High mass H→WW→lνlν search

Yongke Zhao

on behalf of the ATLAS Collaboration

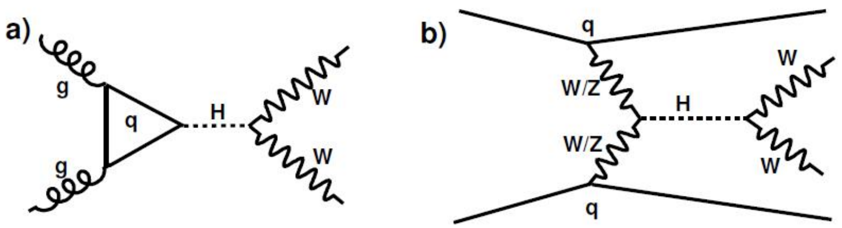
17/8/2016

Introduction

- Motivation: search for a potential extension of the Standard Model(SM) with an extended Higgs sector
- Signals(scenarios) in our study:
  - A heavy Higgs with a narrow(4 MeV) or large(5,10,15% of m_H) width
  - Mass range: 300 GeV - 3 TeV
- Higgs production modes in our study:

  a) ggF

  b) VBF



- The analysis channel
  - H→WW→lvlv μ μ , ee not considered for the moment
- We are using Run 2 data with integrated luminosity of 13.2 fb^{-1} at 13 TeV
- CONF Note: https://cds.cern.ch/record/2206243

Backgrounds

- Top, WW

  Normalization factors determined from final fit using signal regions(SRs) and control regions(CRs)

  Except for VBF WW 2J which was from MC prediction

- W+jets

  Derived from data using "fake-factor" method (from SM $H \rightarrow WW$ analysis)

  See Weimin talk for details

- Z+jets, non-WW diboson

  Small, predicted from simulation

- H125

  Small, predicted from simulation

  ggF component interference with WW considered

Transverse mass

- Transverse missing momentum:

  Negative vectorial sum of the transverse momenta

  - $\mathbf{E}_T^{miss}$: based on calorimeter objects
  - $\mathbf{p}_T^{miss}$: based on charged tracks

- Transverse invariant mass:

$$M_T = \sqrt{(E_T^{ll} + E_T^{miss})^2 - |\,\mathbf{p}_T^{ll} + \mathbf{E}_T^{miss}\,|^2}$$

→ Discriminating variable

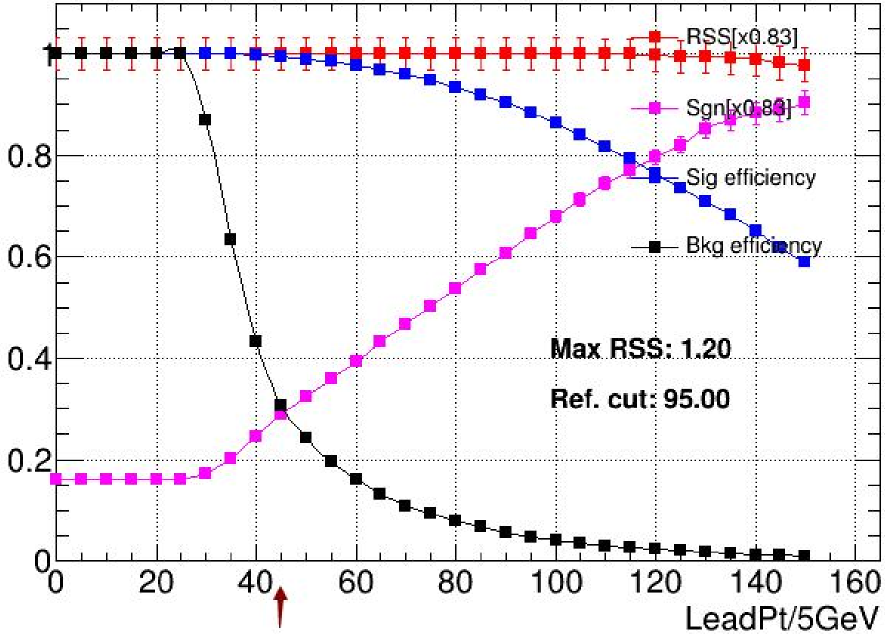
$$where \quad E_T^{ll} = \sqrt{|\,\mathbf{p}_T^{ll}\,| + m_{ll}^2}$$

Event selection optimization

- Developed a simple and general optimization procedure
  1. Select the most discriminating variables from the BDT training

     Remove duplicated variables that are highly correlated

  2. Choose cut values for each variable by maximizing the signal significance
- Scanning on leading lepton p_T



N-1 scanning: also other cuts applied

- Naming on this plot:

$$RSS = \sqrt{\sum_{i}^{N_{bins}} Sgn_i^2}$$

$$Sgn = \sqrt{2\left[(n_S + n_B)\ln\left(1 + \frac{n_S}{n_B}\right) - n_S\right]}$$

Event selection

- Event selection in signal regions(SRs) :
  - VBF 1J phase space: $N_{\text{jet}} = 1:\ |\eta_j| > 2.4 \text{ and } \min(|\Delta\eta_{j\ell}|) > 1.75$
  - VBF 2J phase space: $N_{\text{jet}} \geq 2:\ m_{jj} > 500 \text{ GeV and } |\Delta y_{jj}| > 4$

Preselection cuts: $p_{\mathrm{T}}^{\text{lead}} > 25$ GeV, $p_{\mathrm{T}}^{\text{sublead}} > 15$ GeV, 3rd lepton veto, $m_{\ell\ell} > 10$ GeV		
$\mathrm{SR}_{\mathrm{ggF}}$	$\mathrm{SR}_{\mathrm{VBF1J}}$	$\mathrm{SR}_{\mathrm{VBF2J}}$
$N_{b\text{-jet}} = 0$ $\lvert\Delta\eta_{\ell\ell}\rvert < 1.8$ $m_{\ell\ell} > 55$ GeV $p_{\mathrm{T}}^{\text{lead}} > 45$ GeV $p_{\mathrm{T}}^{\text{sublead}} > 30$ GeV $\max(m_{\mathrm{T}}^{W}) > 50$ GeV		
Inclusive in N_{jet} but excluding VBF1J and VBF2J phase space	$N_{\text{jet}} = 1$ $\lvert\eta_j\rvert > 2.4$ $\min(\lvert\Delta\eta_{j\ell}\rvert) > 1.75$	$N_{\text{jet}} \geq 2$ $m_{jj} > 500$ GeV $\lvert\Delta y_{jj}\rvert > 4$

$$m_{\mathrm{T}}^{W} = \sqrt{2 p_{\mathrm{T}}^{\ell} E_{\mathrm{T}}^{\text{miss}} \left(1 - \cos\left(\phi^{\ell} - \phi^{E_{\mathrm{T}}^{\text{miss}}}\right)\right)}$$

Event selection

- Event selection in control regions(CRs) :
  - VBF 1J phase space: $N_{\text{jet}} = 1: |\eta_j| > 2.4 \text{ and } \min(|\Delta\eta_{j\ell}|) > 1.75$
  - VBF 2J phase space: $N_{\text{jet}} \geq 2: m_{jj} > 500 \text{ GeV and } |\Delta y_{jj}| > 4$

WW CR$_{\text{ggF}}$	Top CR$_{\text{ggF}}$	WW CR$_{\text{VBF1J}}$	Top CR$_{\text{VBF}}$
$N_{b\text{-jet}} = 0$	$N_{b\text{-jet}} = 1$	$N_{b\text{-jet}} = 0$	$N_{b\text{-jet}} \geq 1$
$\|\Delta\eta_{\ell\ell}\| > 1.8$	$\|\Delta\eta_{\ell\ell}\| < 1.8$	($\|\Delta\eta_{\ell\ell}\| > 1.8$ or	–
$m_{\ell\ell} > 55$ GeV		$m_{\ell\ell} < 55$ GeV)	–
$p_{\text{T}}^{\text{lead}} > 45$ GeV		$p_{\text{T}}^{\text{lead}} > 25$ GeV	$p_{\text{T}}^{\text{lead}} > 25$ GeV
$p_{\text{T}}^{\text{sublead}} > 30$ GeV		$p_{\text{T}}^{\text{sublead}} > 25$ GeV	$p_{\text{T}}^{\text{sublead}} > 15$ GeV
$\max(m_{\text{T}}^{W}) > 50$ GeV		–	–
Excluding VBF VBF1J and VBF2J		VBF1J phase space	VBF1J or VBF2J phase space

Dominant systematics

Main experimental uncertainties

- Dominant Top uncertainties from jet energy scale and resolution: 9.8%, 12% in VBF1J, 2J SRs. Also 16%, 23% for WW uncertainties

Main theoretical uncertainties on backgrounds

- Dominant Top uncertainties from generator modelling: 17%, 48% in VBF 1J, 2J SRs. Also 35%, 48% for WW uncertainties
- Similar in CRs, so the extrapolation uncertainties from CRs to SRs remain small

Main theoretical uncertainties on signals

- Dominant scale uncertainties on category migration: 30% increased to 90% from 300 GeV to 3 TeV for VBF 1J, 25% increased to 40% from 300 GeV to 3 TeV

Results: event yields

- Normalization factors(NFs)

$$NF^{Top}_{ggF\,CR} = 0.96^{+0.09}_{-0.08} \qquad NF^{Top}_{VBF\,CR} = 0.96^{+0.12}_{-0.14}$$

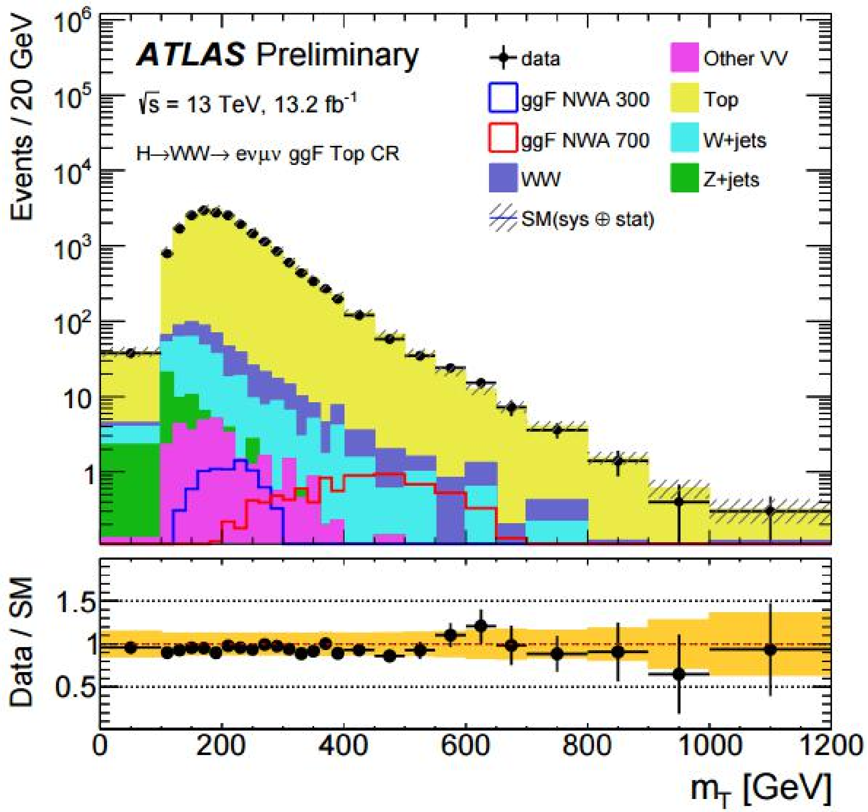
$$NF^{WW}_{ggF\,CR} = 1.3^{+0.1}_{-0.1} \qquad NF^{WW}_{VBF1J\,CR} = 1.2^{+0.5}_{-0.3}$$

- Event yields (statistical and systematic uncertainties combined)

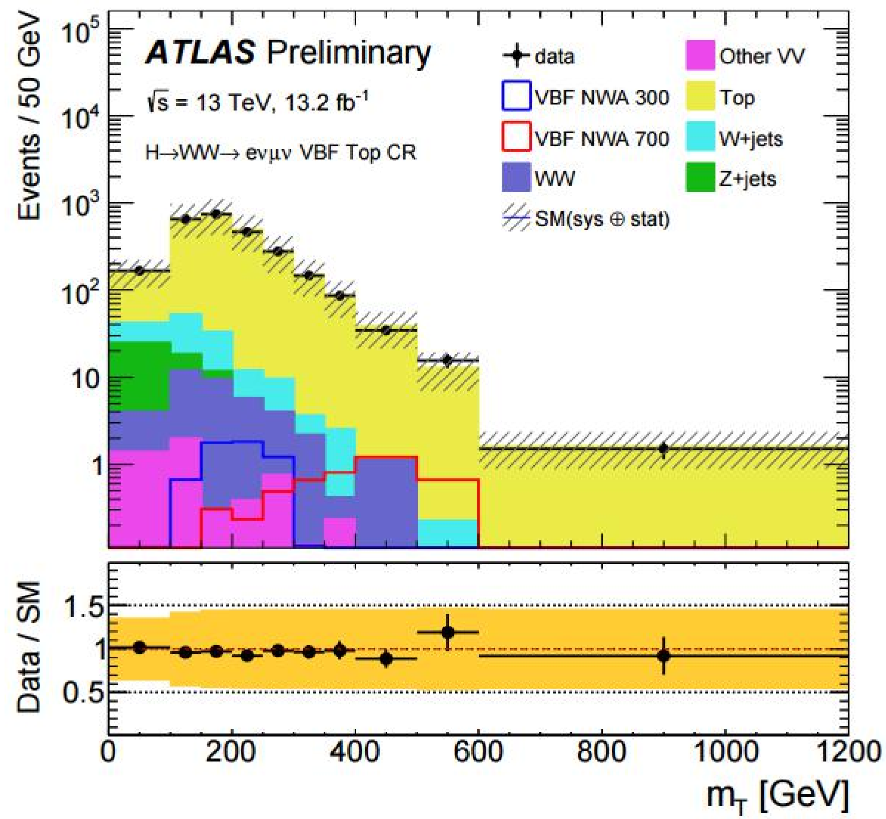
	SR_{ggF}	Top CR_{ggF}	*WW* CR_{ggF}
WW	5300 ± 400	430 ± 90	1430 ± 120
Top-quark	4200 ± 400	20560 ± 210	900 ± 100
Z/γ^*	557 ± 25	46 ± 12	10.7 ± 1.0
W+jets	450 ± 120	260 ± 80	105 ± 30
VV	323 ± 12	37 ± 4	88.5 ± 3.4
Backgrounds	10790 ± 110	21330 ± 180	2530 ± 40
Data	10718	21333	2589

	SR_{VBF1J}	SR_{VBF2J}	Top CR_{VBF}	*WW* CR_{VBF1J}
WW	197 ± 31	53 ± 15	37 ± 4	117 ± 21
Top-quark	141 ± 26	124 ± 19	2650 ± 80	65 ± 14
Z/γ^*	20 ± 7	12 ± 4	40 ± 17	27 ± 5
W+jets	22 ± 6	7.5 ± 2.2	95 ± 25	24 ± 6
VV	9.5 ± 1.0	5.7 ± 0.8	5.2 ± 2.2	11.0 ± 1.5
Backgrounds	389 ± 22	202 ± 14	2830 ± 70	247 ± 16
Data	384	203	2825	253

Results: plots in Top CRs

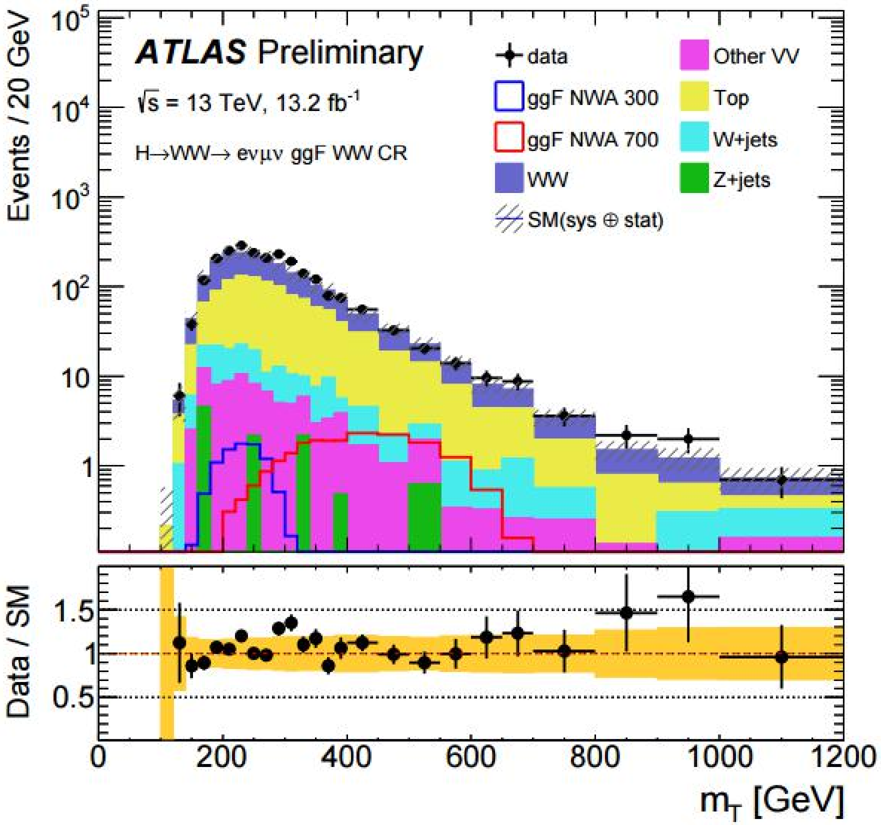


ggF Top CR

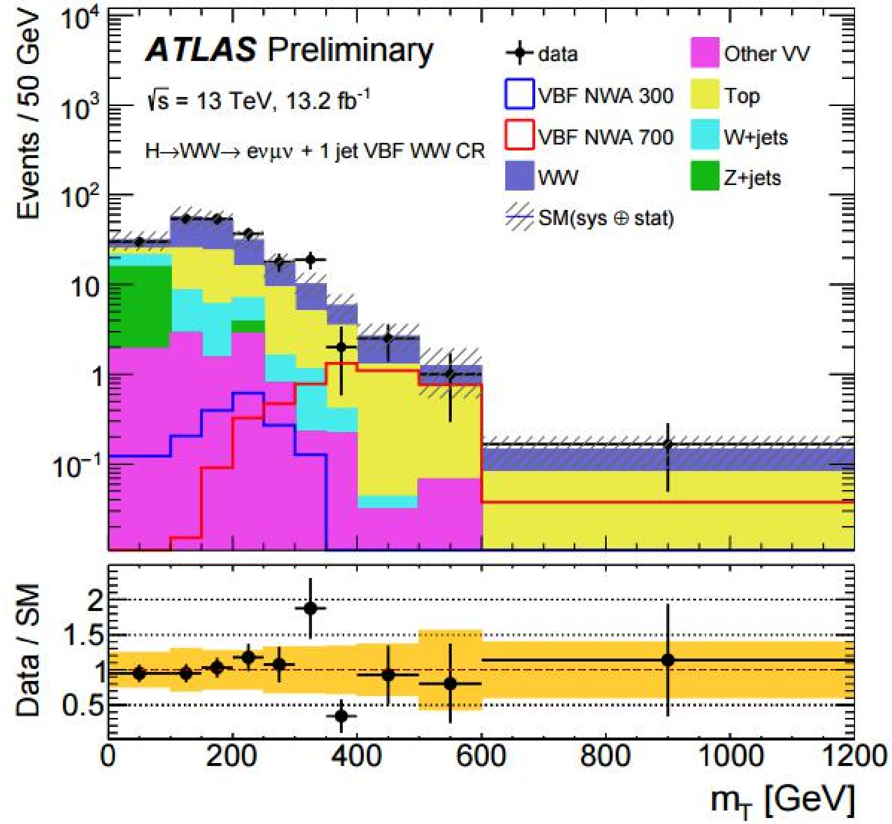


VBF Top CR

Results: plots in WW CRs



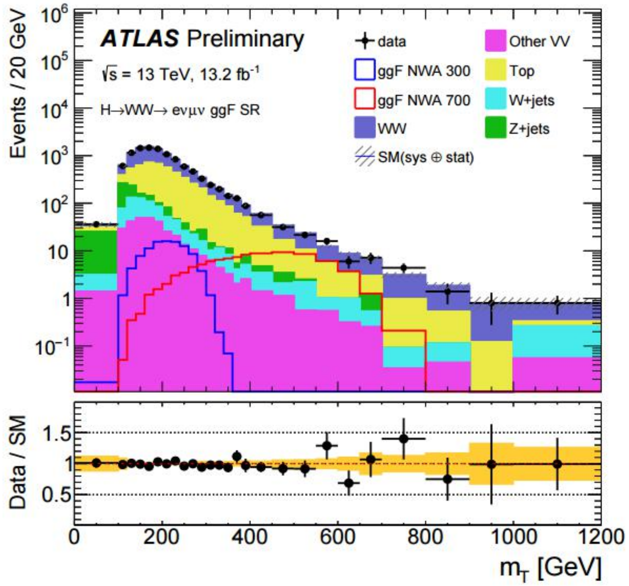
ggF WW CR

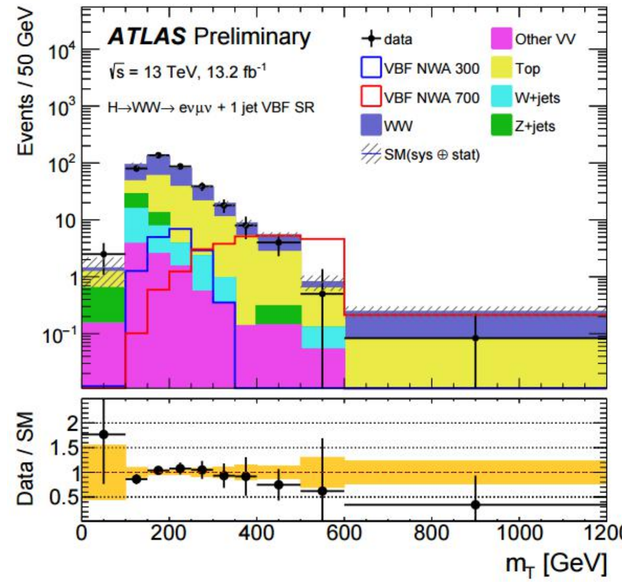


VBF WW 1J CR

Results: plots in SRs

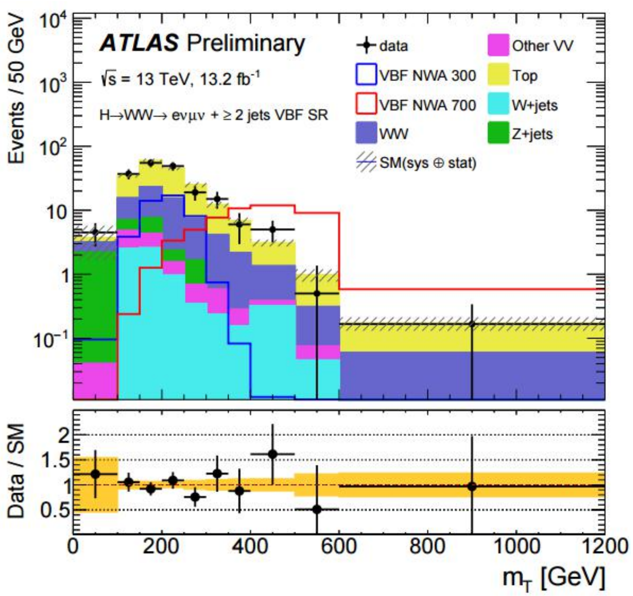
ggF SR





VBF 1J SR

VBF 2J SR



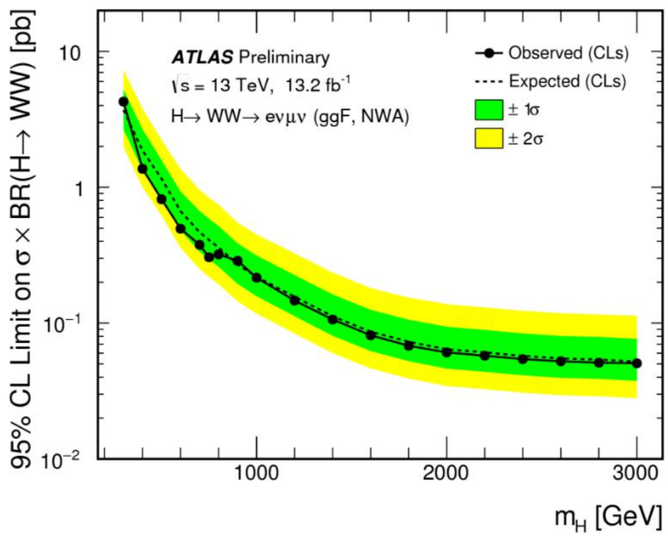
$$\text{Signal scaled to}: \sigma_H \times BR(H \rightarrow WW) = 1\,\text{pb}$$

No evidence for an excess around 750 GeV

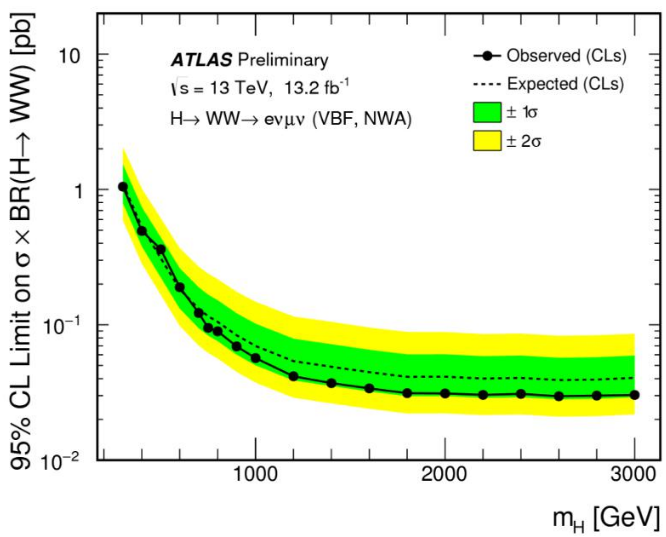
Results: upper limits

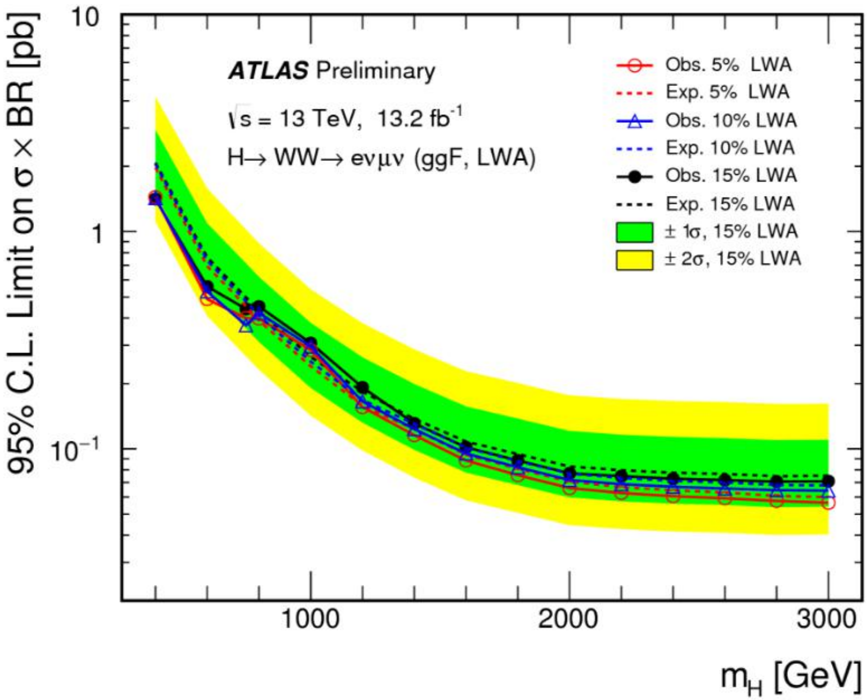
- 95% CL upper limit on $\sigma_H \times BR(H \to WW)$

NWA, ggF



NWA, VBF





LWA, ggF

NWA: SM-like Higgs with narrow width

LWA: wih large width

Conclusion

- A search for heavy Higgs boson performed in the H⟶WW⟶lvlv decay channels at 13 TeV with 13.2 fb^{-1} data(year 2015 + 2016 of Run 2)
- No significant excess found between 300 GeV and 3 TeV
- 95% CL upper limits on $\sigma_H \times BR(H \to WW)$ shown for two SM-like Higgs models with a narrow or large width
- Compared to Run 1, the limits extend well from 1.5 TeV to 3 TeV
- Next plan:

  Continue this search with more 2016 data

BACKUP

BACKUP - object reconstruction

- Primary vertex (PV) = 1
  - ≥ 2 associated tracks with p_T > 400 MeV
  - if > 1 vertex meet the requirement above, the largest track $\sum p_T^2$ chosen as PV
- Electron
  - MediumLH for pT > 25 GeV or TightLH for 15 < pT < 25GeV
  - $|\eta| < 2.47$, except for $1.37 < |\eta| < 1.52$
- Muon
  - Medium for pT > 25 GeV or Tight for 15 < pT < 25GeV
  - $|\eta| < 2.5$
- Jet
  - p_T > 30 GeV, $|\eta| < 4.5$
  - JVT > 0.59 for p_T < 60 GeV
  - Overlap removal with electrons and muons
- B-tagged jet
  - Identified using the MV2c10 b-tagging algorithm, with an efficiency of 85%
  - p_T > 20 GeV, $|\eta| < 2.5$

BACKUP - object reconstruction

- MET: missing transverse momentum
  - E_T^{miss}: calorimeter-based

    Negative vectorial sum of the transverse momenta of all calibrated selected objects

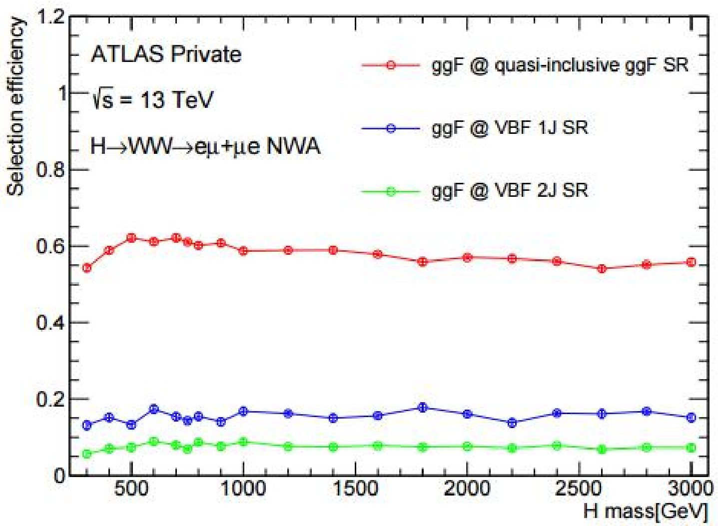
    Tracks compatible with the primary vertex but not matched to the objects also included
  - p_T^{miss}: track-based

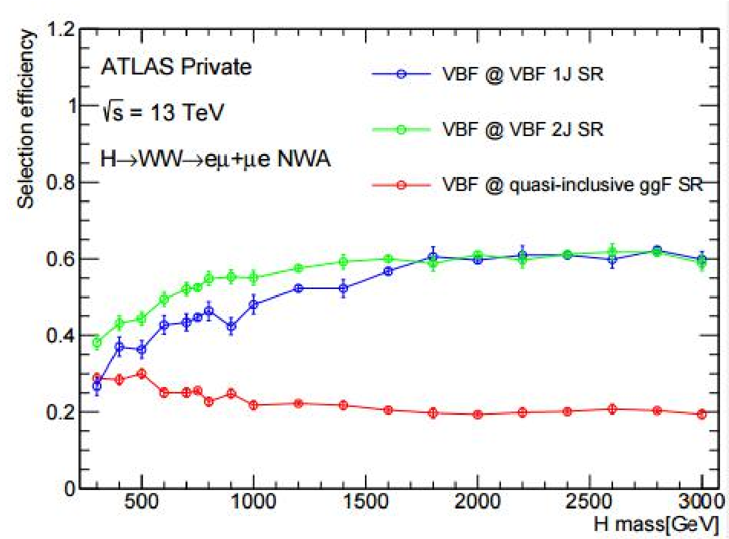
    negative sum of the momenta of ID tracks, satisfying:
    - $d_0 < 1.5$mm, $d_0/\sigma(d_0) < 3$
    - $|\eta| < 2.5$, $p_T > 500$ MeV

    Calorimeter electron p_T used instead of track p_T

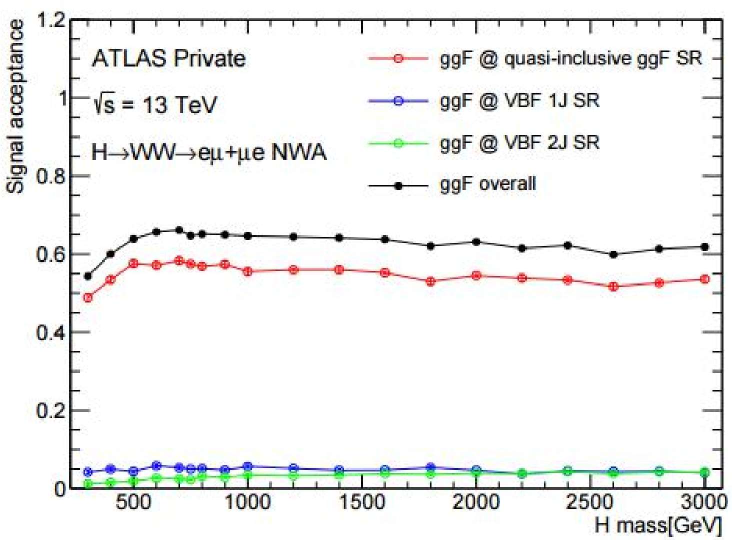
BACKUP - acceptance*efficiency

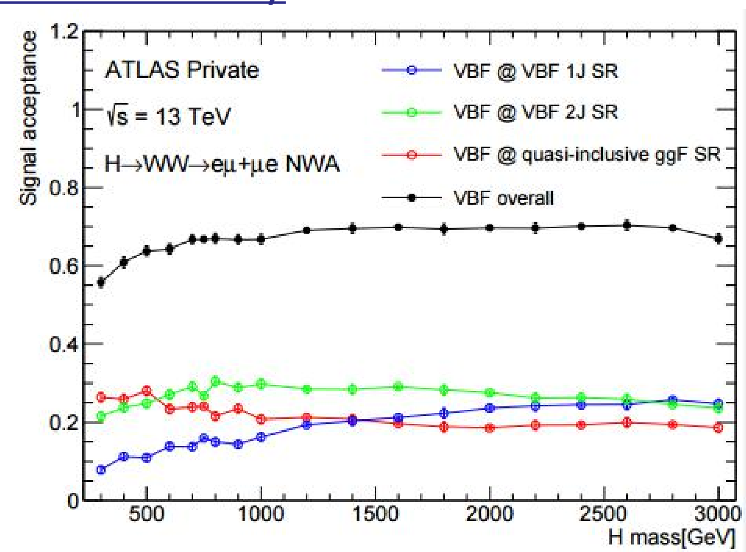
- Selection efficiency (not including pre-selection cuts)





- Acceptane*efficiency (namely the <u>overall efficiency</u> after all selections in SR)





BACKUP - fake-factor method

6.3 W+jets background

The background contribution is estimated using the fake-factor based data-driven method developed for the SM $H \rightarrow WW$ analysis [31] . The relevant information used for this analysis is shown here. The W+ jets background contribution is estimated using a sample of events satisfying all selection criteria but in which one of the two lepton candidates satisfies the identification and isolation criteria used to define the signal samples (these lepton candidates are denoted as "fully identified"), and the other lepton fails to meet these criteria and satisfies a less restrictive selection denoted as "anti-identified" (anti-id) lepton. From this data sample the non-W+ jets contribution based on MC predictions is subtracted. The purity of the samples is 55%, 64% and 38% for the quasi-inclusive ggF, $N_{\text{jet}} = 1$ and $N_{\text{jet}} \geq 2$ VBF categories, respectively.

The W+ jets contamination in the signal region is determined by scaling the number of events in the selected data sample by an extrapolation factor (fake-factor), which is measured in a data sample of di-jets events. The fake-factor is the ratio of the number of fully identified leptons to the number of anti-identified leptons, measured in bins of anti-identified lepton p_{T} and η. The systematic errors associated with the fake factor evaluation are described in Section 7.3.

Systematic uncertainties

- Total luminosity uncertainty: 2.9% (data15: 2.1%, data16: 3.7%)
- Main experimental uncertainties (in %):

	Top-quark			WW		
Source	SR_{ggF}	SR_{VB1J}	SR_{VBF2J}	SR_{ggF}	SR_{VBF1J}	SR_{VBF2J}
Jet	4.6	9.8	12	1.3	16	23
b-tag	17	6.2	13	1.7	0.99	3.3
MET	0.09	0.03	0.37	0.22	0.18	0.46
JVT	2.1	0.73	2.2	1.0	0.45	1.8
MC Stat.	0.42	2.4	2.5	0.58	2.7	4.8

- Top background theoretical uncertainties
  - Dominating uncertainties from ME: 17%, 48% for VBF 1J, 2J SRs
  - Similar in CRs, so the extrapolation uncertainties from CRs to SRs remain small

Error source
ME
PS
Radiation (radHi)
Radiation (radLo)
$Wt - t\bar{t}$ interference
Relative variation of σ_{st} ($\pm$20%)
PDF (down/up)

Systematic uncertainties

- WW background theoretical uncertainties
  - Dominating uncertainties from ME+PS: 35%, 48% for VBF 1J, 2J SRs
  - Extrapolation uncertainties still small

Error source
ME+PS
Renormalisation scale (0.5)
Renormalisation scale (2)
Factorisation scale (0.5)
Factorisation scale (2)
Qsf scale (0.5)
Qsf scale (2)
CKKW matching (down)
CKKW matching (up)
PDF (down/up)

  - NLO EW correction in the dominant qq→WW process was considered as normalization and shape uncertainty
  - Addtional k-factor of 1.7 as higher order correction in gg→(h*)→WW process with 60% uncertainty

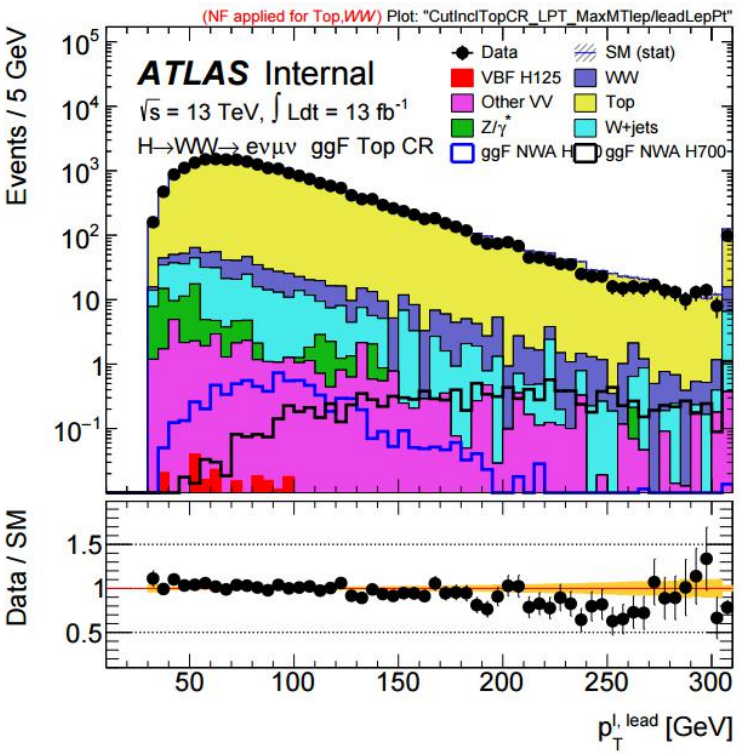
- For both Top and WW, shape uncertainties also considered

Systematic uncertainties

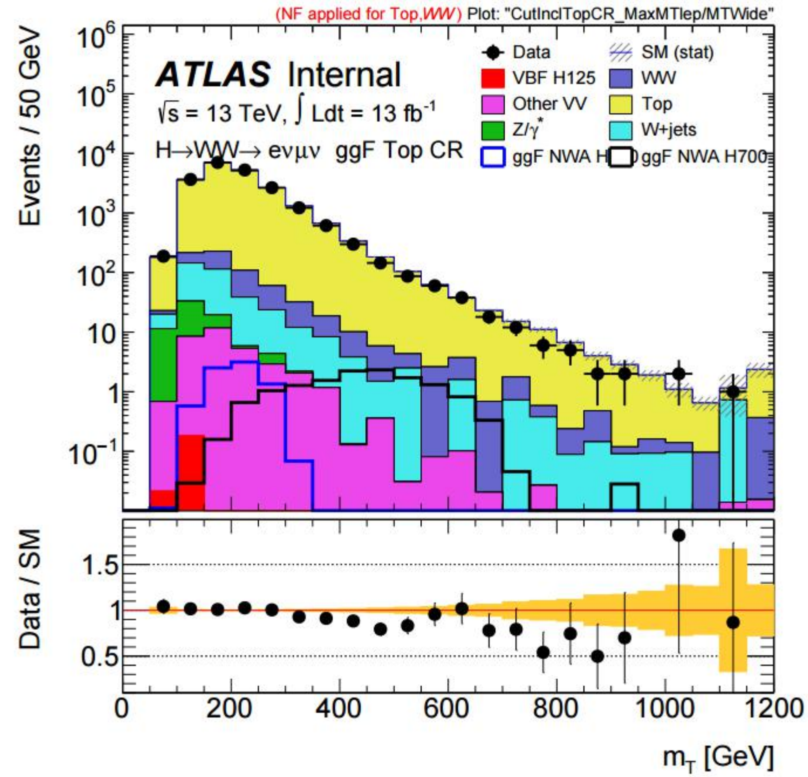
- Dominating W+jets fake-factor estimation uncertainties
  - EW contribution: 15% for both e and μ in VBF 1J and 2J SR
  - Sample composition uncertainties, in ggF SR, VBF 1J and 2J SR respectively:
    - e: 26%, 21% and 21%
    - μ: 12%, 16% and 16%
  - Statistical uncertainty: < 1%
- Dominating signal theory uncertainties
  - Scale uncertainties on category migration
    - ggF: 10% over full mass range
    - VBF 1J: 30% increased to 90% from 300 GeV to 3 TeV
    - VBF 2J: 25% increased to 40% from 300 GeV to 3 TeV
  - Scale uncertainties on acceptance: relatively small
  - PS, underlying event and PDF uncertainties: relatively small
- For signal, we also have a correction on VBF due to PowHeg mismodelling

Top leading lepton p_T shape correction

- Mis-modelling found in ggF inclusive Top CR



leading lepton p_T

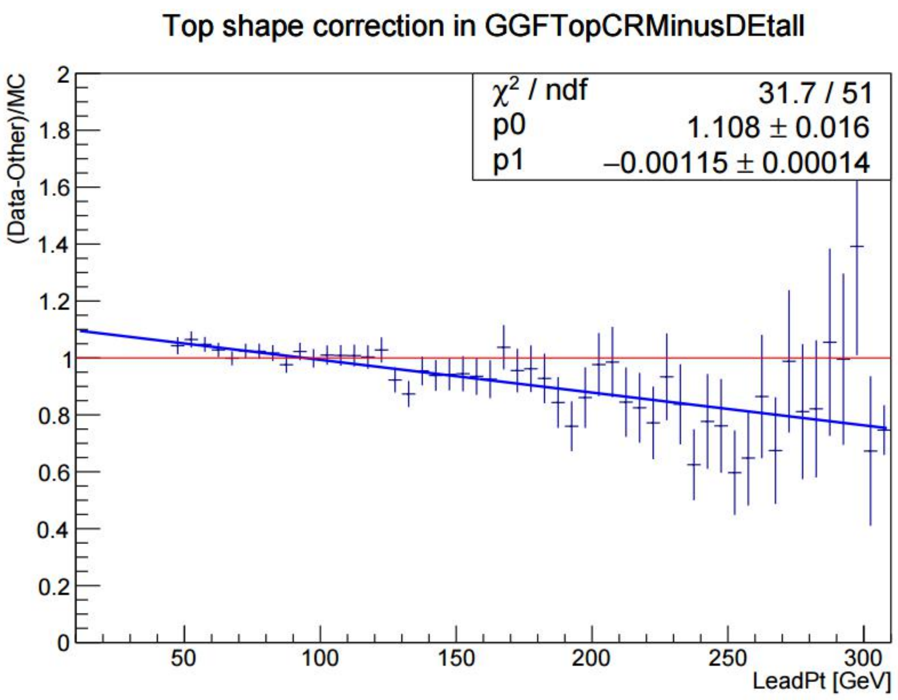


m_T

- This mis-modelling was less pronounced with DS1 data samples, and consistent with the NNLO QCD correction.
- No such mis-modelling observed in the VBF categories

Top leading lepton p_T shape correction

- The correction used a reweighting with linear fit function in ggF Top CR



- No discrepency between em and me
- DEtall cut has no effect on the correction
- Applied to ggF SR, WW and Top CRs
- Shape uncertainties also considered

Fitting

- Likelihood function defined using MT distributions, as a product of Poisson functions over following 7 regions:
  - ggF SR: 26 bins
  - VBF 1J, 2J SRs: 10 bins
  - ggF Top, WW and VBF Top, WW 1J CRs: 1 bin
- Profiled likelihood method was used
- For the observed limits for ggF(VBF), VBF(ggF) cross section treated as a nuisance parameter, and profiled using a flat prior

BACKUP - pull plots