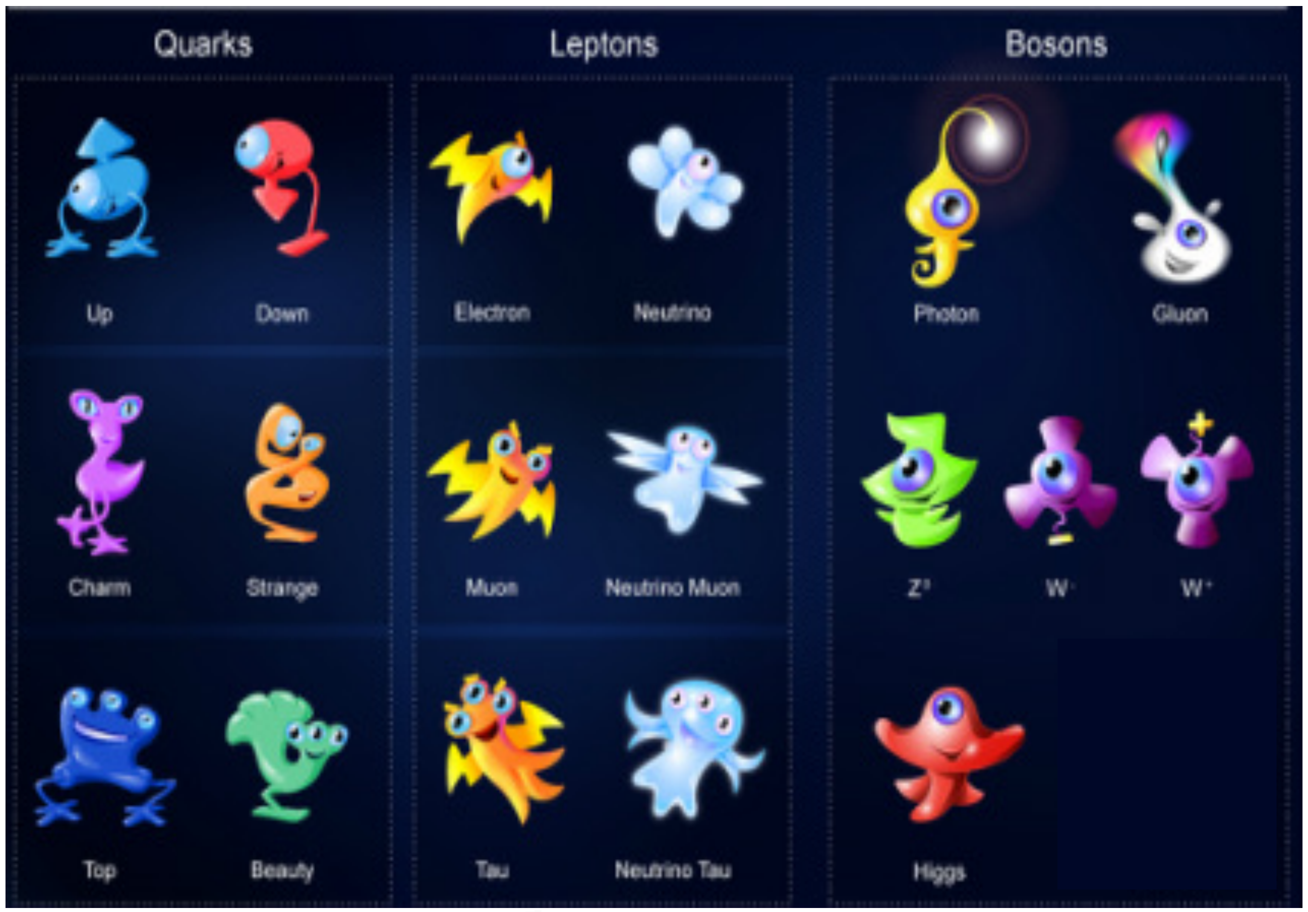
New method to derive systematic uncertainties in VHbb analysis with the ATLAS detector

Konie Al Khoury, Nicolas Morange

Journées de Rencontre des Jeunes Chercheurs 2018

18/10/2018

Higgs in the Standard Model



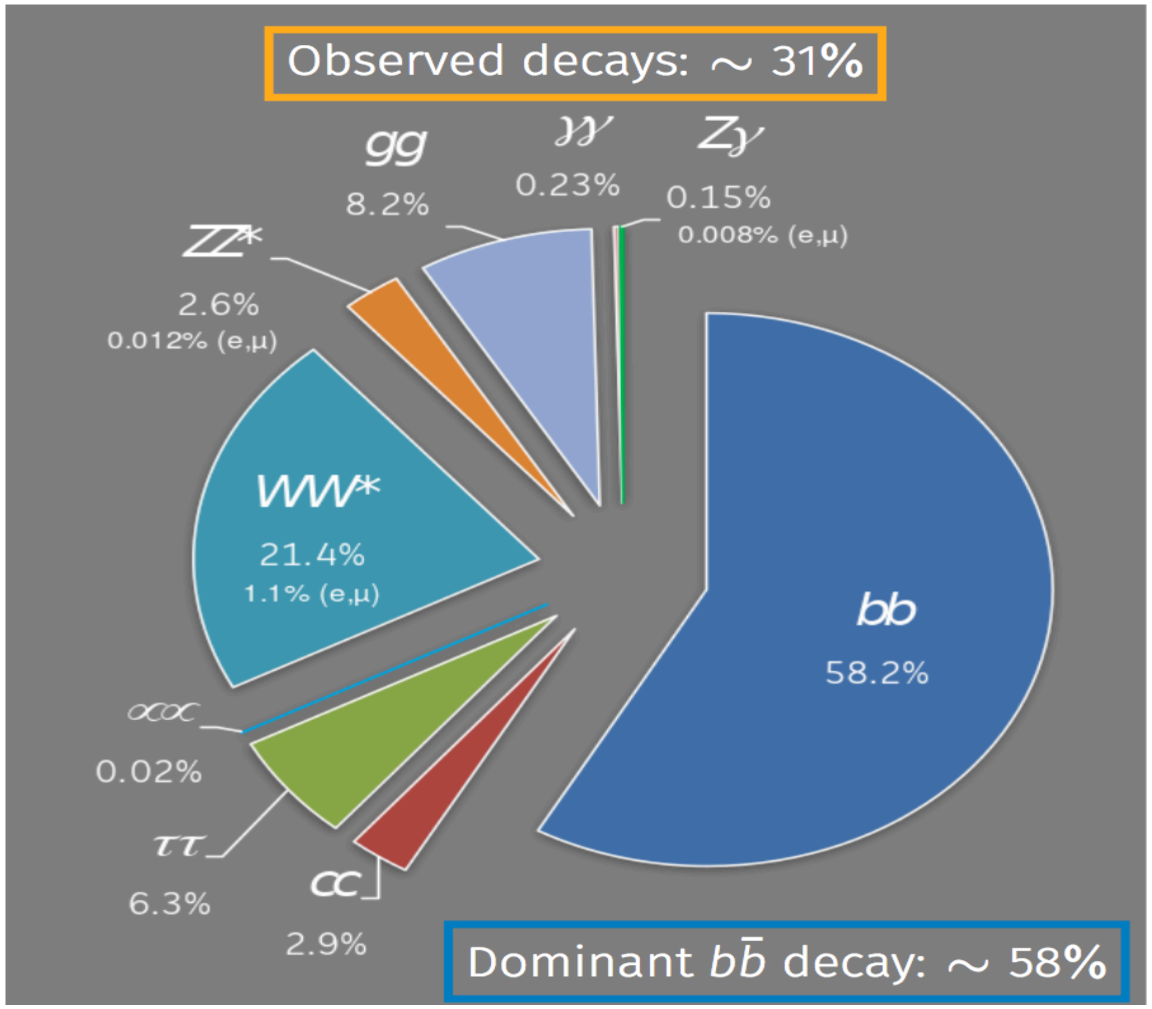
- 2 categories of particles: fermions (building Blocks of matter) and boson (forces mediators)
- The SM model describes 12 fermions (6 quarks and 6 leptons) and 6 bosons
- The Higgs mechanism gives mass to the fermions and bosons
- Higgs boson discovered in 2012 by ATLAS and CMS experiments
- A whole field of study of coupling between Higgs boson and different particles

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$
$$+ i\bar{\psi}\not{D}\psi + h.c$$
$$+ \psi_i y_{ij}\psi_j\phi + h.c$$
$$+ |D_\mu\phi|^2 - V(\phi)$$

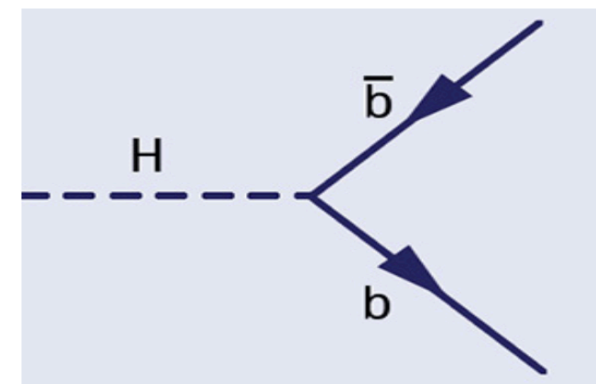
Coupling to Fermions

Coupling to bosons

H→bb Decay

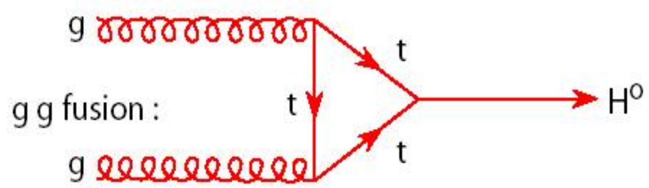


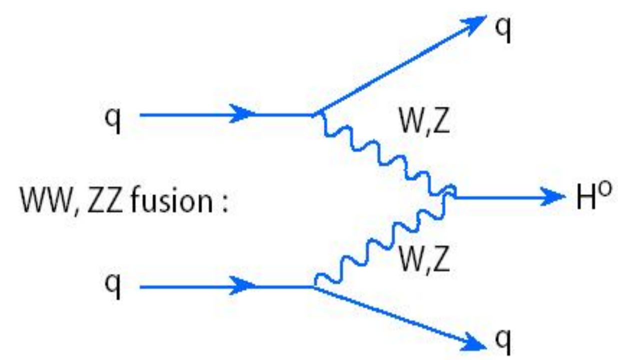
OBSERVED !

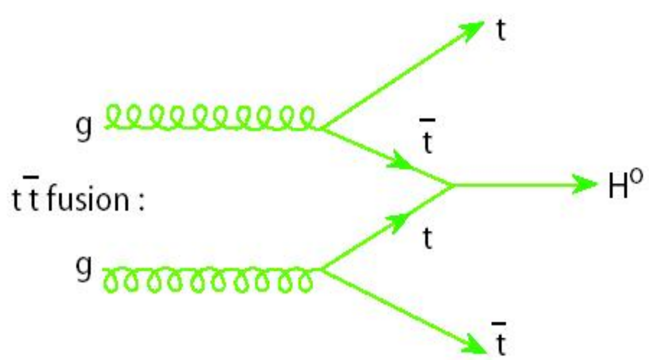


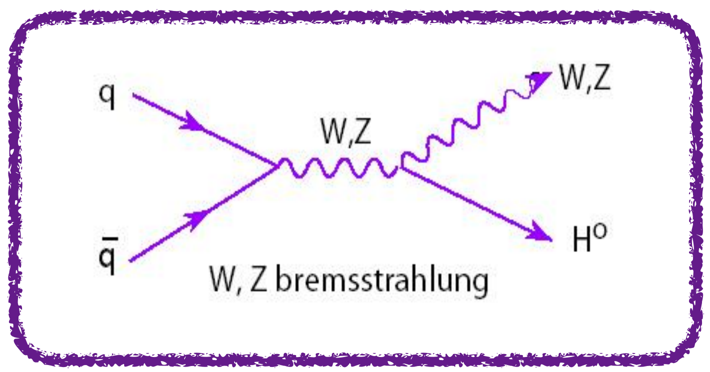
- Many decay modes: H→bb is the dominant with a BR ~58%
- Was recently observed by both ATLAS (https://arxiv.org/abs/1808.08238) and CMS (https://arxiv.org/abs/1808.08242) experiments
- This channel allows direct measurement of the Higgs coupling to b quarks
- Challenge: strong contamination by EW and QCD background

VHbb channel



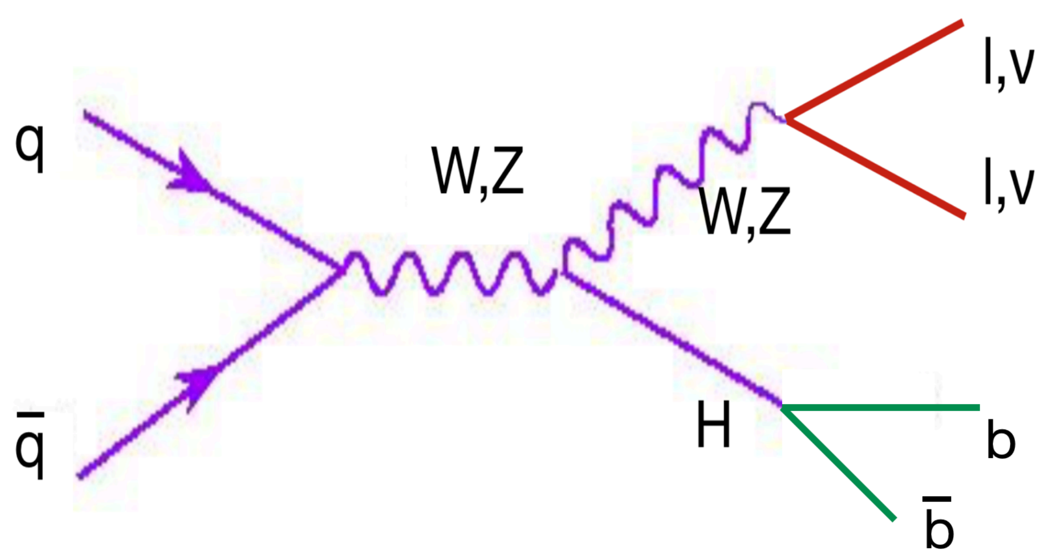






- 4 main Higgs production modes
- VHbb is the channel of interest for Hbb analysis: V (W/Z) boson created with the Higgs

- VH channel is interesting when W/Z decays leptonically (QCD background suppression)
- 3 channels of study: 0lepton (νν), 1lepton(lν) and 2leptons(ll)

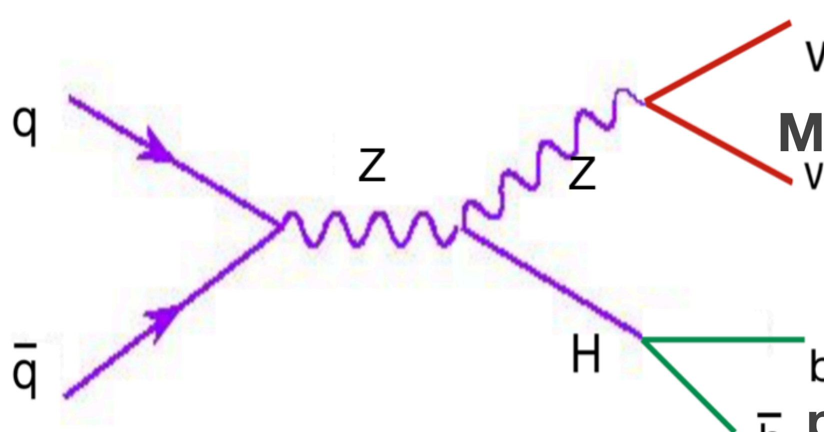


The most favourable channel since leptons have a clean signature in the detector and are easy to trigger on

VHbb Analysis event selection

- Large data set with Run2: 80 fb-1 at 13 TeV
- backgrounds: ttbar, multijet, single top, Diboson, V+jets
- Selection of exactly 2 b-jets events using an algorithm of b-tagging to eliminate additional light-(q,g) and c-jets contamination (~70% efficiency per b-jet)
- Events with 2 jets and 3 jets (with an additional jet) are treated separately
- Cuts are applied on the events to eliminate background and increase s/b ratio (cuts depend on the channel)

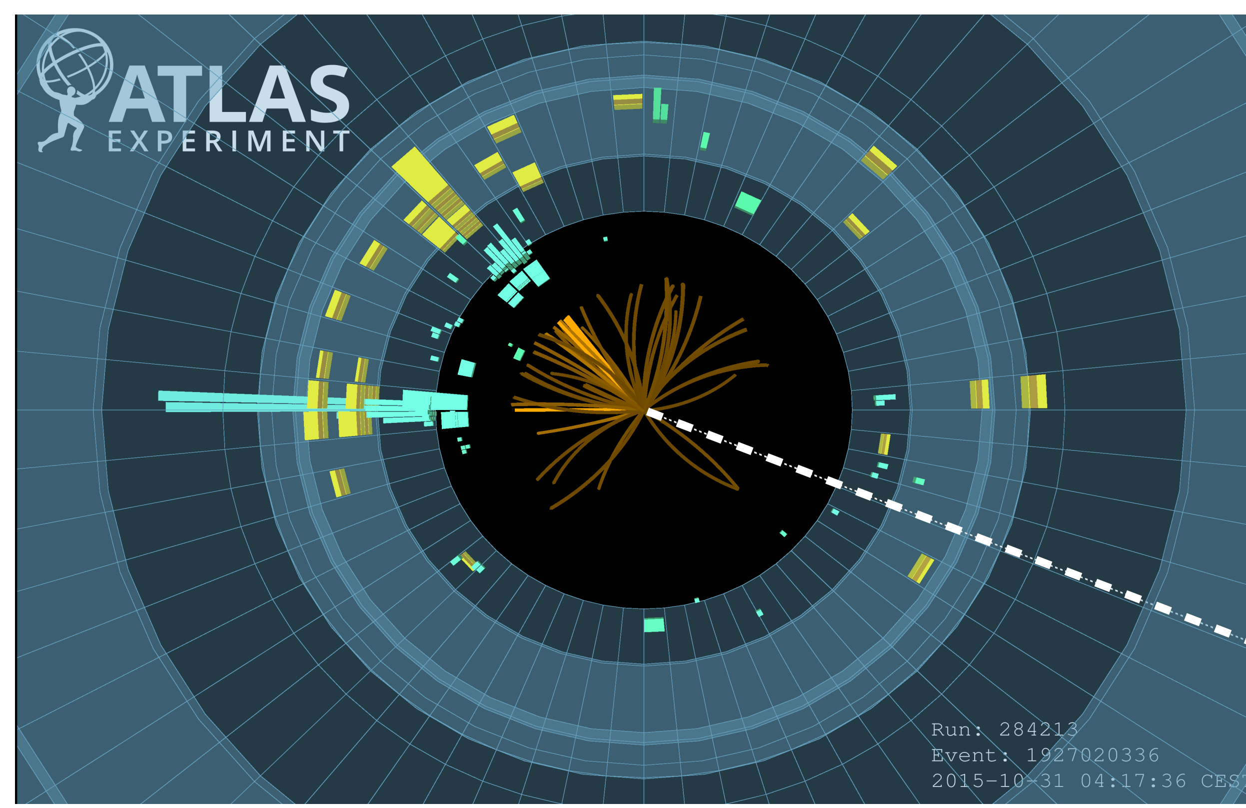
0lepton channel



MET >150GeV

pT b-jet1 >45GeV
pT b-jet2 >20GeV

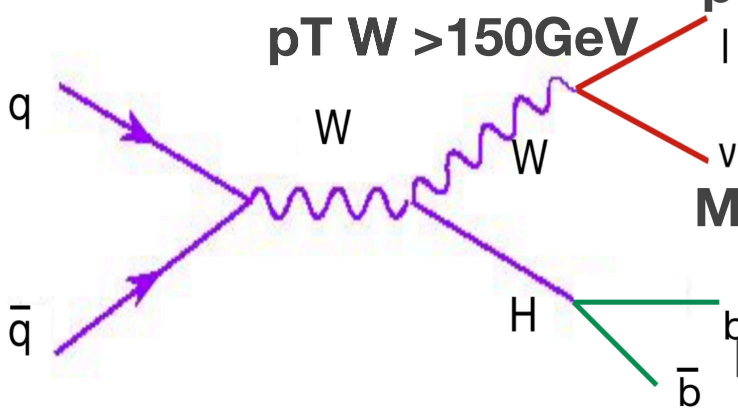
Angular cuts to remove events with fake MET



VHbb Analysis event selection

1lepton channel

pT W >150GeV



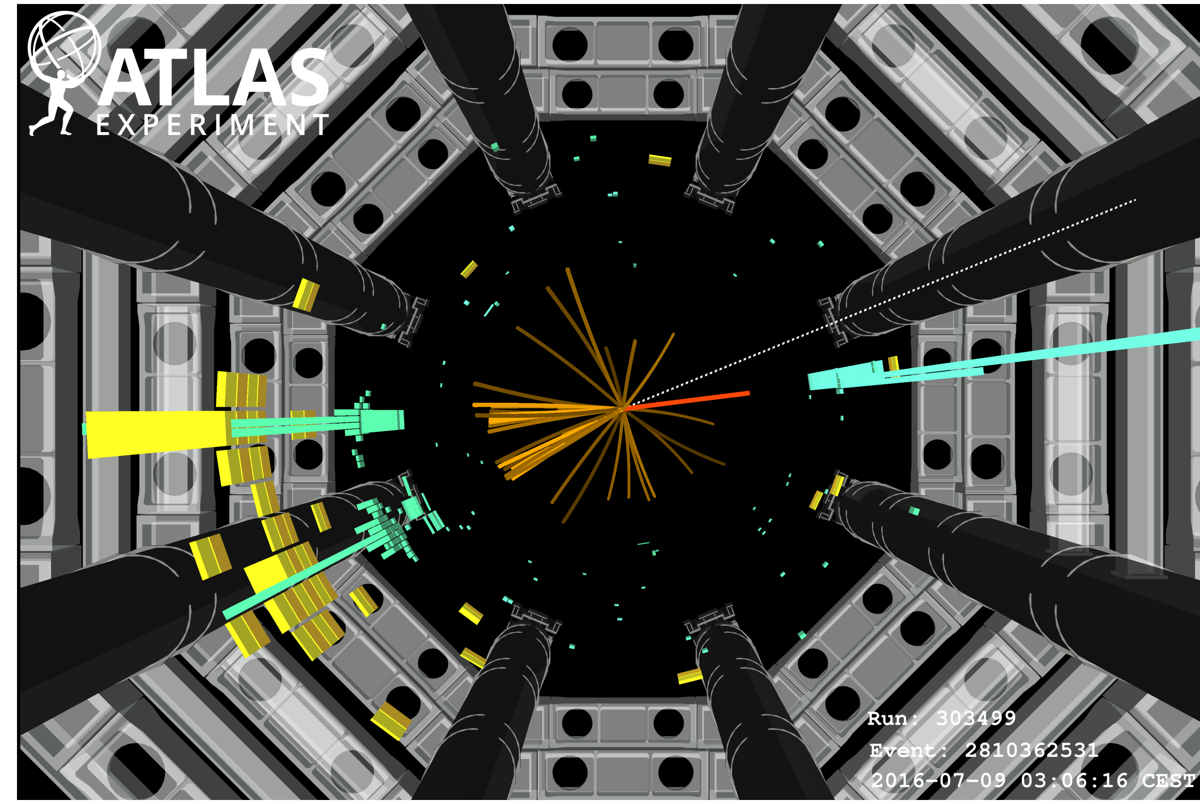
pT electron >27GeV

pT muon >25GeV

MET electron >30GeV

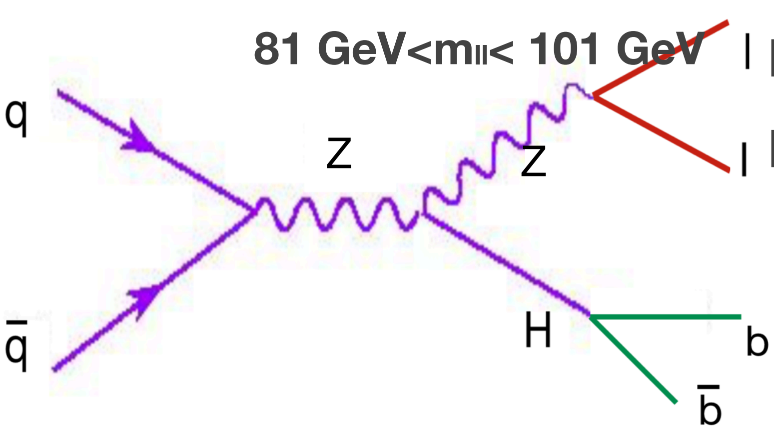
pT b-jet1 >45GeV

pT b-jet2 >20GeV



2lepton channel

$81\ \text{GeV} < m_{ll} < 101\ \text{GeV}$

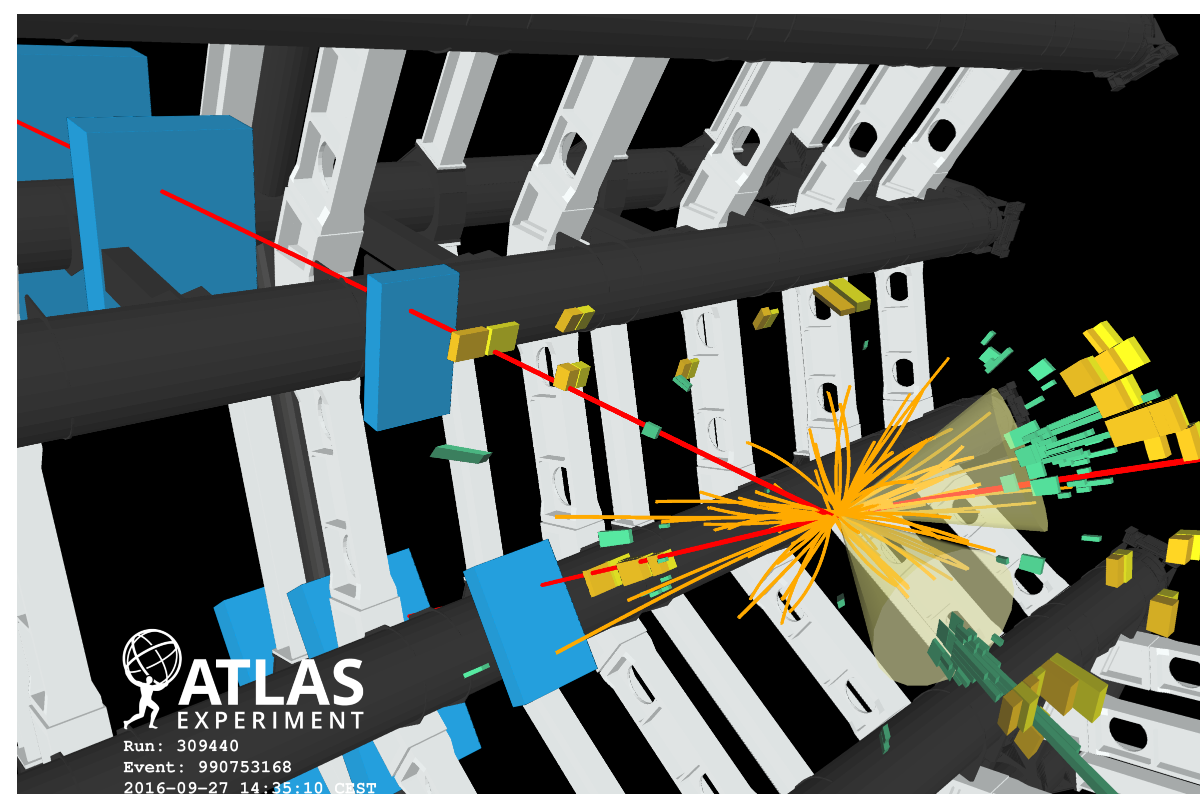


pT L1 >27GeV

pT L2 >7GeV

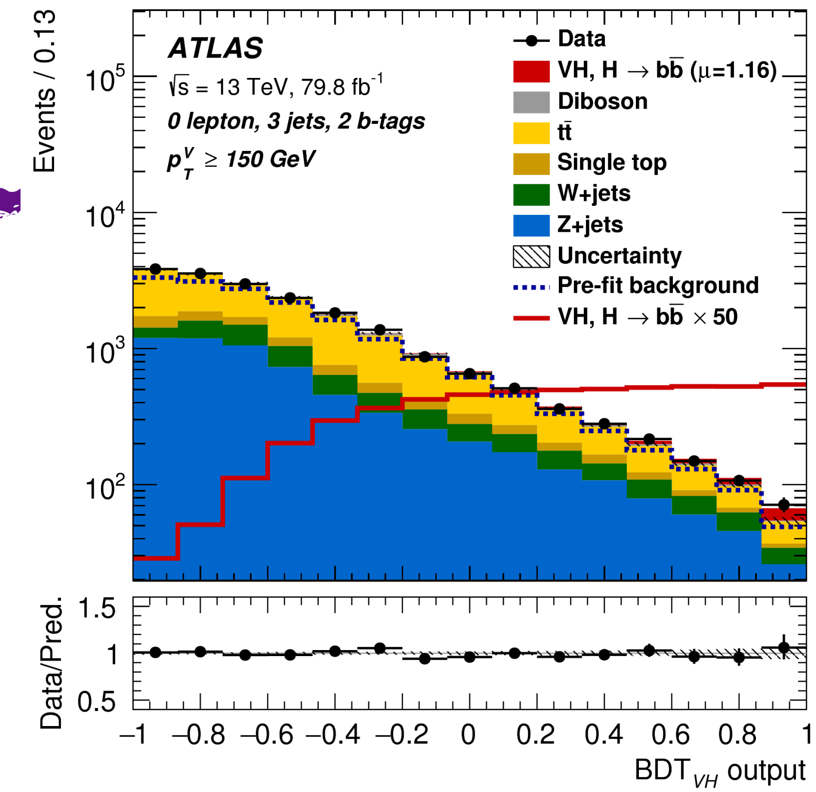
pT b-jet1 >45GeV

pT b-jet2 >20GeV



Mva Analysis

- The Analysis uses a mva analysis BDT method (Boosted Decision Trees)
- 14 kinematic variables used to build the BDT
- Separation into categories (2,3 jets) and regions (Signal Region and Control Region)
- mBB, dRBB and pTV being the most important variables for the classification
- BDT is well tuned for each channel (0,1and 2 lepton)
- A binned likelihood fit is done on mva to extract the significance
- Statistical and systematic uncertainties determine the measurement uncertainty of the significance
- Systematic uncertainties account for jets and MET calibrations, b-tagging efficiencies, pile-up corrections, luminosity uncertainties and MC modelling predictions
- Modelling systematics have a significant impact

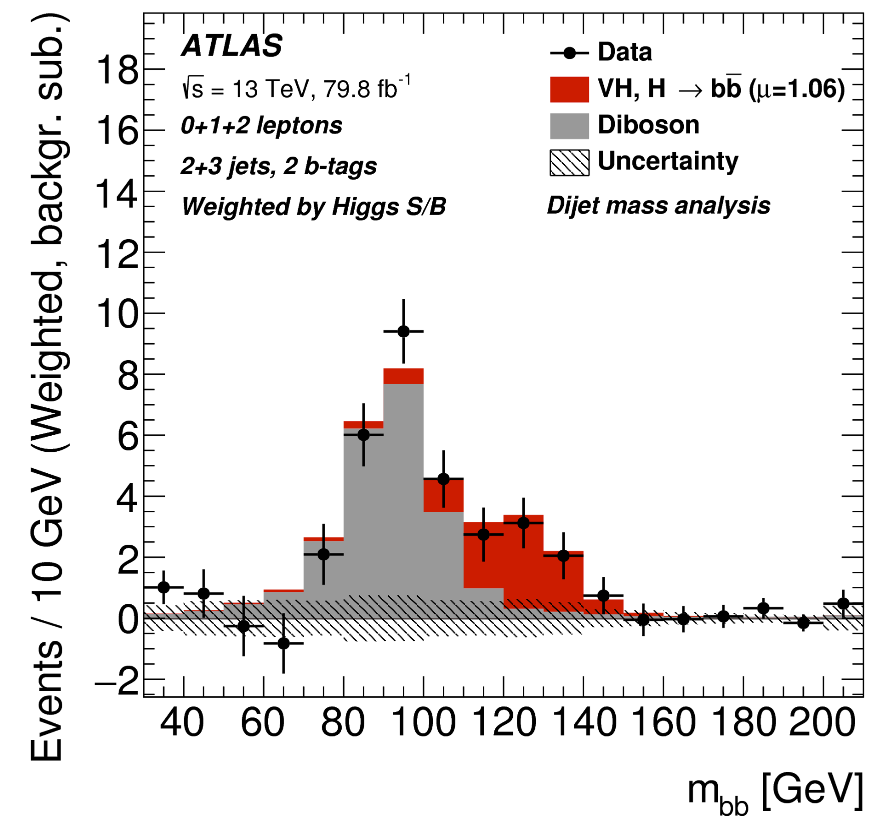


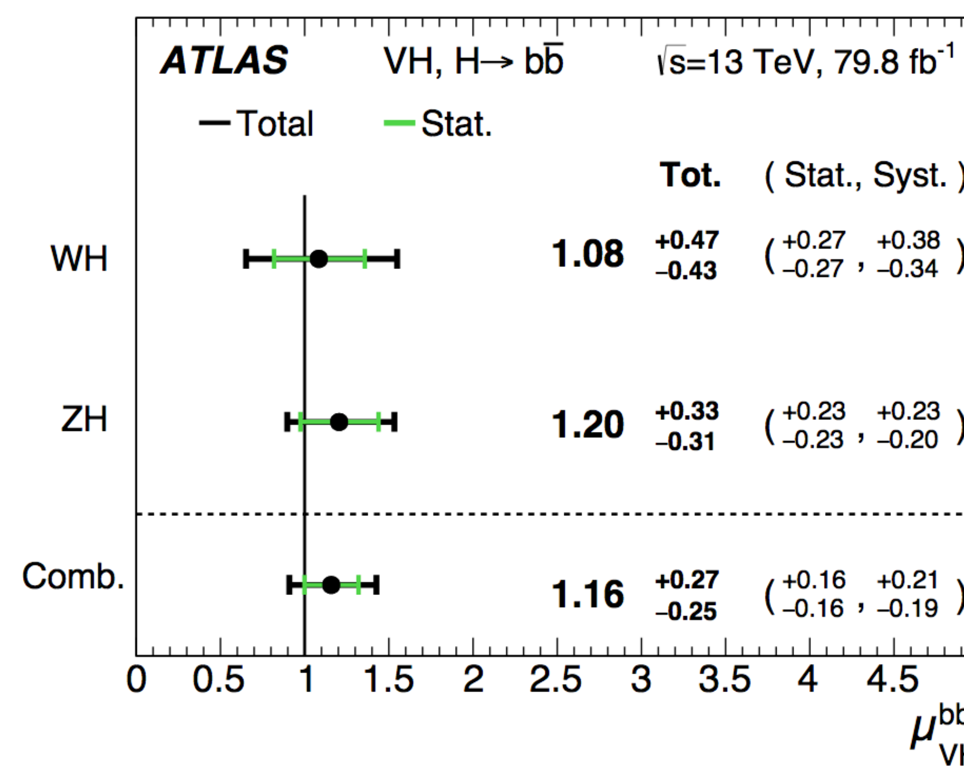
Contribution of the systematics on measurement uncertainty of the significance

Source of uncertainty		σ_μ
Total		0.259
Statistical		0.161
Systematic		0.203
Experimental uncertainties		
Jets		0.035
E_T^{miss}		0.014
Leptons		0.009
b-tagging	*b*-jets	0.061
	c-jets	0.042
	light-flavour jets	0.009
	extrapolation	0.008
Pile-up		0.007
Luminosity		0.023
Theoretical and modelling uncertainties		
Signal		0.094
Floating normalisations		0.035
Z + jets		0.055
W + jets		0.060
$t\bar{t}$		0.050
Single top quark		0.028
Diboson		0.054
Multi-jet		0.005
MC statistical		0.070

Observation of VH and Hbb with Run2 @ 13TeV

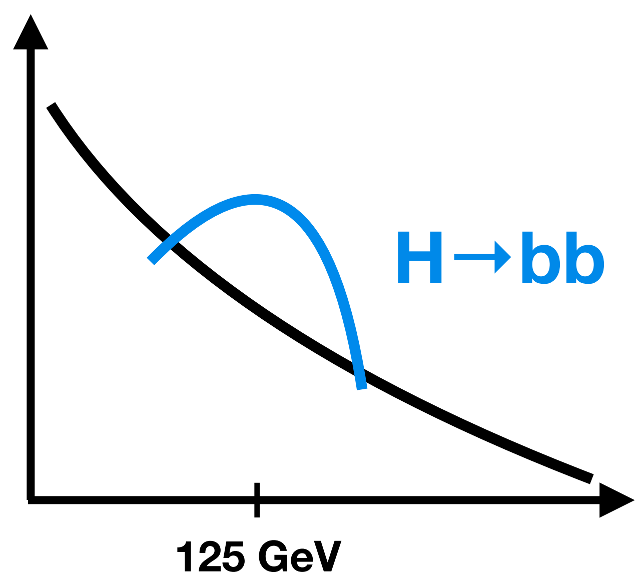
- Results cross checked with cut-based method (fit to mBB mass)
- Same analysis tested and checked by measuring VZ(bb) process
- VHbb analysis:
    - 4.9σ significance
    - μ =1.16 ± 0.26
- Observation of Hbb decay:
    - 5.4σ with combination with ttH and VBF production modes
    - μ =1.01 ± 0.2
- Observation of VH production channel:
    - 5.3σ with combination with γγ and 4l channel
    - μ =1.13 ± 0.24

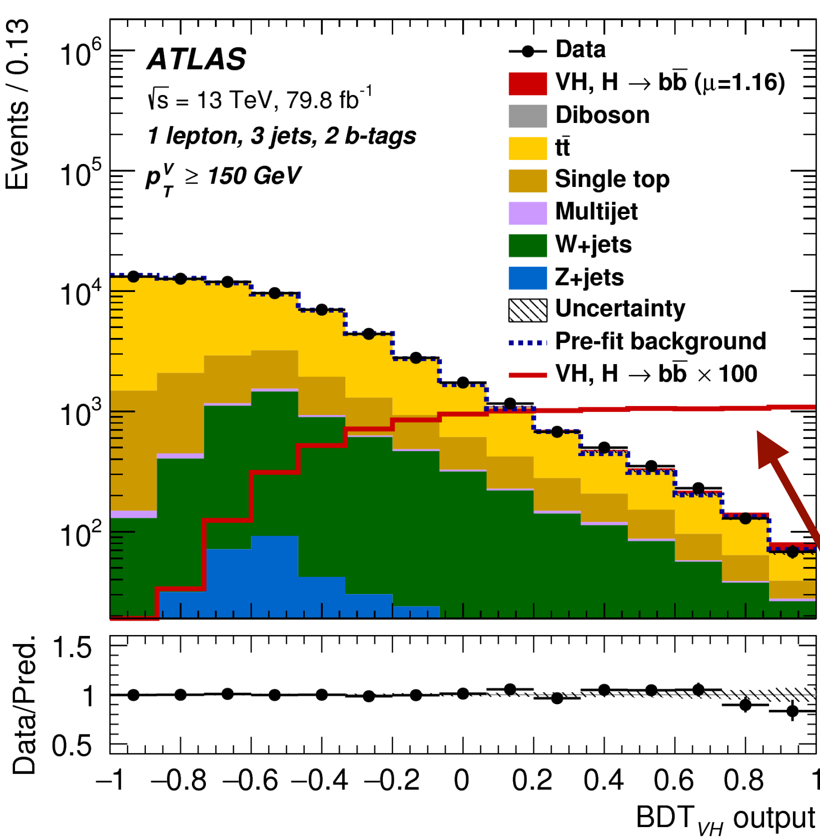




Why this long?

H→bb decay mode is the most dominant so why this long to be observed?

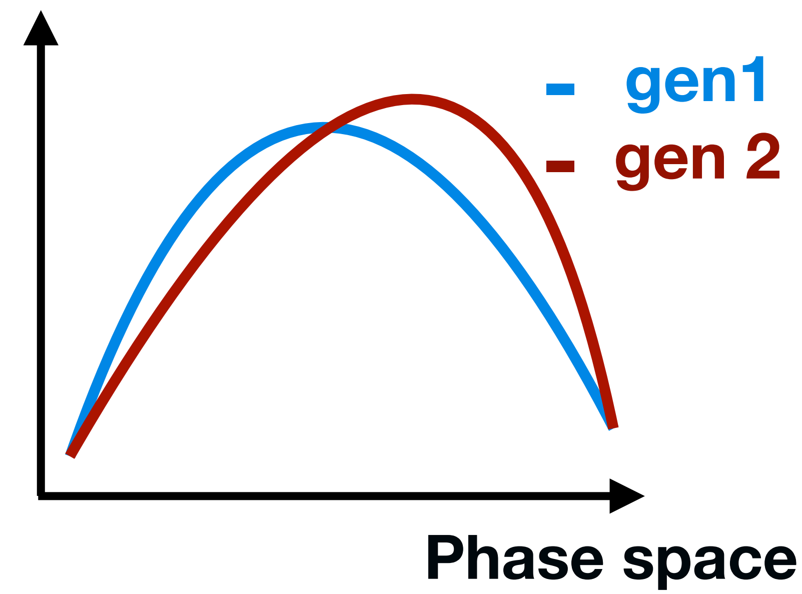




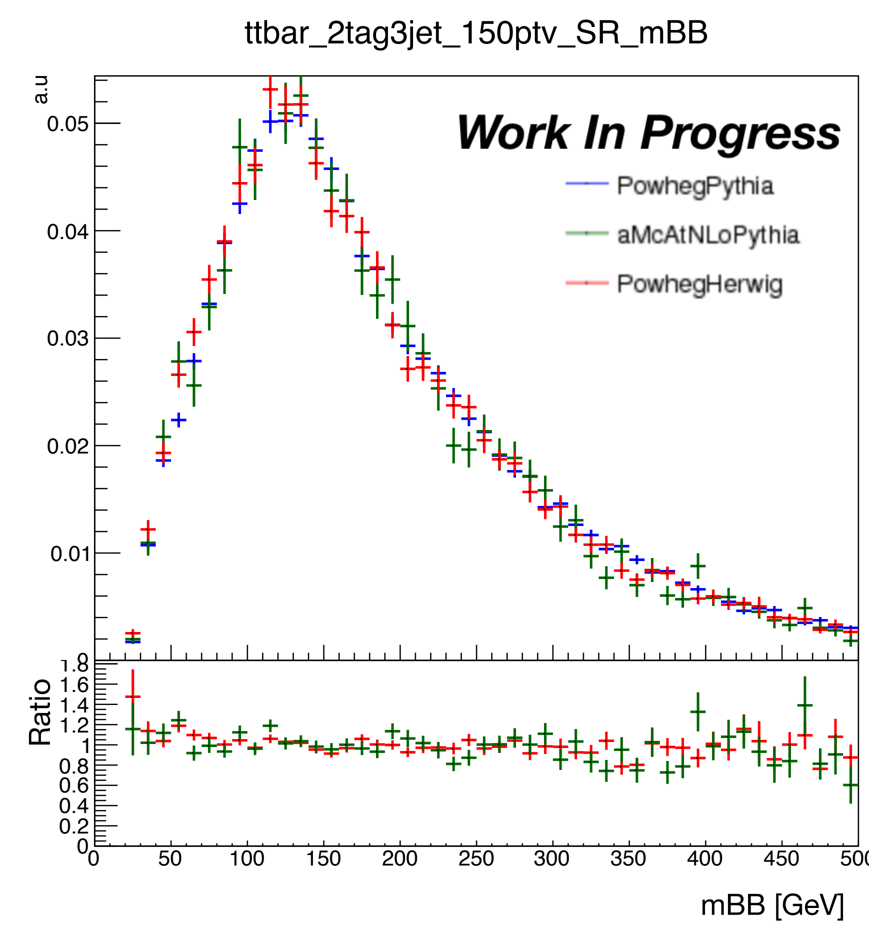
Signal contribution x100

- The analysis of this channel is not easy because s/b ratio is small (few percents) even with the mva method

Background modelling systematics



- MC samples are used to model background events for the analysis
- Systematic uncertainties are assigned to these predictions
- Comparing the nominal generator to another generator with different Matrix Element or Parton Shower (considered as the variation)
- The usual method is to compare bin-by-bin to all possible variation on the final discriminant of the analysis
- Does not apply in our case due to lack of MC statistics to see effect at the percent level (and we need need systematics at few percent level because s/b is low)
- The number of events passing all cuts and selections in the alternative sample is very low compared to the nominal



Background modelling systematics

- An example of the number of events in each generator of ttbar samples in 2tag 2/3jets region:

Generator	2tag 3jets	2tag 2ets
PowhegPythia	63590	8666
aMCAtNLoPythia	23989	3319
PowhegHerwig	25631	2552

- ~3 times less events in alternative samples compared to nominal. Statistics decent but not enough to access variations at the percent level

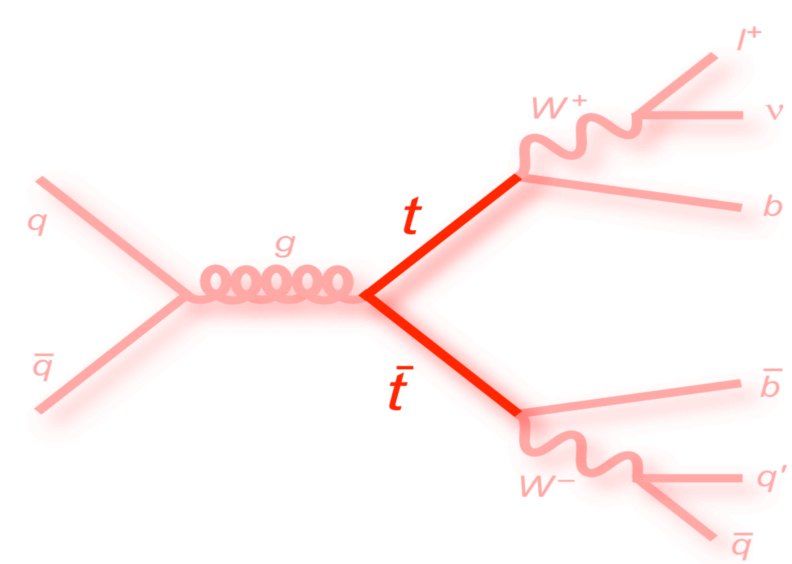
- **Can't we just reproduce more MC events?**

  - Nominal samples have already events of the order of 100millions: the acceptance of VHbb analysis is low with respect to the full ttbar phase space

  - Expensive even for ATLAS to reconstruct as many events for alternative generators (both in disk space and CPU)

Background modelling systematics

- **Solution?**
    - Look at difference between generators at truth (-particle) level, where we can have much more MC statistics for all generators
    - Parametrise the ratios alternative/nominal
    - This gives event weights, which are then applied on the nominal samples at reconstruction level
    - The use of event weights smooths the impact of MC statistical fluctuations
- The weights are obtained from kinematic variables distributions considered important or that show a difference between generators
- The weights are derived to each of the background samples independently and applied to the corresponding nominal generator
- Study presented here focuses on ttbar 1 lepton channel

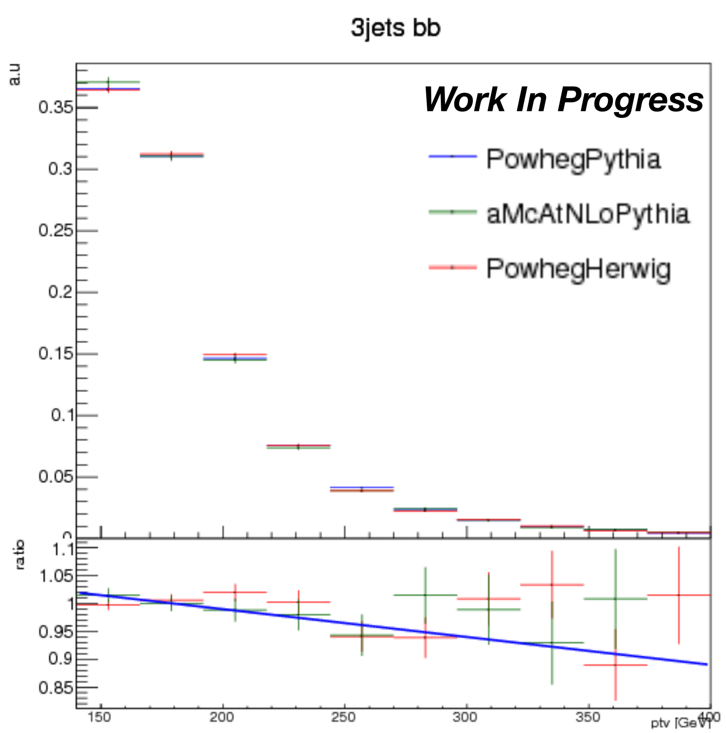


Systematics in ttbar 1lep channel

- The systematics for ttbar modelling are derived from pTV and mBB distribution (for being the most important variables for the analysis)

- These systematics come from comparing mBB and pTV distribution between PowhegPythia (the nominal generator for the analysis) and aMCAtNLoPythia (alternative generator)

- The weights are computed as follows: ratio of the distribution (mBB and pTV separately) is fitted to get the weighting function



- This method is proven reliable in computing systematics

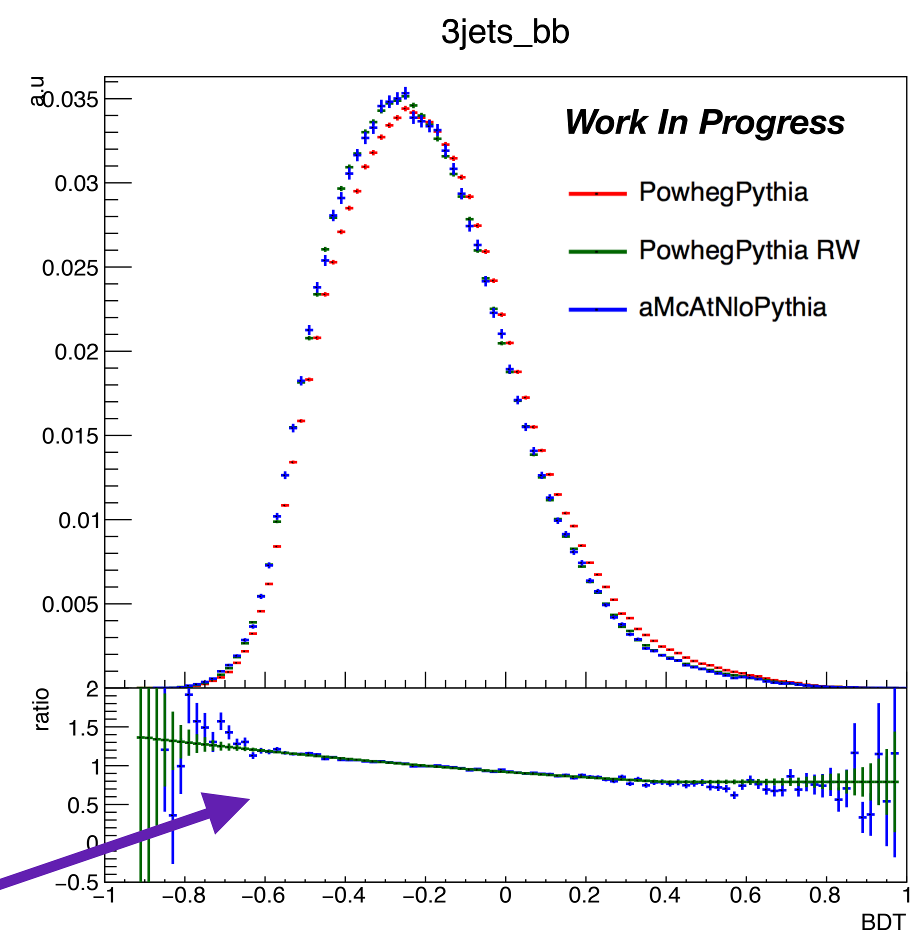
- Difficulties: need to look at all kinematic variables to check for non closure

- A new method is proposed to not focus on two variables but rather use many variables to represent the whole phase space

- New method: BDT to use one variable “BDT score” (instead of two) in assigning systematics

How to derive BDT Based Systematics

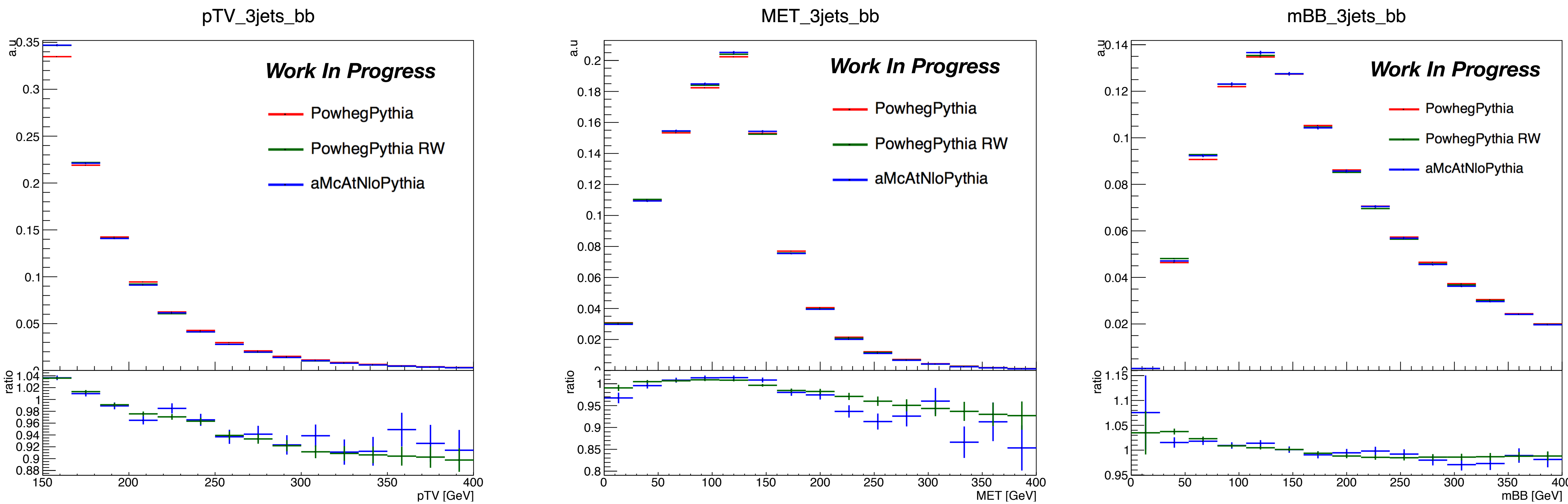
- 11 kinematic variables (at truth level) are used to construct the BDT
- The dedicated BDT is trained by taking PowhegPythia as signal input and aMCAtNLoPythia as background input
- Want also to introduce a systematics related to the PS generator -> Also train a BDT to compare PowhegPythia and PowhegHerwig
- Different categories of events are treated separately
- The BDT score (is a value between -1 and 1) is assigned to each event by being more signal like or background like
- The ratio of the BDT distributions of the two generators is then computed for ME and PS comparison



The fitting function allows to get the weights

First application of the weights

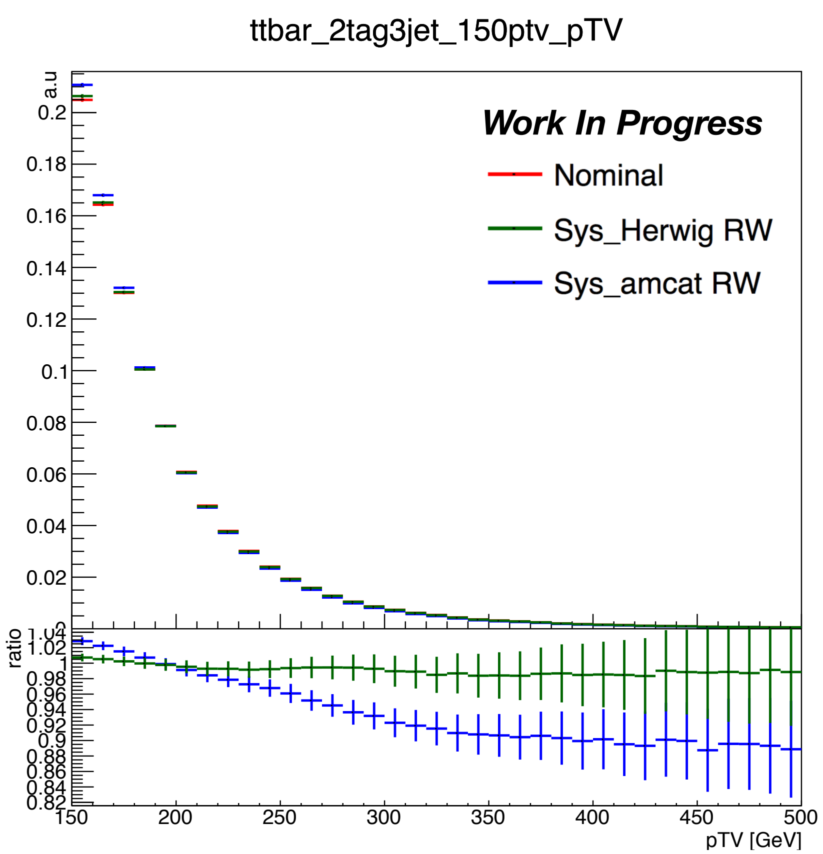
- First test for the new method is to apply the weights on the variables used for the BDT training
- The weights are applied on the nominal sample to morph it into the alternative generator

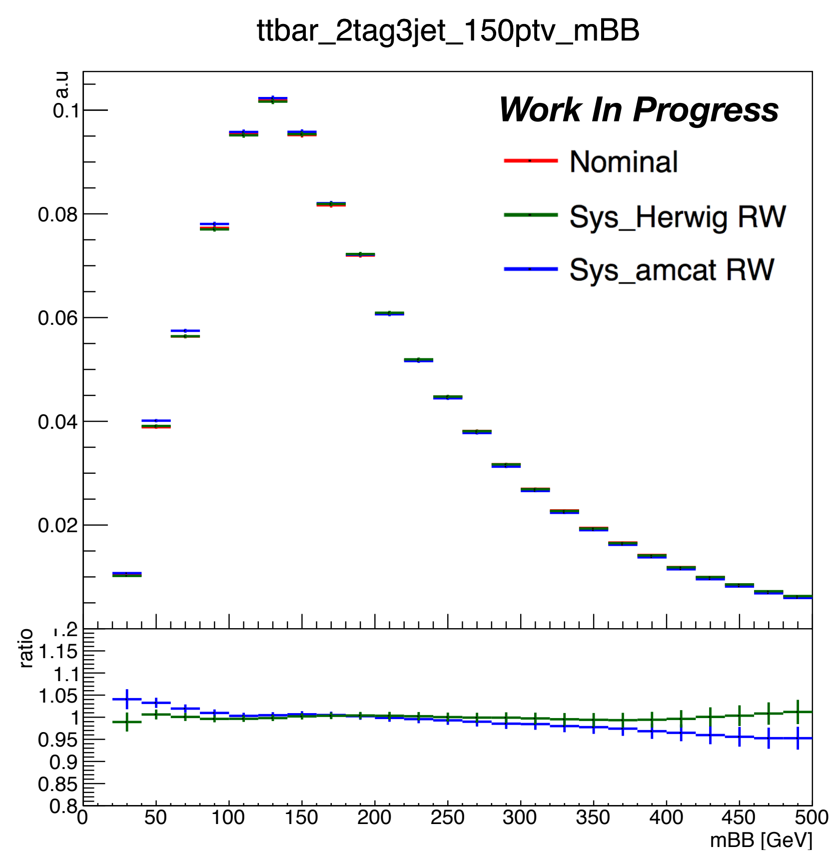


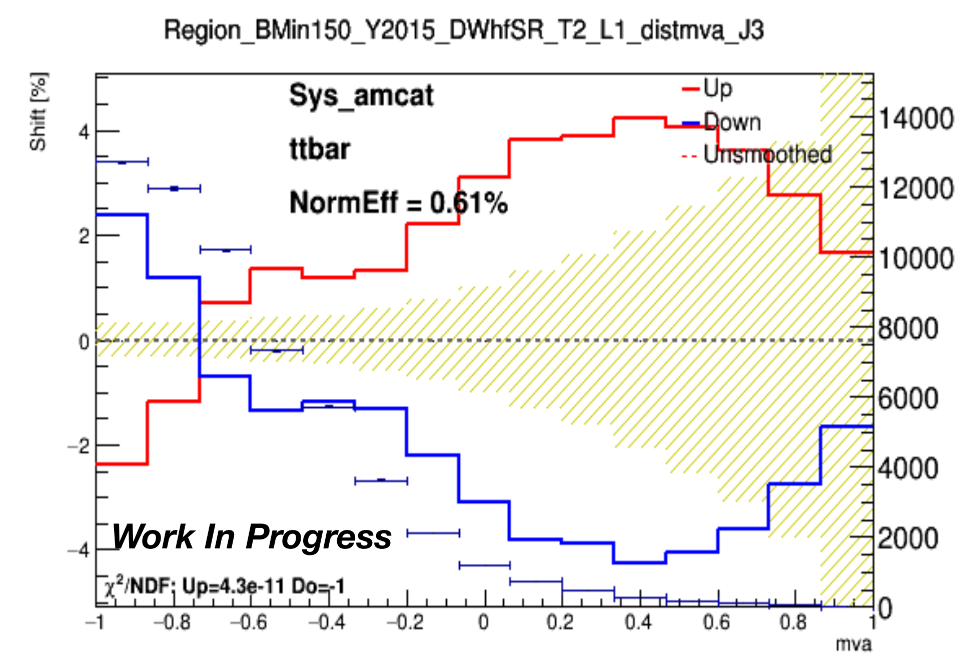
- The reweighting does a pretty good job in transforming the nominal into the alternative
- The effect is seen on many of the variable distributions
- Reweighting was also applied to test PowhegPythia vs PowhegHerwig reweighting

Reweighting applied on reconstructed events

- Reweighting validated at truth level -> Need to apply to events after ATLAS reconstruction
- The BDT that was trained at truth level was implemented at reco level
- To be coherent with the training -> Access the truth information of the reconstructed events
- The truth variables are then passed to the BDT to compute the weights







- The impact of both Sys_Herwig and Sys_amcat can be seen on all variables
- Not able to compare to alternative samples at reco level -> Not yet available

Performance of the new Systematics

- Two models of ttbar-modelling systematics were compared: the pTV and mBB reweighting vs Sys_Herwig and Sys_amcat reweighting

- Method : investigate the contribution of the systematics to the signal strength uncertainty after performing an Asimov fit (μ=1)

	Sys_Herwig & Sys_amcat	TTbarPTV &TTbarMBB
Total	0.44	0.44
DataStat	0.27	0.27
FullSyst	0.35	0.35
Jets MET	0.06	0.06
BTag	0.16	0.13
Leptons	0.0065	0.0070
Luminosity	0.02	0.02
Model ttbar	0.05	0.07
MC stat	0.12	0.12

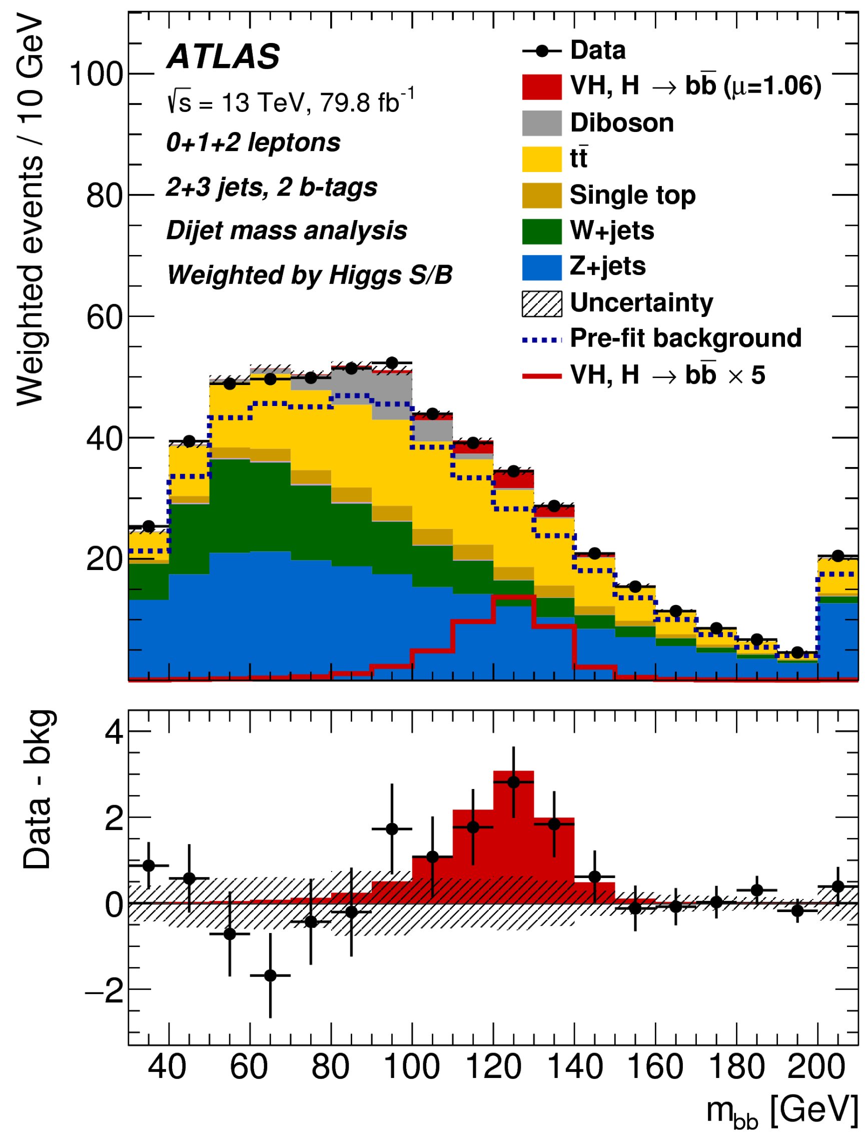
- The impact on the significance of the ttbar model is reduced when Sys_Herwig and Sys_amcat are introduced

- However the impact of the total uncertainties is the same

- The impact of the new systematics seems reasonable and fit to data behaves well

Conclusion

- VH production and H→bb decay were observed with the ATLAS detector with Run2 data
- A new method to compute systematic uncertainties is being studied for VHbb analysis
- The new method uses BDTs to parametrise the difference between two generators
- It is a new method using one variable to represent all the phase space
- So far the results show promising
- The study will be propagated to other backgrounds and other channels
- Hopefully we will have a more precise estimation of our background modelling systematics for the analysis of the full Run2 data in 2019



Thank you !

Fun fact: the biggest oyster in the world is the size of a man's shoe