$VH \rightarrow b\bar{b}$ Searches At ATLAS And The Estimation Of Theoretical Systematic Uncertainties

IOP 2013

Ben Harry Smart
The University of Edinburgh

$H \to b\bar{b}$

$H \rightarrow b\bar{b}$

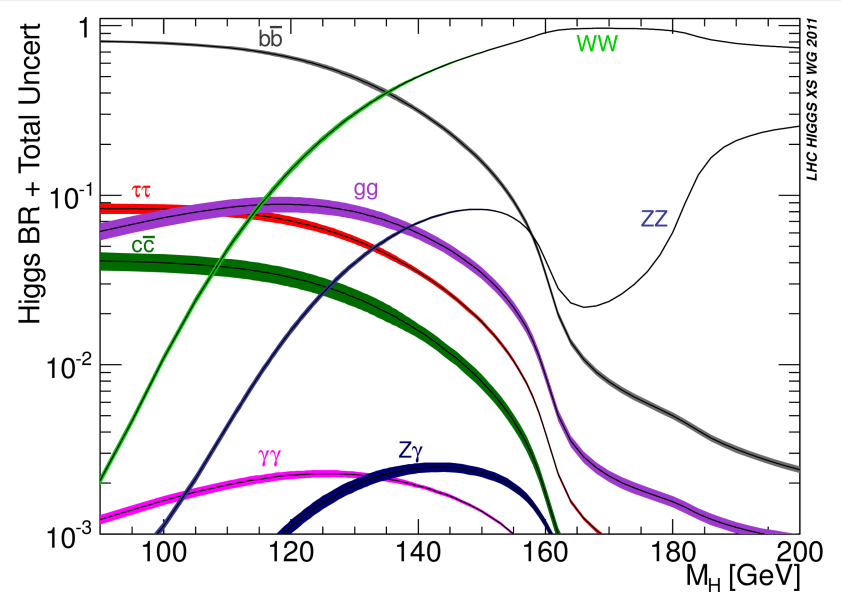
- Why are we searching for this?

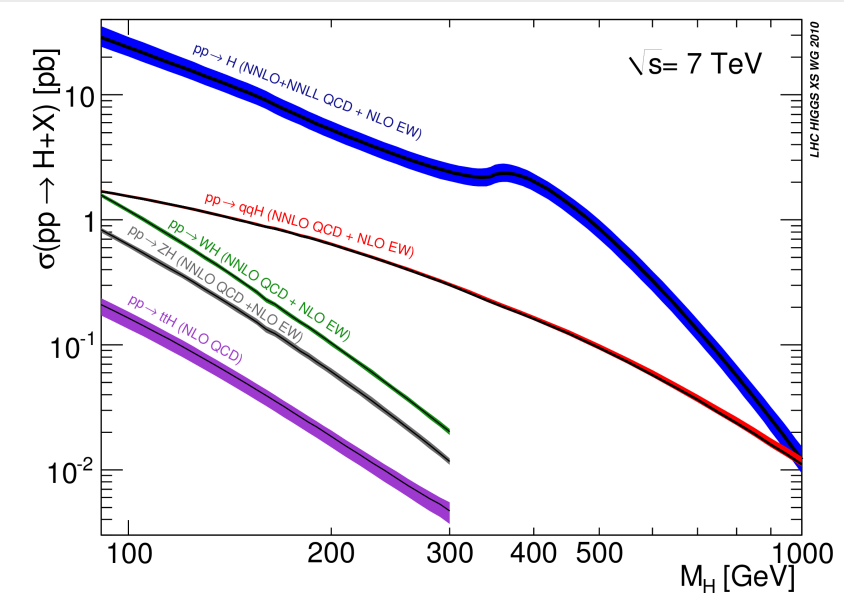
$H \rightarrow b\bar{b}$

- Why are we searching for this?
  - We believe that a Standard Model Higgs boson should decay into two b quarks.
  - Yet to observe a Higgs boson decaying to fermions.
  - Whether we observe this or not will tell us a lot about the new particle that has been observed.

$H \to b\bar{b}$

- It's not easy though...
- Largest branching ratio for a 125 GeV Standard Model Higgs boson (~58%).
- But very large backgrounds! (QCD $b\bar{b}$ is 9 orders of magnitude larger)

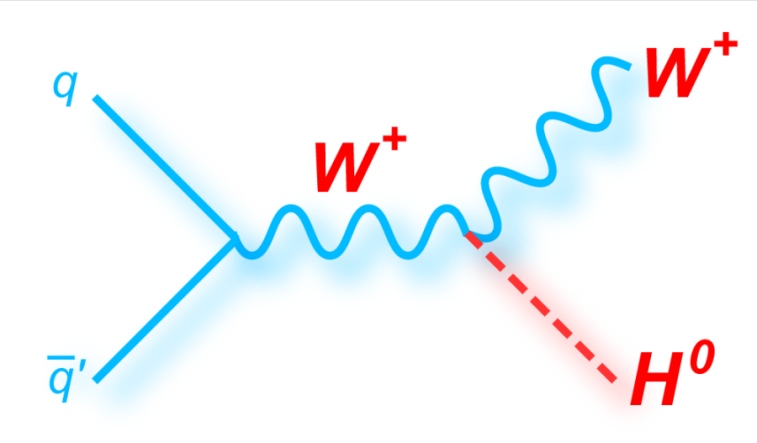


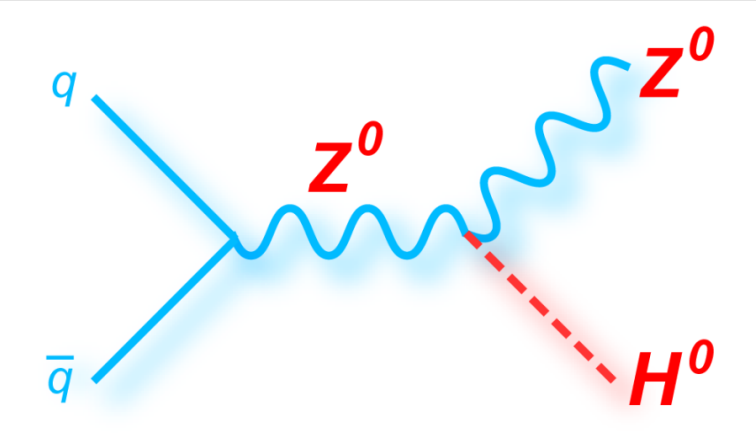


- To combat these backgrounds we use features of the production methods to help identify Higgs signals.
- The main production channels used are VH (WH and ZH) and $t\bar{t}H$.
- Low production cross section.

VH→b$\bar{b}$

• Make use of both VH production channels:





• Separate analysis into three identified final states based on the number of observed leptons (e or mu):

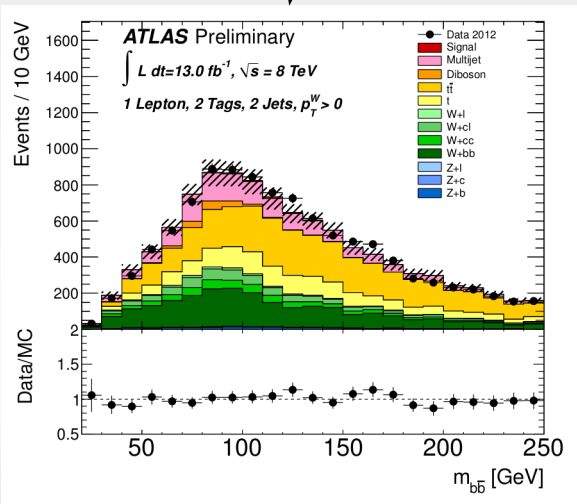
0 lepton	1 lepton	2 lepton
• Mostly ZH→vvbb • Some WH contribution • Large MET	• Mostly WH→lvbb • Reconstruct W from lepton + MET	• Mostly ZH→llbb • 2 charged leptons – clean signature

Also identify 2 b-tagged jets, with 70% efficiency and 1% fake rate

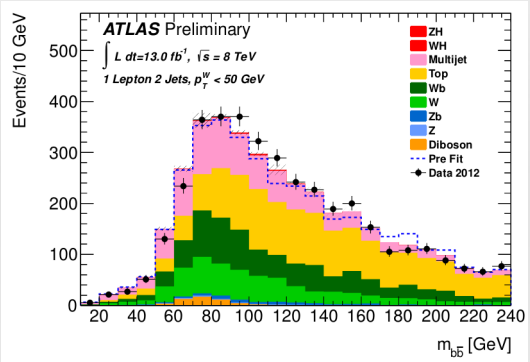
Signal regions by VpT

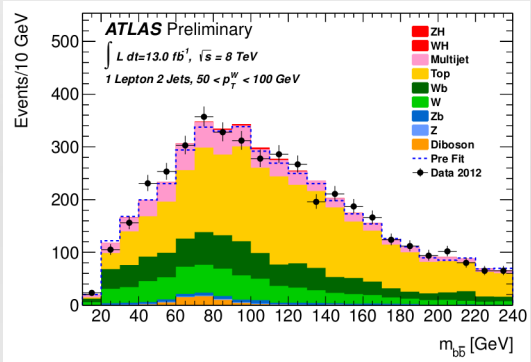
- Analyses separated into different bins based on the pT of the vector boson.
- This increases overall sensitivity as S/√B increases with vector boson pT.

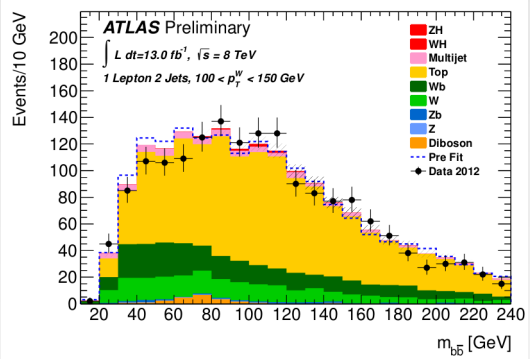
Combined

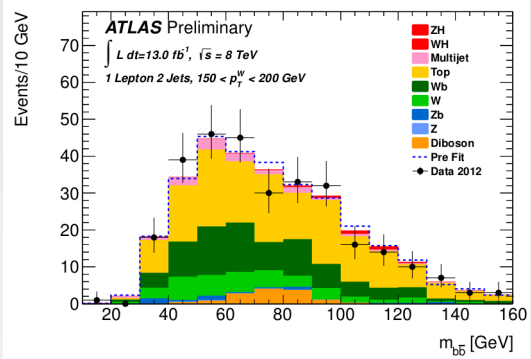


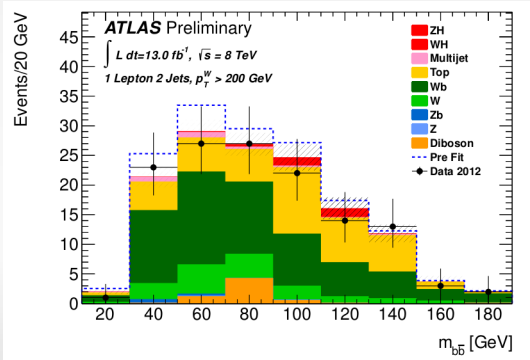
ATLAS-CONF-2012-161







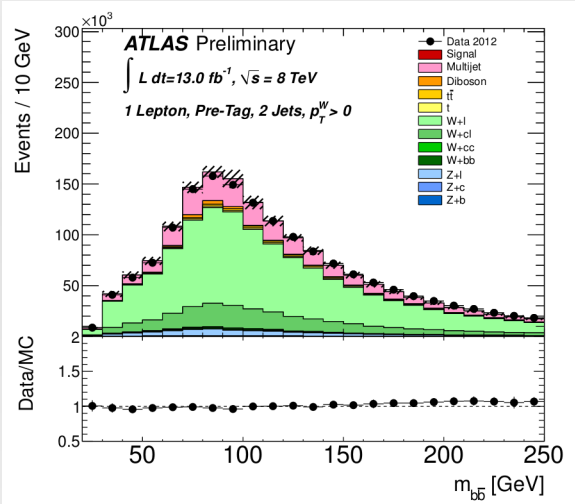




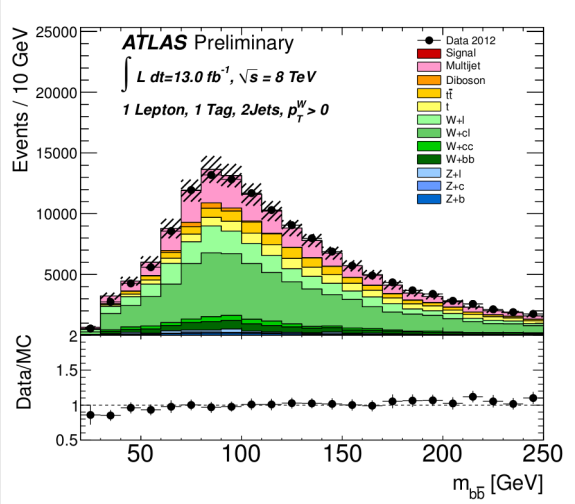
Backgrounds and control regions

- Define control regions to help isolate backgrounds.

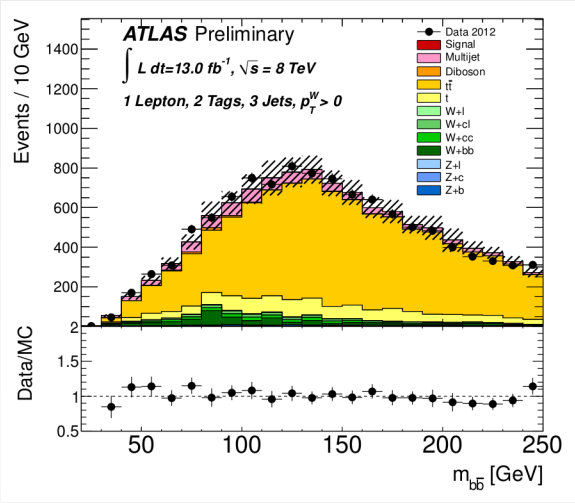
- Most background shapes taken from MC.
- Background normalisation taken from fits to data using all WH and ZH signal + control regions.
- m_{bb} is used as discriminating variable in fits.
- Multijet background estimated from data.
- Diboson shape and normalisation taken from MC.



2 jet, pre-tag



2 jet, 1 b-tag

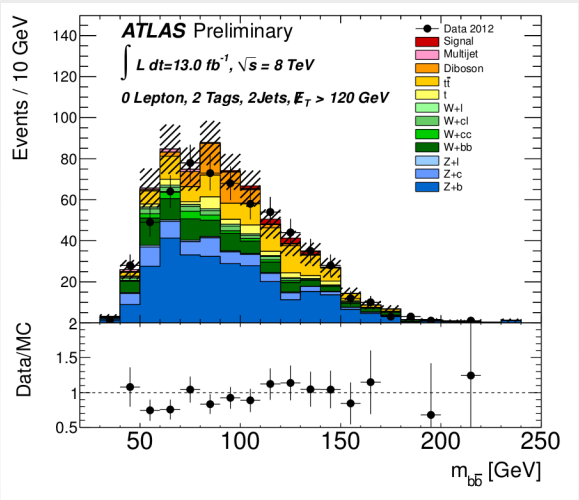


3 jet, 2 b-tag

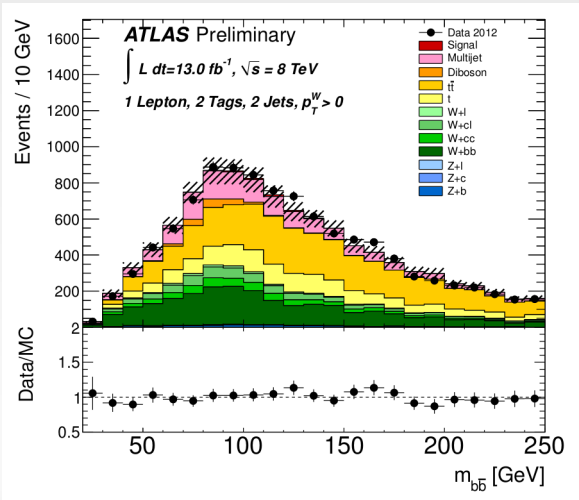
Backgrounds and signal regions

- Background composition in each of the 3 channels:
- m_{bb} signal region for each channel shown below.

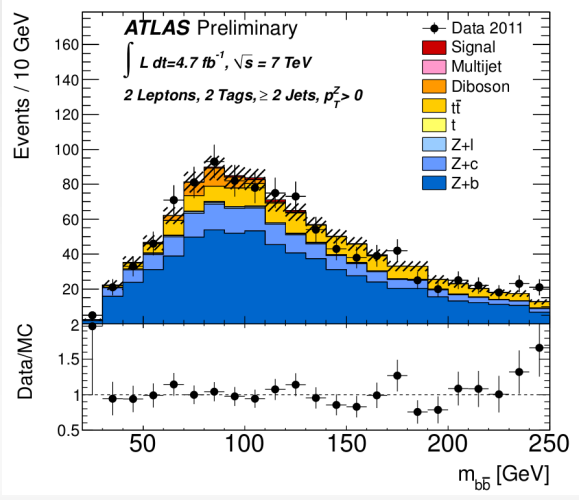
• Signal	Pythia8
• Multijet	Data driven
• Diboson	Herwig
• $t\bar{t}$	MC@NLO
• t	AcerMC/MC@NLO
• W+l	Alpgen
• W+c	Alpgen
• W+b	Powheg
• Z+l	Sherpa
• Z+c	Sherpa
• Z+b	Sherpa



0 lepton



1 lepton



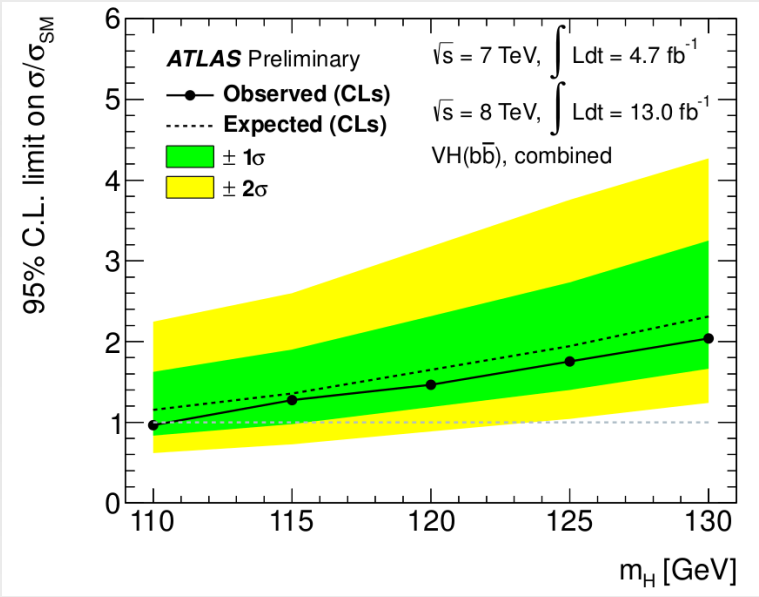
2 lepton

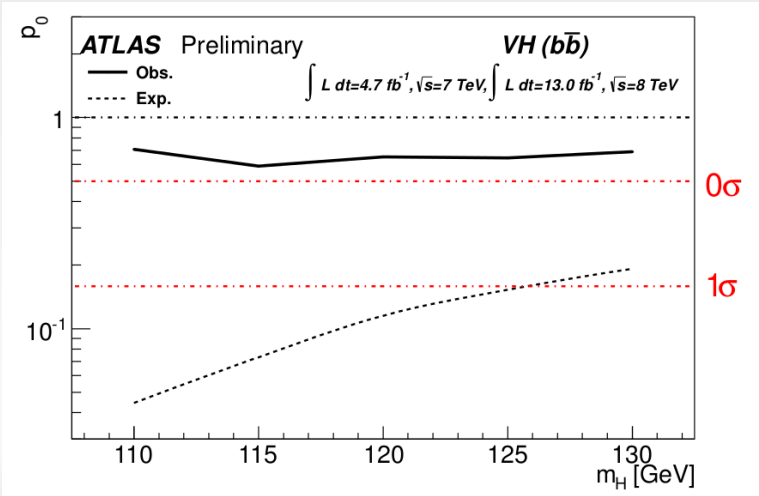
Combined HCP results

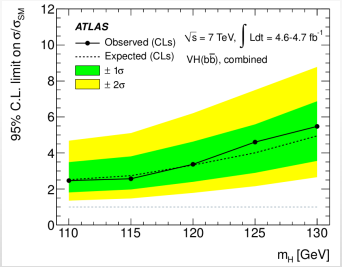
- Combining 7 TeV and 8 TeV 'HCP' data we get the following results:
- For m_H = 125 GeV, *expected* (**observed**)

  Limits: *1.9* (**1.8**) x SM
  p_0: *0.15* (**0.64**)

  σ/σ_{SM}: **$\mu = -0.4 \pm 0.7(\text{stat.}) \pm 0.8(\text{syst.})$**
- SM Higgs boson decaying to $b\bar{b}$ with m_H = 110 GeV has been excluded at 95% confidence level.
- Analysis sensitivity has been more than doubled. (was 4.0 in previous analysis: http://arxiv.org/abs/1207.0210)







Systematic uncertainties

- But what about systematic uncertainties?
- I will focus on current work to improve the estimates of the signal theoretical uncertainties.

- Below are tables of systematic uncertainties from the HCP analysis. ATLAS-CONF-2012-161
- Theoretical systematics are large (as large as statistical uncertainties below).
- Up until now, the difference between Pythia6 and Herwig++ with Powheg has been used as the of the size of the systematic uncertainty.

Uncertainty [%]	0 lepton		1 lepton	2 leptons
	ZH	WH	WH	ZH
b-tagging	8.9	9.0	8.8	8.6
Jet/Pile-up/$E_{\mathrm{T}}^{\mathrm{miss}}$	19	25	6.7	4.2
Lepton	0.0	0.0	2.1	1.8
$H \to bb$ BR	3.3	3.3	3.3	3.3
VH p_T-dependence	5.3	8.1	7.6	5.0
VH theory PDF	3.5	3.5	3.5	3.5
VH theory scale	1.6	0.4	0.4	1.6
Statistical	4.9	18	4.1	2.6
Luminosity	3.6	3.6	3.6	3.6
Total	24	34	16	13

Signal uncertainties

Uncertainty [%]	0 lepton	1 lepton	2 leptons
b-tagging	6.5	6.0	6.9
c-tagging	7.3	6.4	3.6
light tagging	2.1	2.2	2.8
Jet/Pile-up/$E_{\mathrm{T}}^{\mathrm{miss}}$	20	7.0	5.4
Lepton	0.0	2.1	1.8
Top modelling	2.7	4.1	0.5
W modelling	1.8	5.4	0.0
Z modelling	2.8	0.1	4.7
Diboson	0.8	0.3	0.5
Multijet	0.6	2.6	0.0
Luminosity	3.6	3.6	3.6
Statistical	8.3	3.6	6.6
Total	25	15	14

Background uncertainties

Theoretical systematic uncertainties

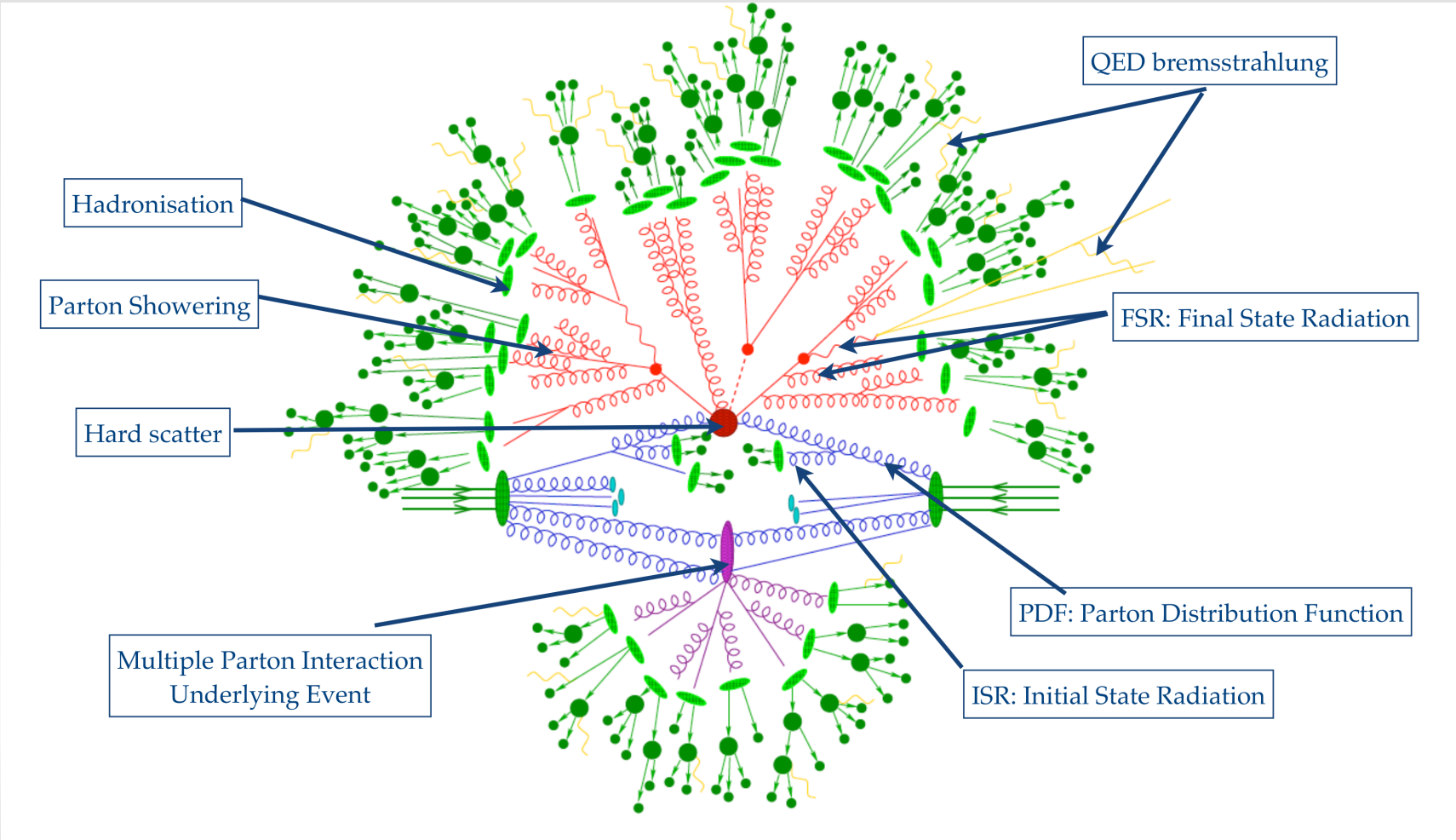
- What should be considered when estimating theoretical systematic uncertainties?

Theoretical systematic uncertainties

- What should be considered when estimating theoretical systematic uncertainties?
  - Parton Density Functions (PDFs)
  - Perturbative matrix element and shower scales
  - Parton shower models
  - Hadronization
  - Underlying event
  - Electroweak corrections

Theoretical systematic uncertainties

- What should be considered when estimating theoretical systematic uncertainties?



Thank you to Dr Veronique Boisvert!

Theoretical systematic uncertainties

- What should be considered when estimating theoretical systematic uncertainties?
    - Parton Density Functions (PDFs)
    - Perturbative matrix element and shower scales
    - Parton shower models
    - Hadronization
    - Underlying event
    - Electroweak corrections
- I will just give an overview of a few of the above.

PDF uncertainties

- Cross sections of processes are dependent on the PDFs of the incoming particles:

$$\sigma = \sum_{flav} \int dx_1 dx_2 \, PDF(x_1,Q^2,flav_1) \, PDF(x_2,Q^2,flav_2) \, \underbrace{\sigma(p_1,p_2,\alpha_s(Q^2),Q^2)}_{\text{Hard subprocess}}$$

- We want to see how distributions change when different PDFs are used.
- It is impractical however to generate events separately for all available PDFs.
- A solution is PDF reweighting:

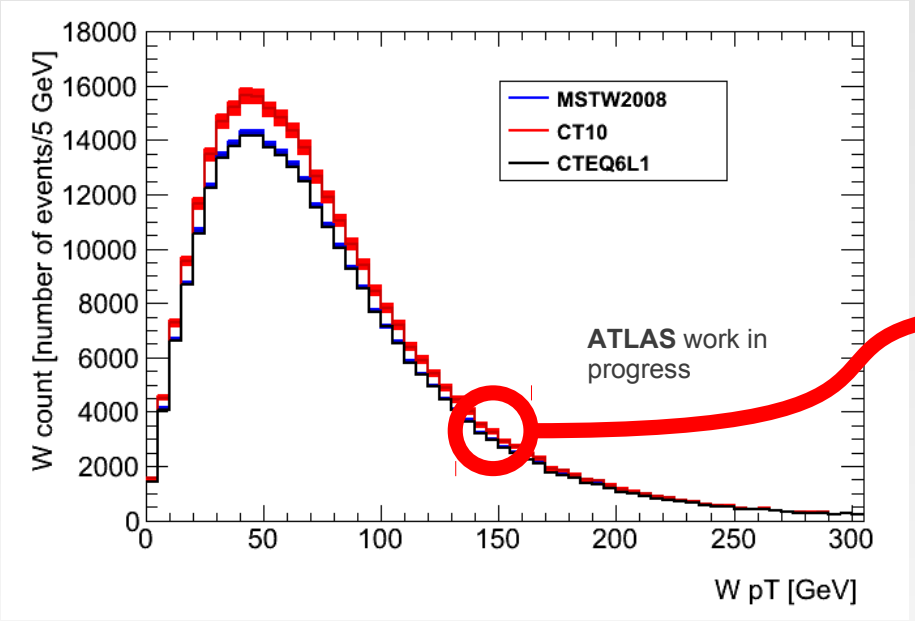
$$\text{Weight} = \frac{PDF_{new}(x_1,Q^2,flav_1) \, PDF_{new}(x_2,Q^2,flav_2)}{PDF_{old}(x_1,Q^2,flav_1) \, PDF_{old}(x_2,Q^2,flav_2)}$$

- Generate events once.
- Reweight generated distributions, event by event, for different PDFs.

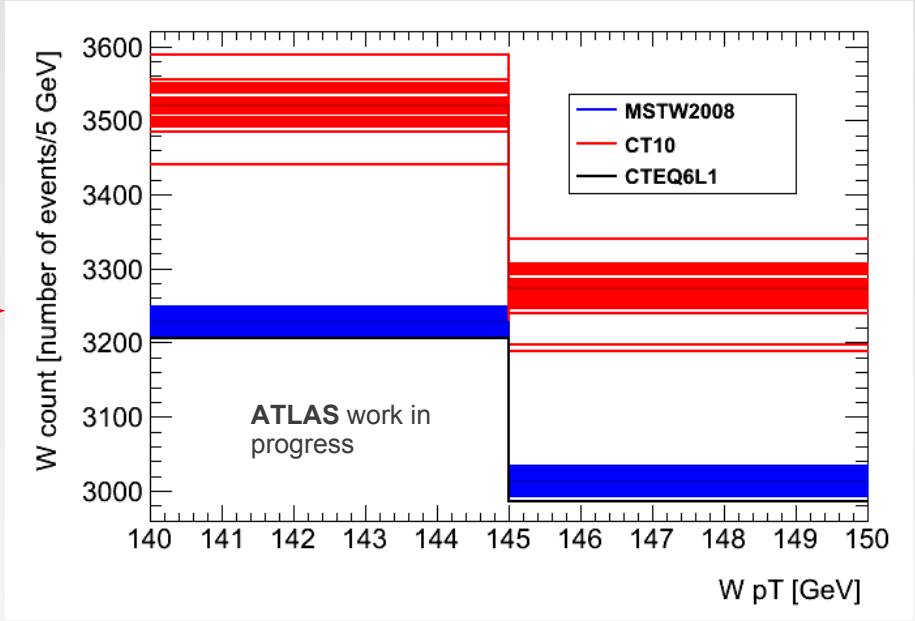
PDF uncertainties

- Reweighted WH→lvbb events. Generator-level simulated data.

W pT

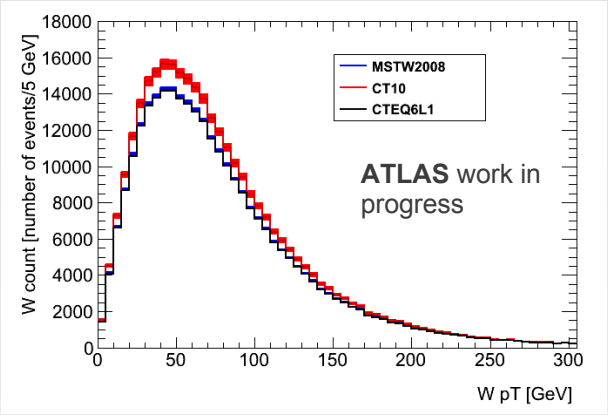


W pT (zoomed in)

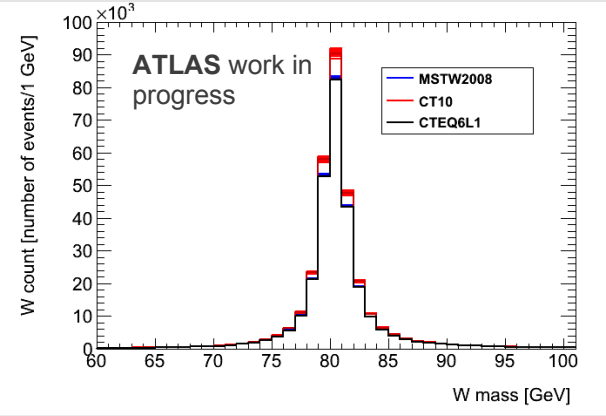


PDF uncertainties

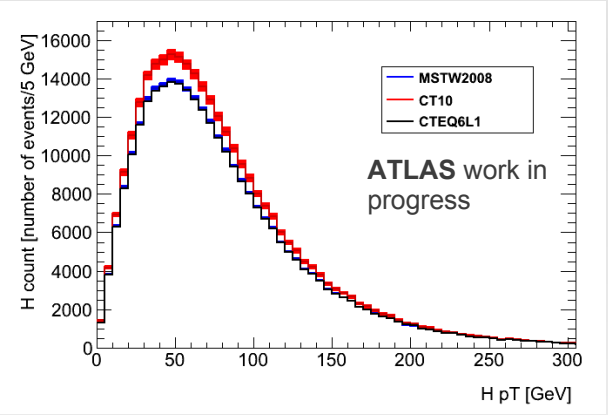
- Reweighted WH→lvbb events. Generator-level simulated data.



W pT



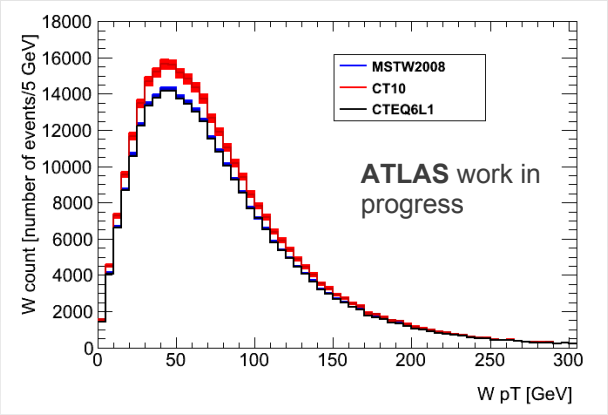
W mass

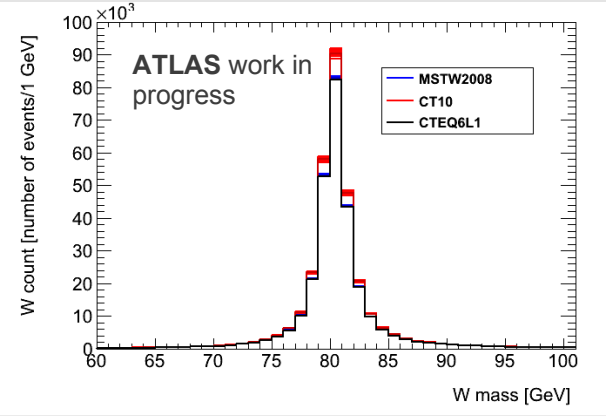


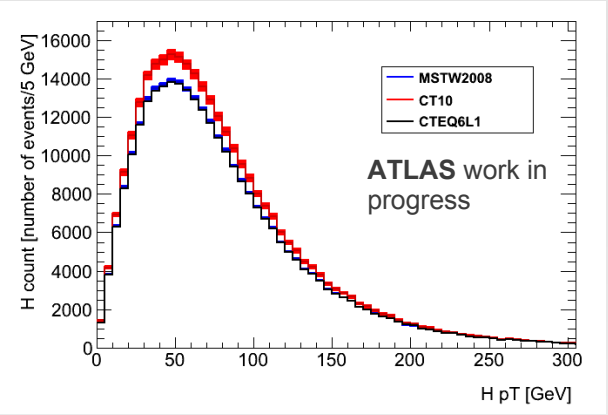
H pT

PDF uncertainties

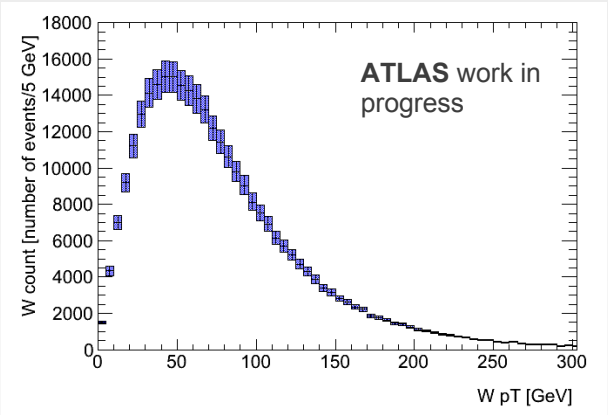
- Reweighted WH→lvbb events. Generator-level simulated data.



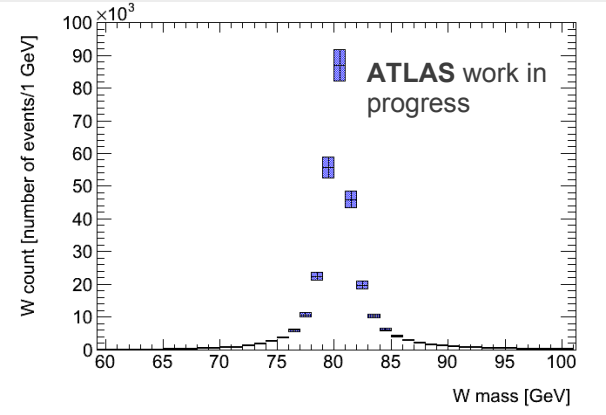




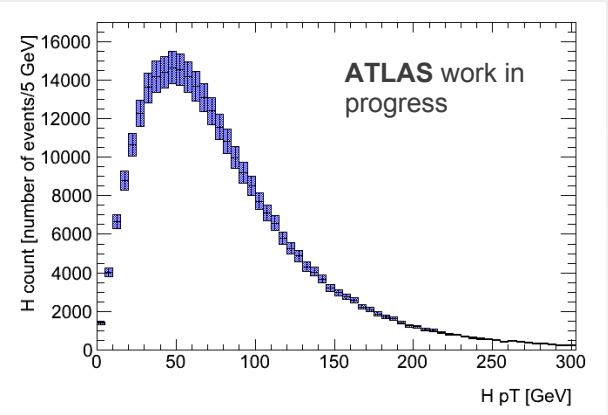
W pT



W mass



H pT



Dressing uncertainties

- 'Event dressing':

  Collective term for the underlying event (MPI), hadronization and showering processes.

- Different models and tunes of the above proceses will introduce uncertainties.

- Isolating each process is difficult:

  Comparing details of event records between different generators is dangerous.

  If one process is switched off to study its effect, another process may compensate.

  The effects of these models are anticorrelated, so naively adding the uncertainties from each separate process will overestimate the total uncertainty.

- All event dressing processes should be treated together

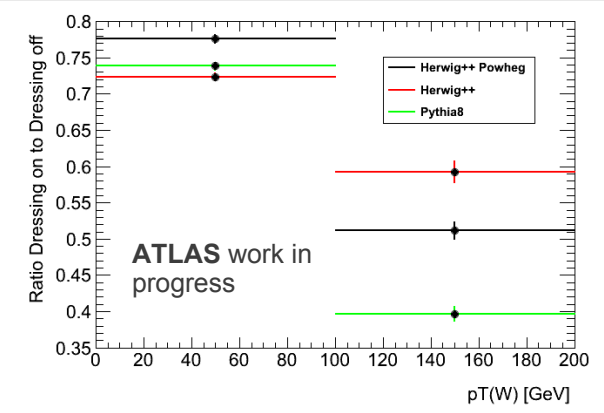
- Switch on and off all event dressing processes and study the difference in effect between different generators.

Dressing uncertainties

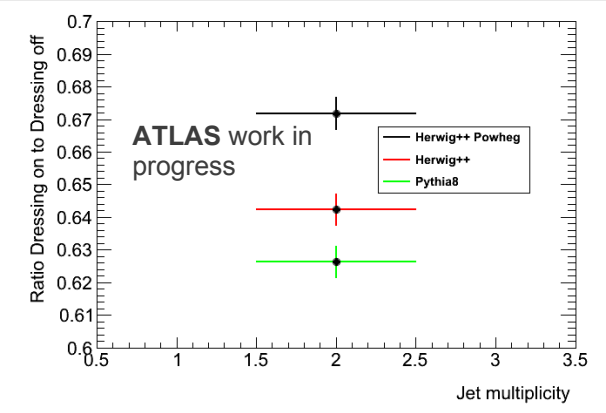
- By comparing the change from switching on and off processes, between different generators, we can get an estimate of the uncertainty from those processes.

- All event dressing processes are collectively switched on and off.
- The ratio of on/off is shown:

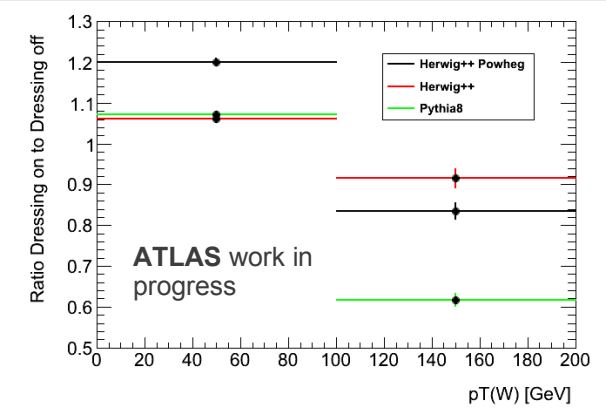
W pT



Signal jet multiplicity



2-jet cut efficiency



- Caveat: These plots are preliminary and not meant as a final estimation of the uncertainty.

Conclusion

- The search for H→bb at ATLAS is progressing rapidly.
- Aim to present results using full 2011 + 2012 data at the LHCP conference in May in Barcelona.

- Work is ongoing to estimate theoretical systematic uncertainties from various sources.
- Aim to have this ready for LHCP.
- This will provide a much more robust estimation of theoretical systematic uncertainties for the H→bb analyses.

Conclusion

- The search for H→bb at ATLAS is progressing rapidly.
- Aim to present results using full 2011 + 2012 data at the LHCP conference in May in Barcelona.

- Work is ongoing to estimate theoretical systematic uncertainties from various sources.
- Aim to have this ready for LHCP.
- This will provide a much more robust estimation of theoretical systematic uncertainties for the H→bb analyses.

Watch this space!...

Thank you!

Backup slides

$H \to b\bar{b}$ in the WH channel – Analysis cuts

0-lepton channel			
E_T^{miss} (GeV)	120-160	160-200	>200
$\Delta R(b, \bar{b})$	0.7-1.9	0.7-1.7	<1.5

1-lepton channel					
p_T^W (GeV)	0-50	50-100	100-150	150-200	>200
$\Delta R(b, \bar{b})$	>0.7			0.7-1.6	<1.4
E_T^{miss} (GeV)	> 25				> 50
m_T^W(GeV)	> 40			-	

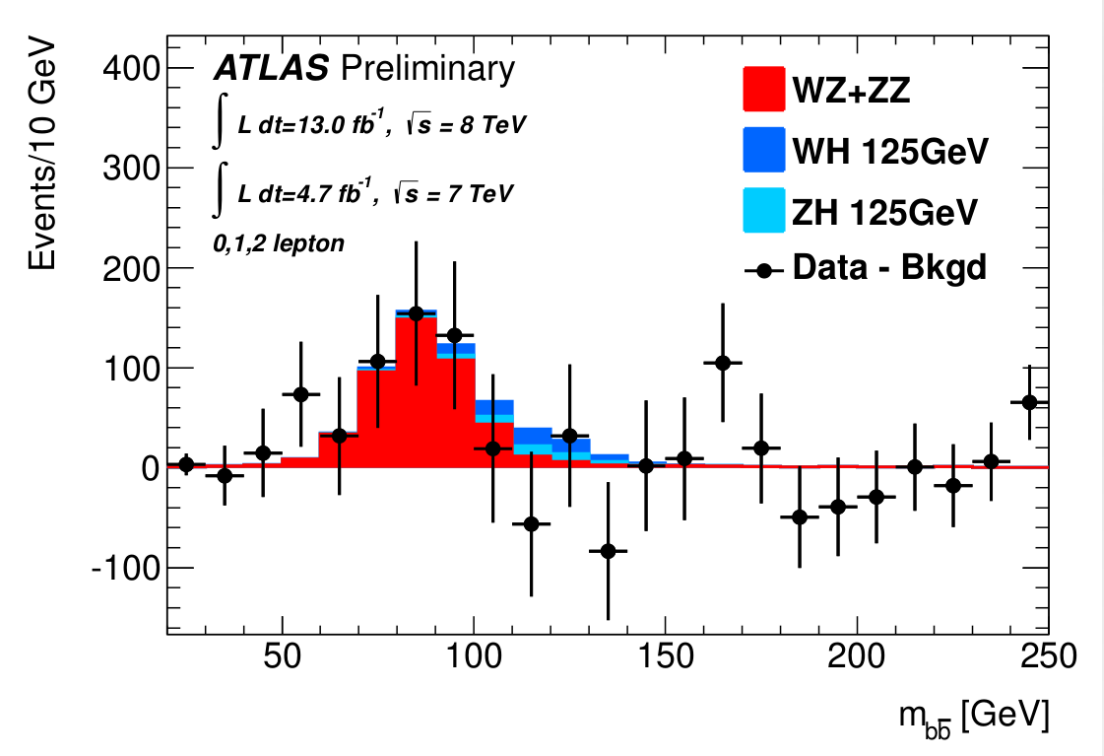
2-lepton channel					
p_T^Z(GeV)	0-50	50-100	100-150	150-200	>200
$\Delta R(b, \bar{b})$	>0.7			0.7-1.8	<1.6

Object	0-lepton	1-lepton	2-lepton
Leptons	0 loose leptons	1 tight lepton + 0 loose leptons	1 medium lepton + 1 loose lepton
Jets	2 b-tags $p_T^1 > 45$ GeV $p_T^2 > 20$ GeV + ≤ 1 extra jets	2 b-tags $p_T^1 > 45$ GeV $p_T^2 > 20$ GeV + 0 extra jets	2 b-tags $p_T^1 > 45$ GeV $p_T^2 > 20$ GeV -
Missing E_T	$E_T^{miss} > 120$ GeV $p_T^{miss} > 30$ GeV $\Delta\phi(E_T^{miss}, p_T^{miss}) < \pi/2$ $\mathrm{Min}[\Delta\phi(E_T^{miss}, \mathrm{jet})] > 1.5$ $\Delta\phi(E_T^{miss}, b\bar{b}) > 2.8$	-	$E_T^{miss} < 60$ GeV
Vector Boson	-	$m_T^W < 120$ GeV	$83 < m_{\ell\ell} < 99$ GeV

- For object selection please see paper: https://cds.cern.ch/record/1493625

Diboson validation

- To validate the $H \rightarrow b\bar{b}$ analysis a seperate fit was performed to search for diboson signal (WZ or ZZ with $Z \rightarrow b\bar{b}$).
- Similar signature to $H \rightarrow b\bar{b}$ but with 5 times larger cross-section.
- All backgrounds subtracted except diboson and 125 GeV Higgs.
- All three lepton channels combined (0, 1 and 2 leptons).



- Profile likelihood fit with full systematics performed.
- Clear excess is observed in data at expected mass.
- Resulting $\sigma/\sigma SM = \mu D = 1.09 \pm 0.20$ (stat) $\pm$ 0.22 (syst).
- This corresponds to a significance of 4.0 σ.

Combined results – Expected event yields

- Expected and observed number of events in 8 TeV data after profile likelihood fit.
- Number of expected signal events also shown for mH = 125 GeV.
- Quoted errors represent 1 standard deviation and include both systematic and statistical uncertainties.

	0-lepton, 2 jet			0-lepton, 3 jet			1-lepton					2-lepton				
Bin	E_T^{miss} [GeV]						p_T^W[GeV]					p_T^Z[GeV]				
	120-160	160-200	>200	120-160	160-200	>200	0-50	50-100	100-150	150-200	> 200	0-50	50-100	100-150	150-200	>200
ZH	2.9	2.1	2.6	0.8	0.8	1.1	0.3	0.4	0.1	0.0	0.0	4.7	6.8	4.0	1.5	1.4
WH	0.8	0.4	0.4	0.2	0.2	0.2	10.6	12.9	7.5	3.6	3.6	0.0	0.0	0.0	0.0	0.0
Top	89	25	8	92	25	10	1440	2276	1120	147	43	230	310	84	3	0
$W + c$,light	30	10	5	9	3	2	580	585	209	36	17	0	0	0	0	0
$W + b$	35	13	13	8	3	2	770	778	288	77	64	0	0	0	0	0
$Z + c$,light	35	14	14	8	5	8	17	17	4	1	0	201	230	91	12	15
$Z + b$	144	51	43	41	22	16	50	63	13	5	1	1010	1180	469	75	51
Diboson	23	11	10	4	4	3	53	59	23	13	7	37	39	16	6	4
Multijet	3	1	1	1	1	0	890	522	68	14	3	12	3	0	0	0
Total Bkg.	361	127	98	164	63	42	3810	4310	1730	297	138	1500	1770	665	97	72
	± 29	± 11	± 12	± 13	± 8	± 5	± 150	± 86	± 90	± 27	± 14	± 90	± 110	± 47	± 12	± 12
Data	342	131	90	175	65	32	3821	4301	1697	297	132	1485	1773	657	100	69

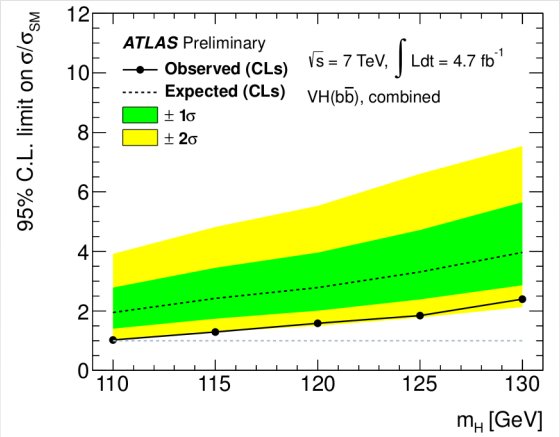
Combined HCP results – 7 TeV and 8 TeV HCP data separately

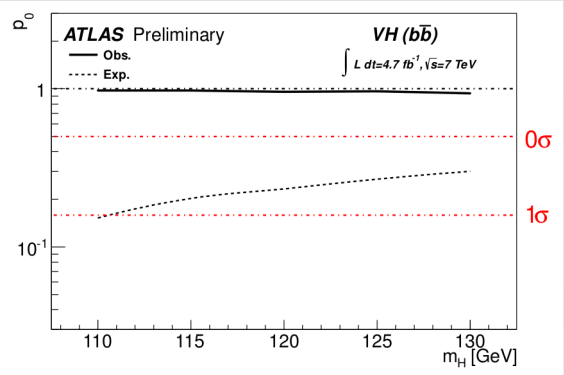
- Plots show results for all three associated production $H \to b\bar{b}$ channels, $ZH \to \nu\bar{\nu}b\bar{b}$, $ZH \to l\bar{l}b\bar{b}$ and $WH \to l\nu b\bar{b}$, separately for 7 TeV and 8 TeV HCP data.

- For m_H = 125 GeV in 7 TeV data, *expected* (**observed**):

  Limits: *3.3* (**1.8**) x SM
  p_0: *0.26* (**0.97**)

  σ/σ_{SM}: **μ = −2.7 ± 1.1(stat.) ± 1.1(syst.)**



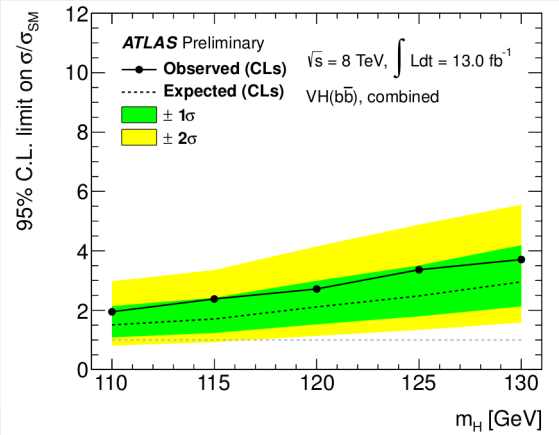


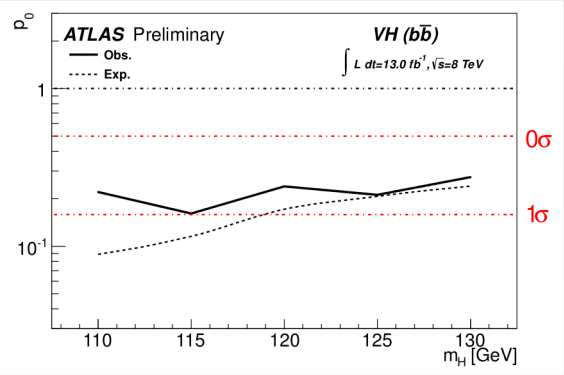
7 TeV data

- For m_H = 125 GeV in 8 TeV data, *expected* (**observed**):

  Limits: *2.5* (**3.4**) x SM
  p_0: *0.20* (**0.17**)

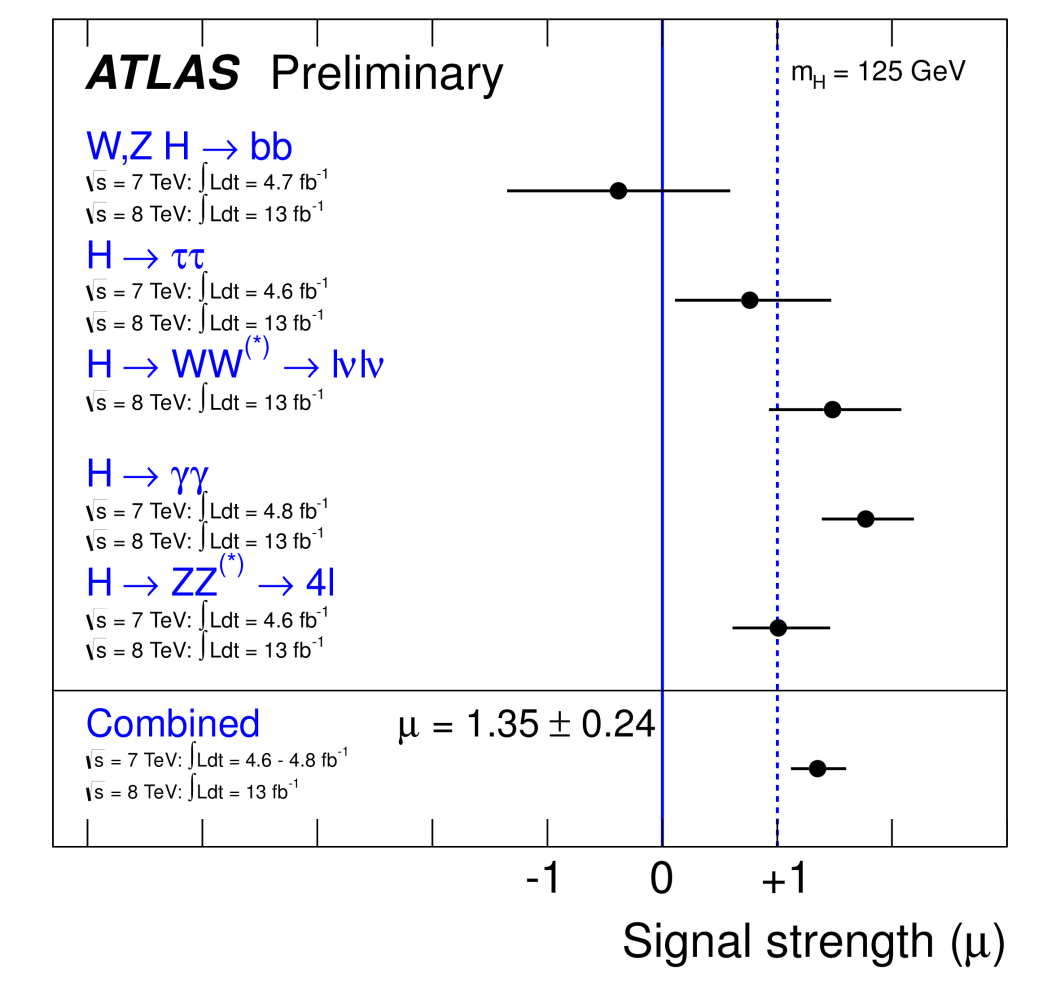
  σ/σ_{SM}: **μ = 1.0 ± 0.9(stat.) ± 1.1(syst.)**





8 TeV data

Negative μ value?



- From the following link
  https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2012-170/

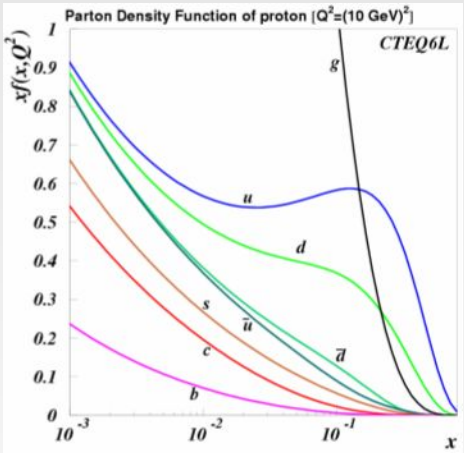
PDFs

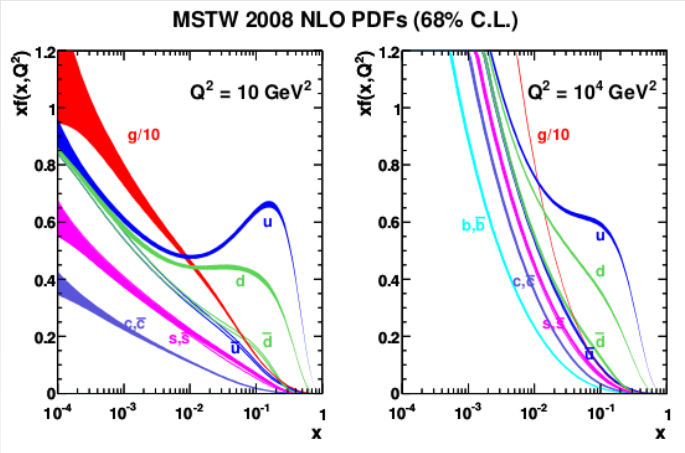
- What are PDFs?
  - Quark and gluon distributions inside hadrons cannot be calculated analytically.
  - PDFs describe how partons are distributed inside composite particles.
  - These distributions vary depending on parameters such as:

    The flavour of the parton.

    x = fraction of proton energy that a given parton has.

    Q = the energy scale of the collision.





(arXiv:0901.0002)

  - Each parton distribution is given by a function that can have free parameters.
  - These parameters are set by fitting the distributions in a PDF to data.