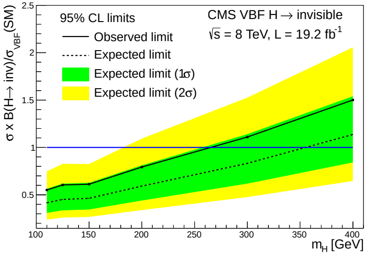


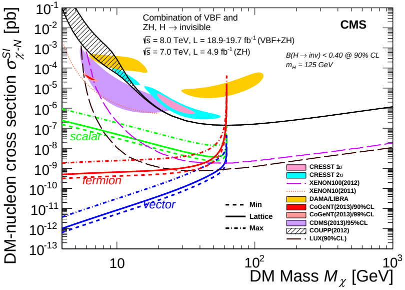




Results

Background	$N_{est} \pm (stat) \pm (syst)$
$Z \to \nu\nu$	157.3 $\pm$ 37.6 $\pm$ 38.3
$W \to \mu\nu$	101.8 $\pm$ 6.1 $\pm$ 11.9
$W \to e\nu$	57.4 $\pm$ 7.3 $\pm$ 7.0
$W \to \tau\nu$	98.0 $\pm$ 13.2 $\pm$ 25.4
top	4.4 $\pm$ 1.0 $\pm$ 1.4
VV	3.8 $\pm$ 0.0 $\pm$ 0.7
QCD multijet	17 $\pm$ 0 $\pm$ 14
Total Background	439.7 $\pm$ 41.0 $\pm$ 55.8
Signal(VBF)	273.4 $\pm$ 0.0 $\pm$ 31.2
Signal(ggH)	22.6 $\pm$ 0.0 $\pm$ 15.6
Data observed	508





Goals of the long exercise

- Analyse 13 TeV MC starting from miniAOD samples.
- Re-optimise selection with new trigger MET threshold.
- Extract expected yields for W(e,μ,τ) and Z.
- Extract 95% confidence limits on branching ratio Higgs$\rightarrow$invisible for $m_H = 125$ GeV, as function of Run II luminosity.
- Interpret limit in Dark Matter models.

Proposed list of tasks

- MiniAOD -> LightTree step.
- Trigger: VBF trigger not available yet, apply threshold by hand on L1MET (known to be dominant component of the turn-on).
- Selection: apply trigger-driven and QCD-killer selections.
- Use Run I scale factors control->signal regions for W,Z,top.
- Use TMVA cut-based optimisation on 5 variables: MET significance, min$\Delta\phi$(all jets $p_T > 30$ GeV, MET), p_T^{jets}, M_{jj}using W/Z vs signal: choose point with highest purity (should translate to best 95% CL limit).
- Study dependence to pile-up on signal sample: efficiency vs number of vertices to check impact of PU jetID in 25ns scenario.
- Plots of main kinematic variables, in control and signal regions.
- Table of expected yields in control and signal region for 10 fb^{-1}.
- Expected 95% limit on BR using systematics from run I, for $m_H = 125$ GeV, extrapolated to different integrated luminosity.
- Conversion to DM models.