trying to minimise
my ATLAS bias

H->bb (cc) measurements (searches) at the LHC

Valerio Dao *(CERN)*

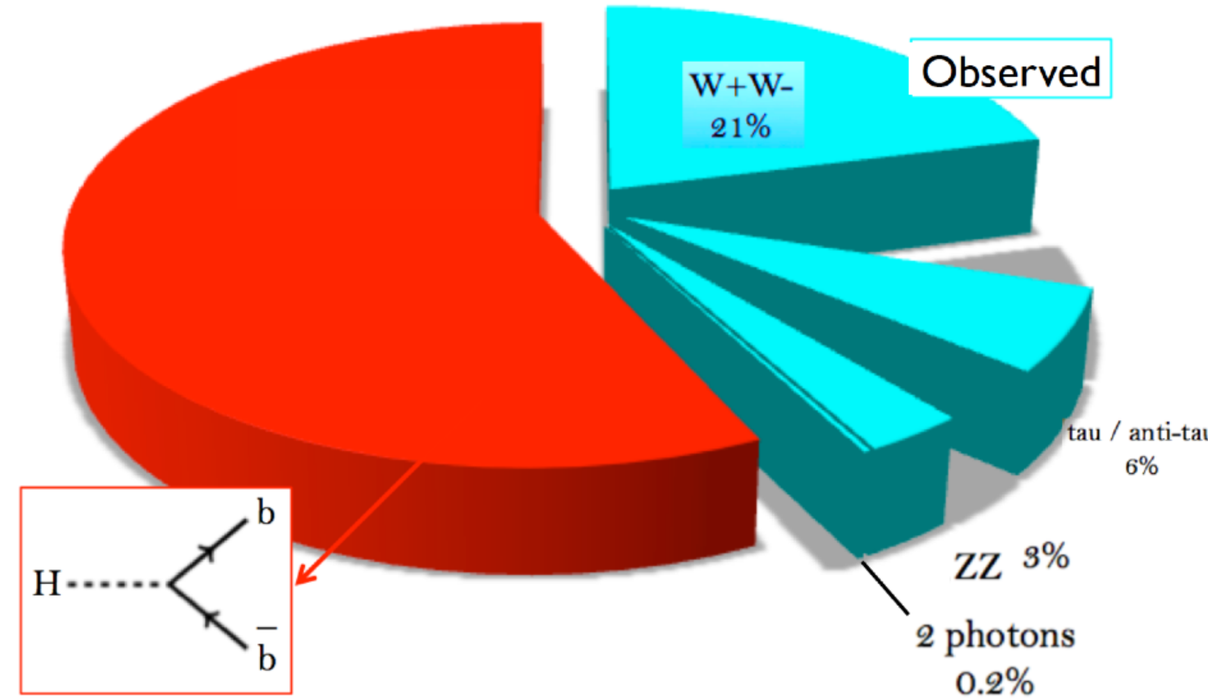
HiggsDiscovery@10 symposium
Birmingham - June 2022

ATLAS EXPERIMENT

un: 338349
vent: 616525246
017-10-16 20:24:46 CEST

A different story: this talk will mainly covers LHC Run 2

- ***Higgs decays snapshot after LHC Run 1:***
  - observation of the Higgs boson driven by channels with much lower branching ratio
  - H→bb is the largest decay and his knowledge is also important to understand contribution to the width
  - **…. so why wasn't it observed earlier?**

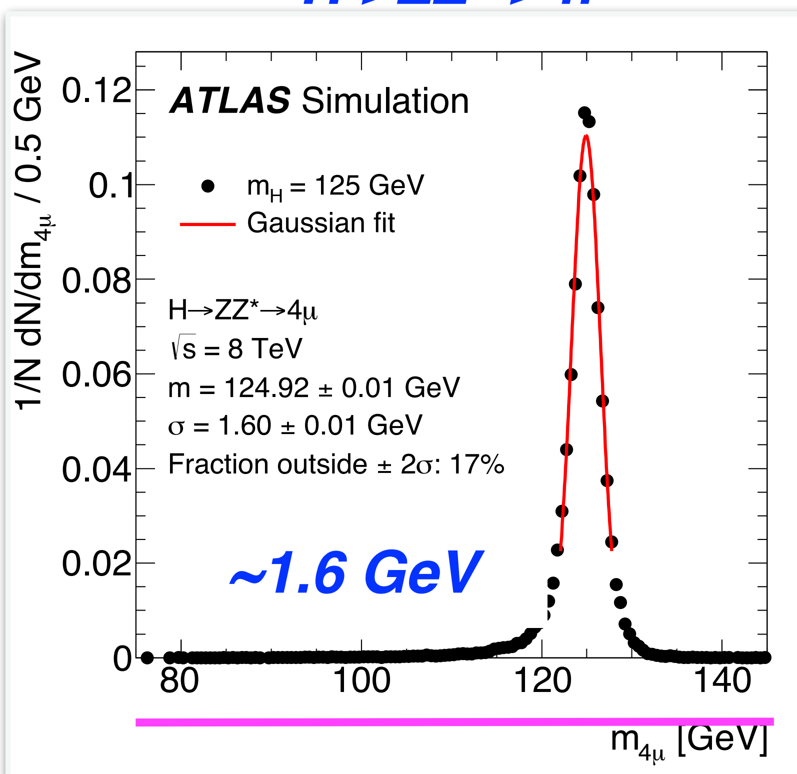


- For a "fair" ATLAS-CMS story I had to restrict myself to public plots

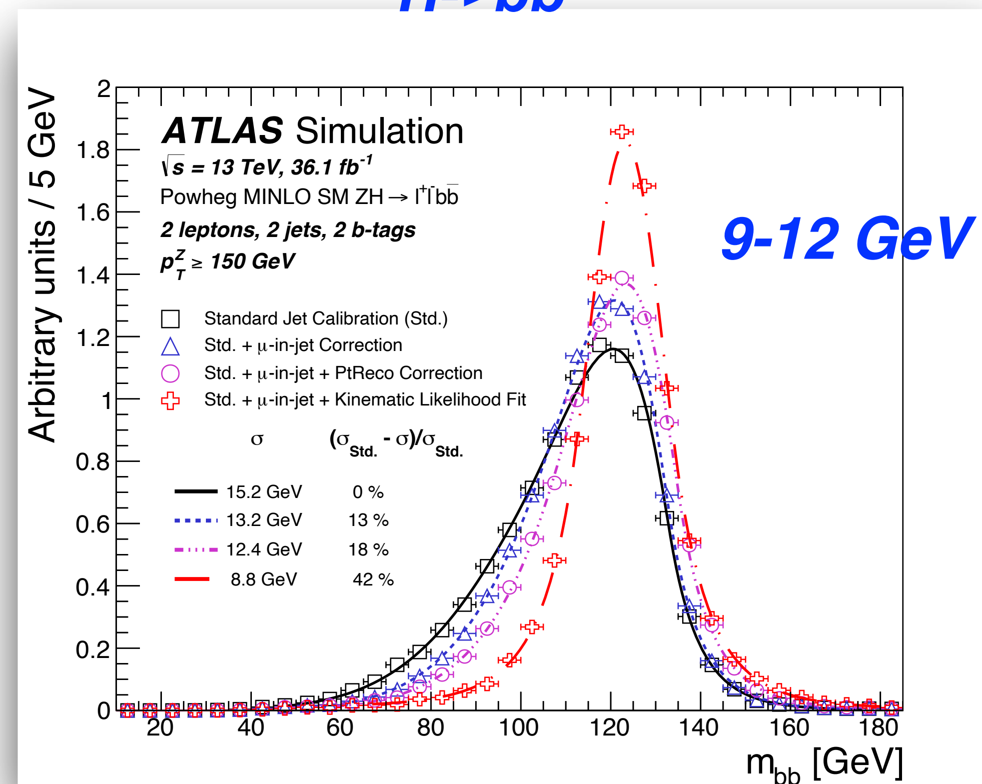
- **Curiosity**: only evidence made it to the CERN courier front page. See **Sept '18** for articles on observation

H->bb: challenges

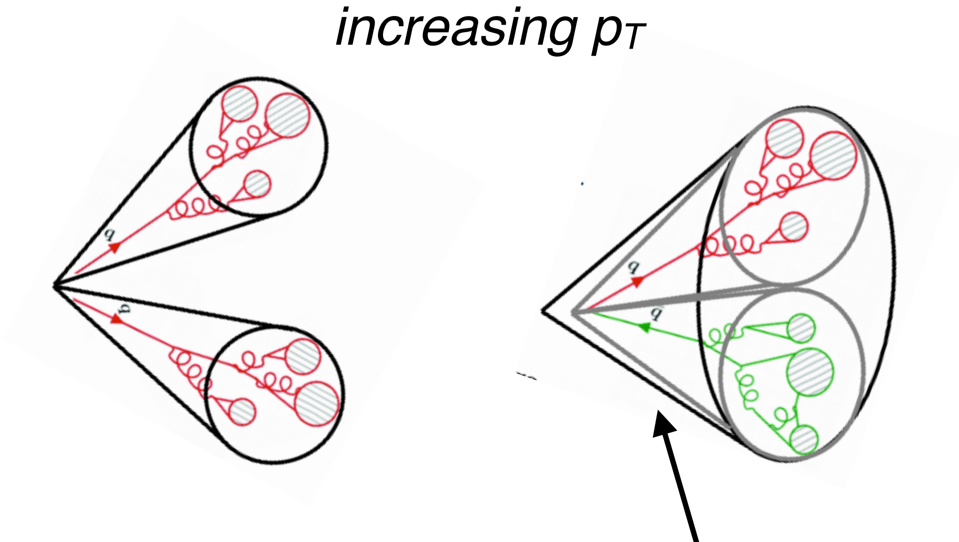
H->ZZ*->4l



H->bb



increasing p_T

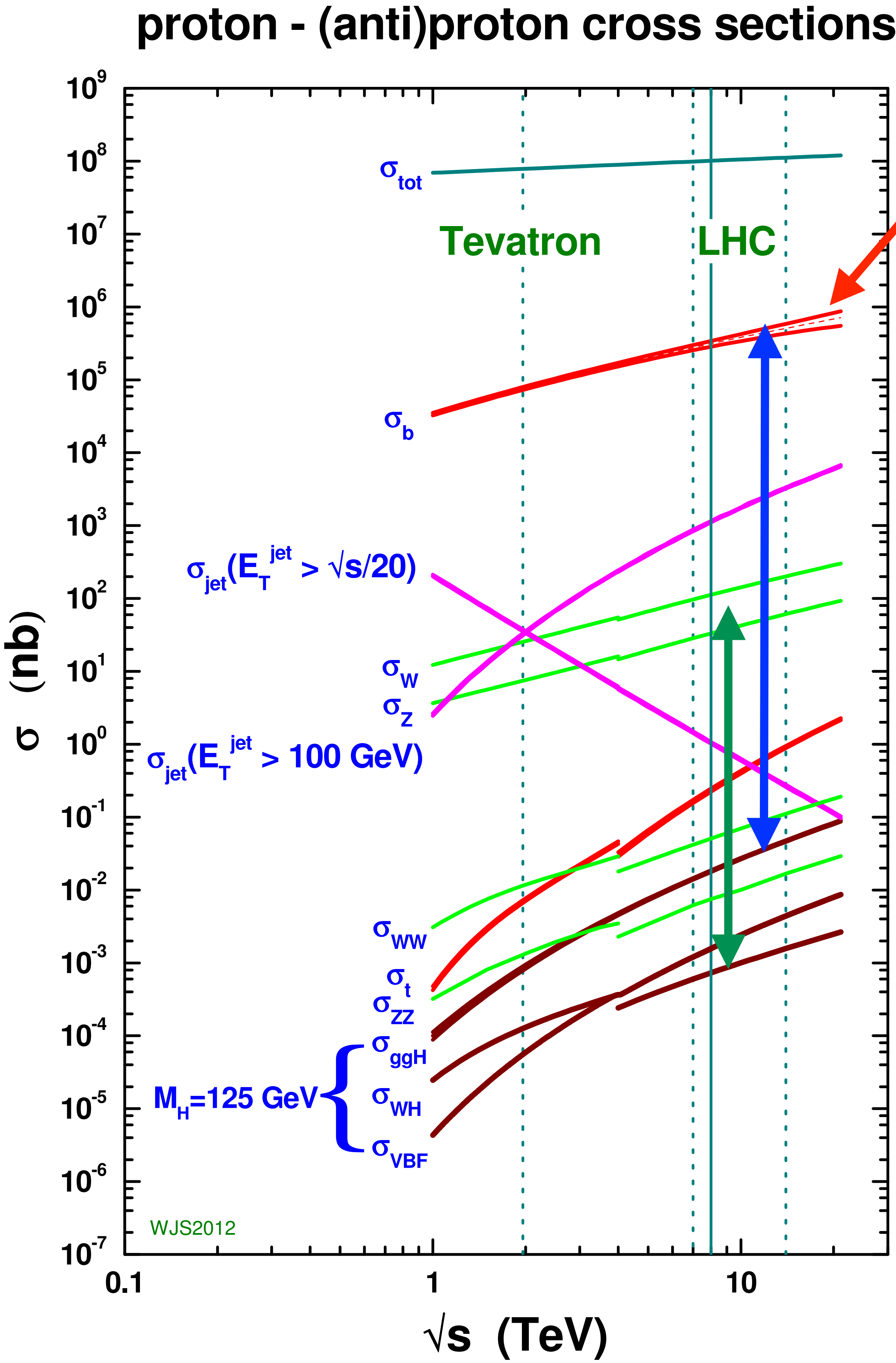


"Jet substructure as a new Higgs search channel at the LHC",
Phys. Rev. Lett. 100 (2008) 242001

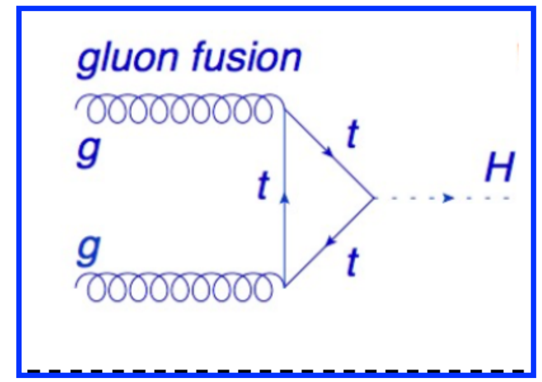
[in production 10 years later]

- ***Reconstructing H->bb decays:***
  - mainstream (historical) reconstruction uses 2 separated R=0.4 jets
  - reconstruction in a single jet is becoming more and more popular
  - in both cases **mass resolution is ~ 12-15 GeV ….compares to 1.5-2 GeV for ZZ->4l and yy**

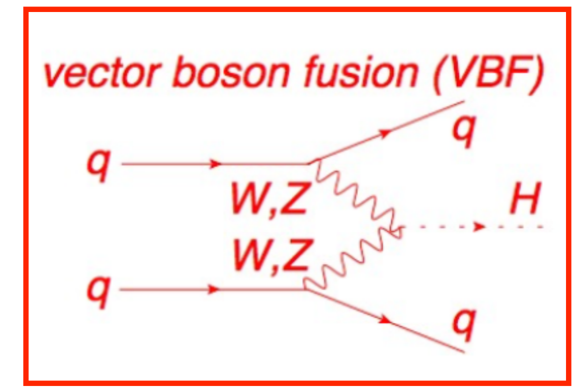
H->bb: challenges



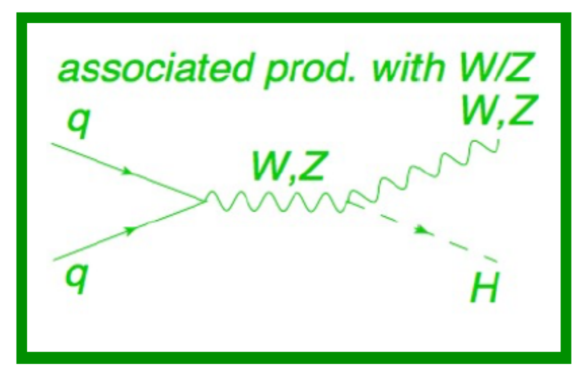
Overwhelming background



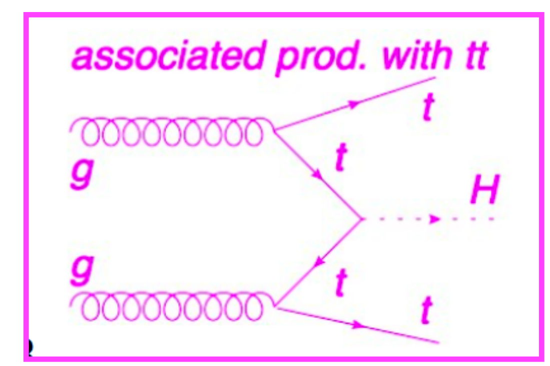
88%



7%

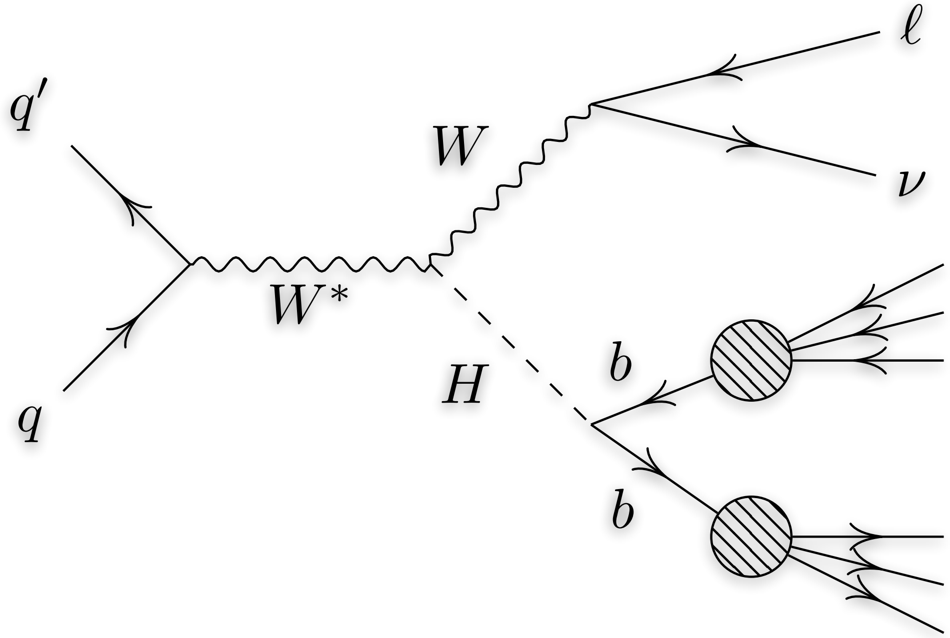


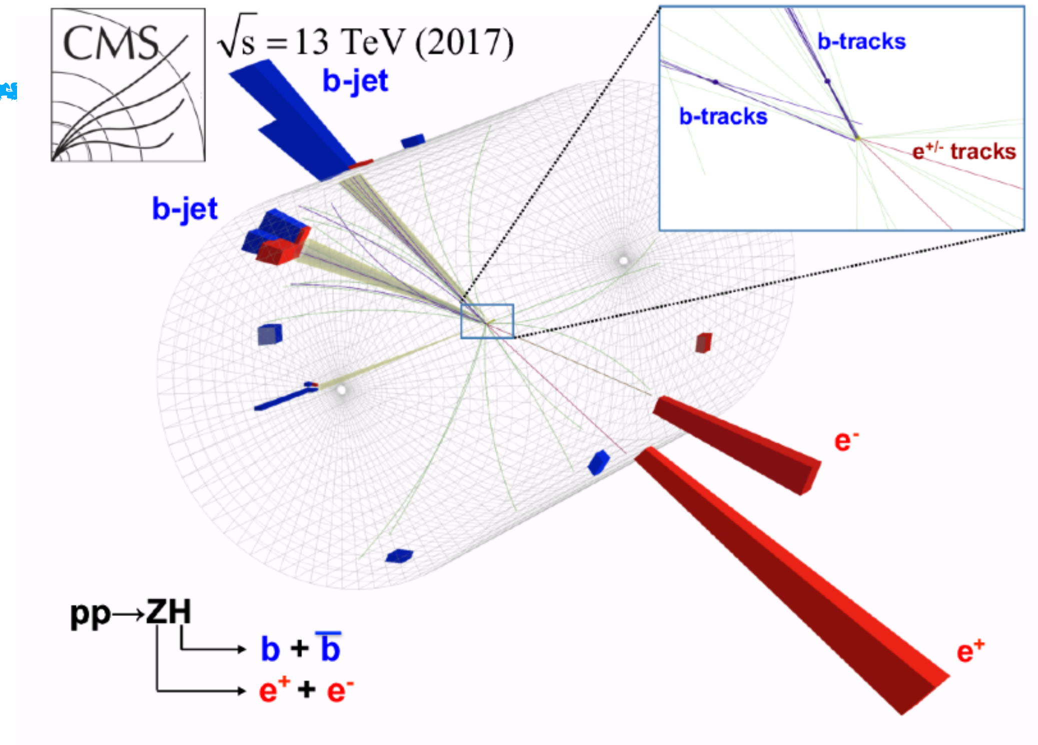
4%



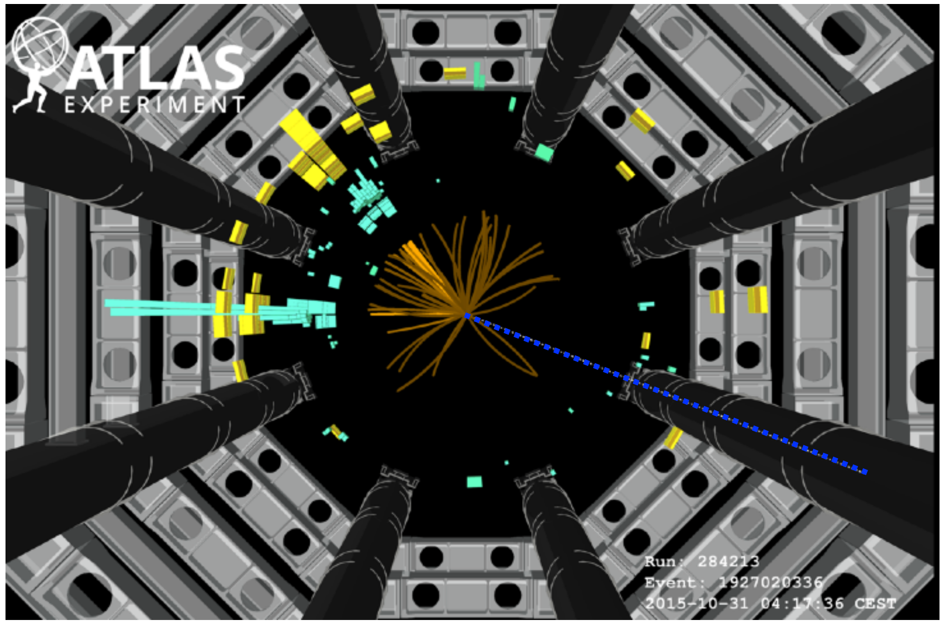
1%

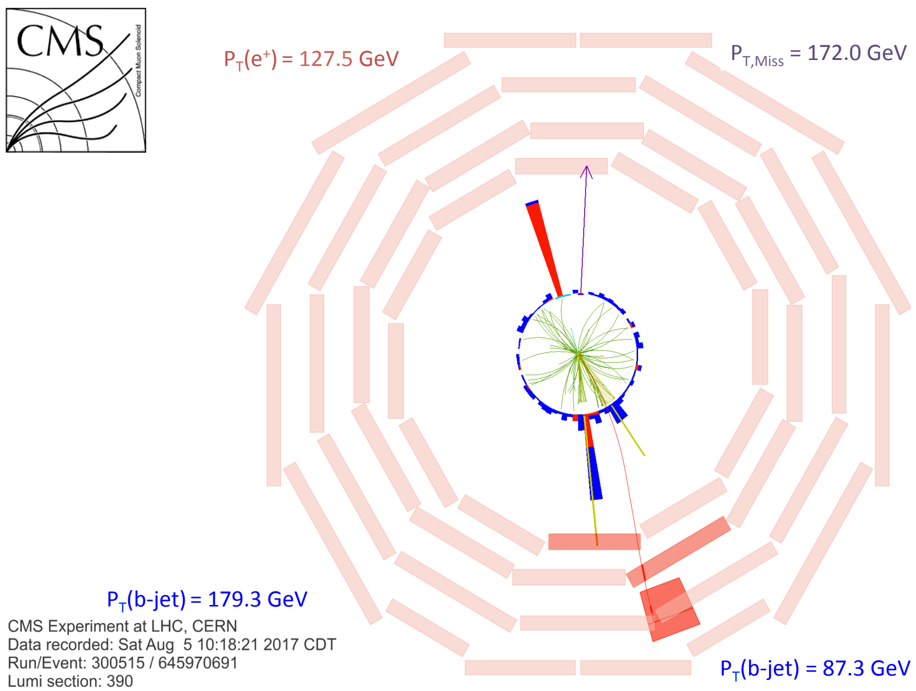
- **VH production** is the ***golden channel*** at hadron colliders. Leptonic V decays helps with:
  - triggering, background rejection , additional handles in signal extractions
- … but you loose factor 20 in XS … **and S/B no way near the one of H->ZZ**

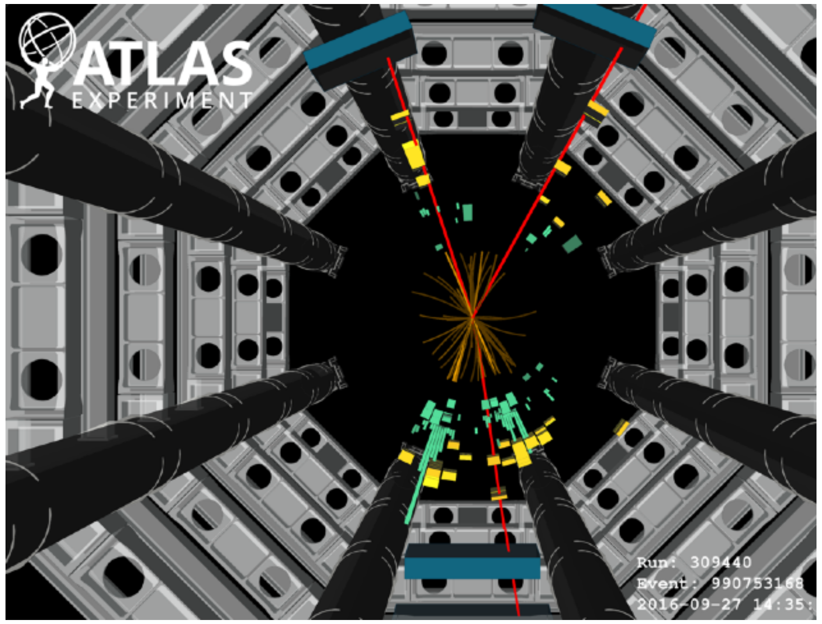




VH H->bb





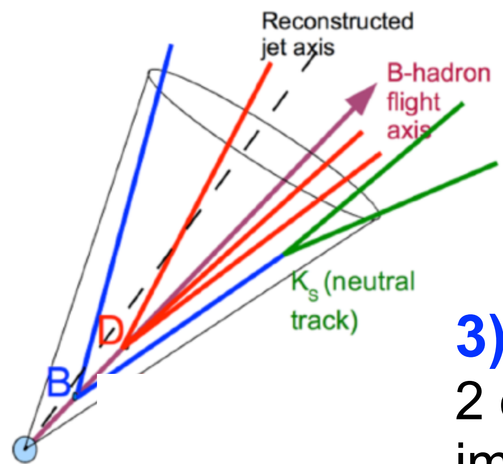


0-lepton ***1-lepton (l=e,mu)*** ***2-lepton (l=e,mu)***

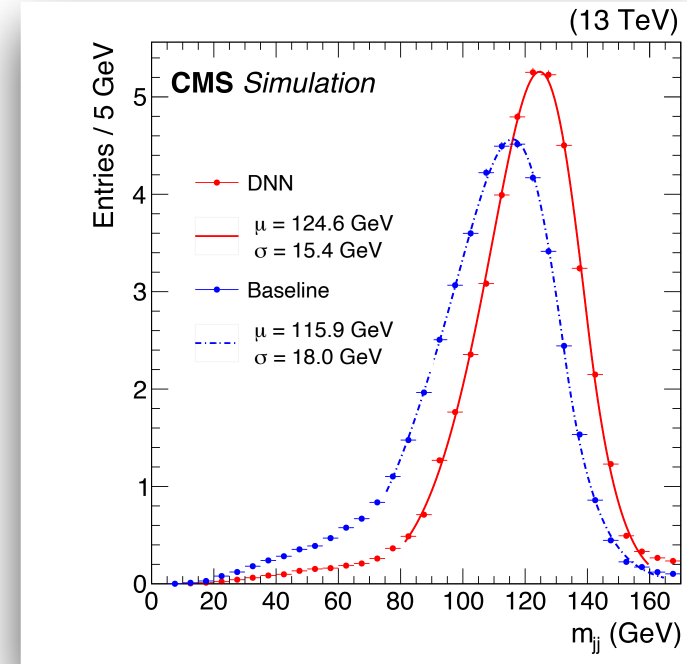
VH H->bb: ingredients*

(*) many in common to other analyses

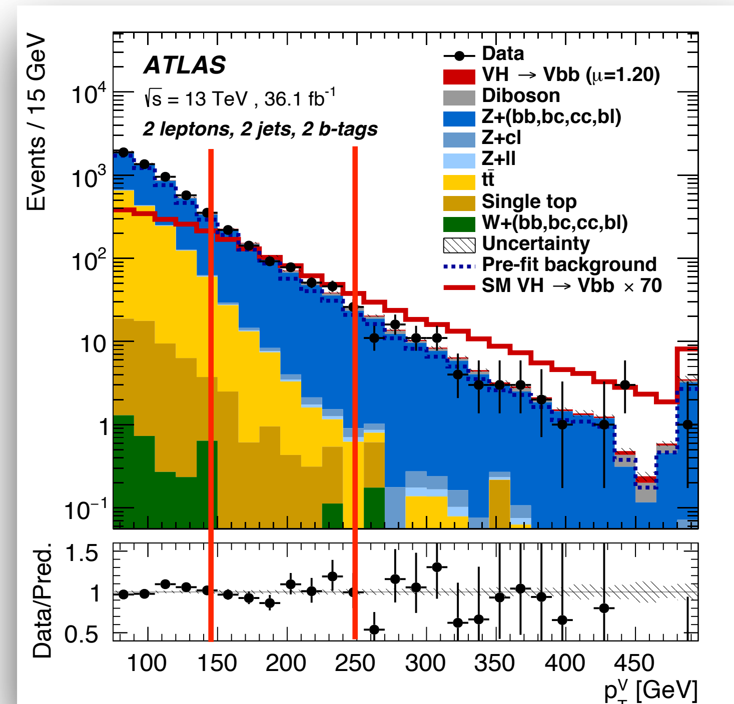
1) b-tagging: very sophisticated algorithms, reaching <1% fake rate with >70% b-efficiency



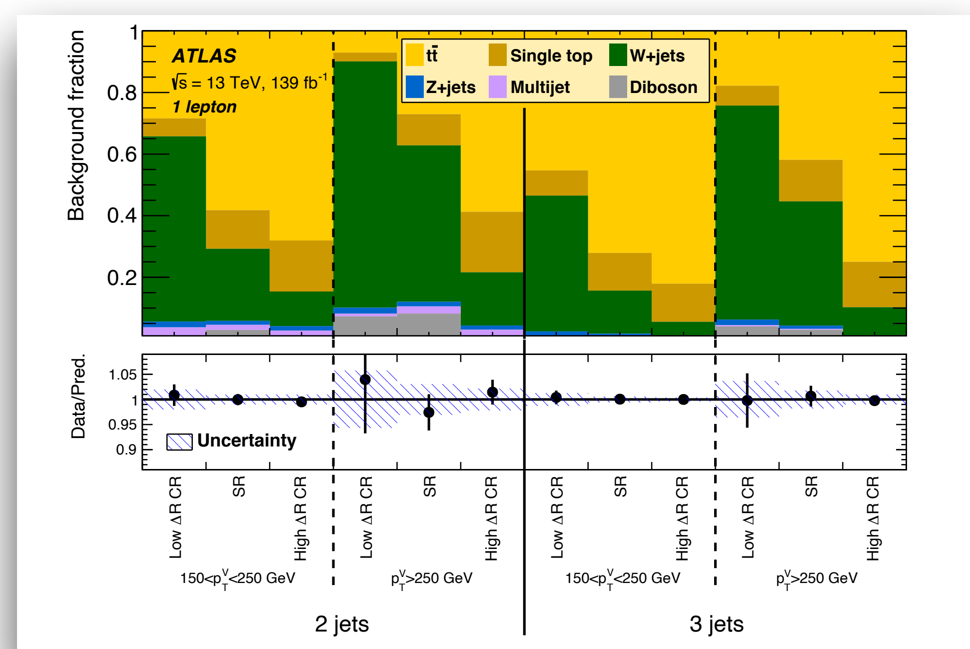
2) reconstruction: additional correction to improve the mass resolution



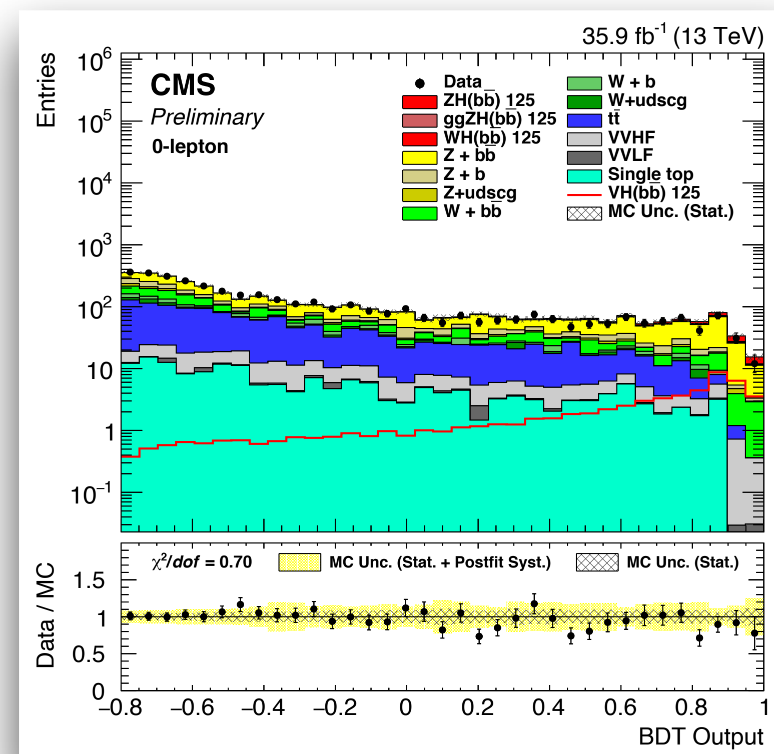
3) event selection: 2 or 3 jets, high pT improves S/B



4) event categorisation: dedicated CR to target leading backgrounds



5) event selection: many MVA discriminants for signal extraction (careful validation of input variables)

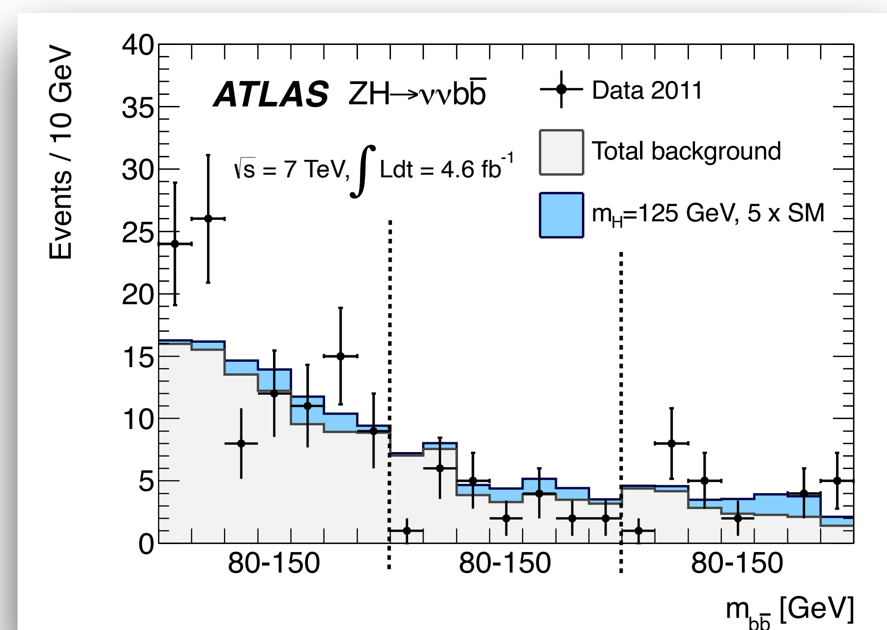


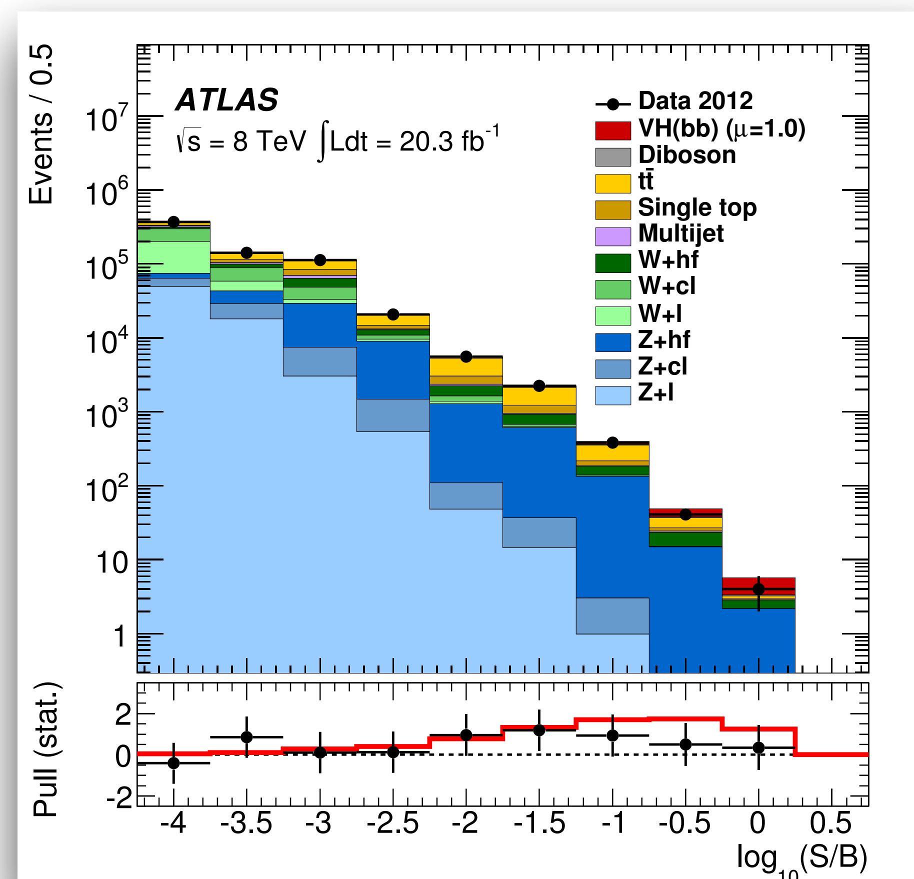
- **Caveat**: I was not involved in VHbb in Run1 (I was doing ttHbb … which is another story all-together)

from yesterday's talk

- **Analysis approaches created at the Tevatron live on today in LHC searches and Higgs measurements**

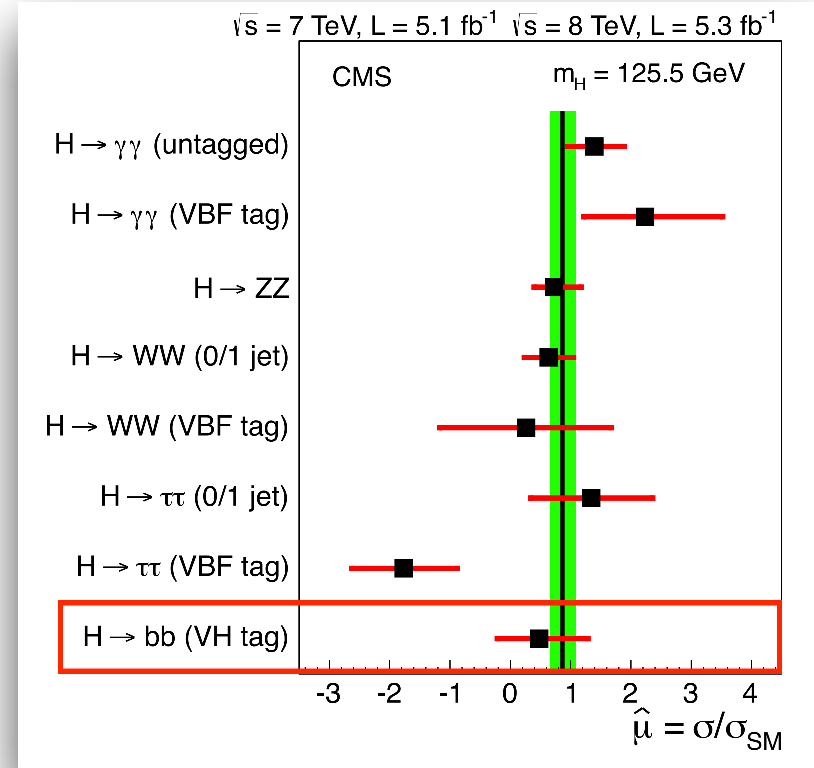
- **Run1 analyses heavily inspired by Tevatron's searches and put in place many aspects that we now consider trivial / for granted:**
  - lepton channels, advanced usage of b-tagging
  - correction to m_{bb}, many MVAs
  - etc

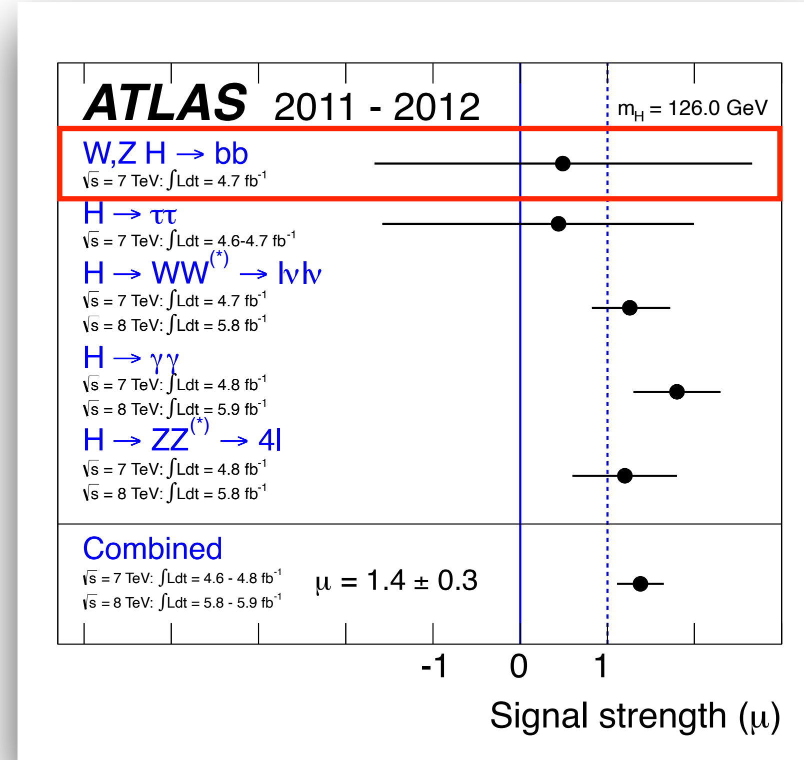




5 orders of magnitude

VH H->bb: Run I





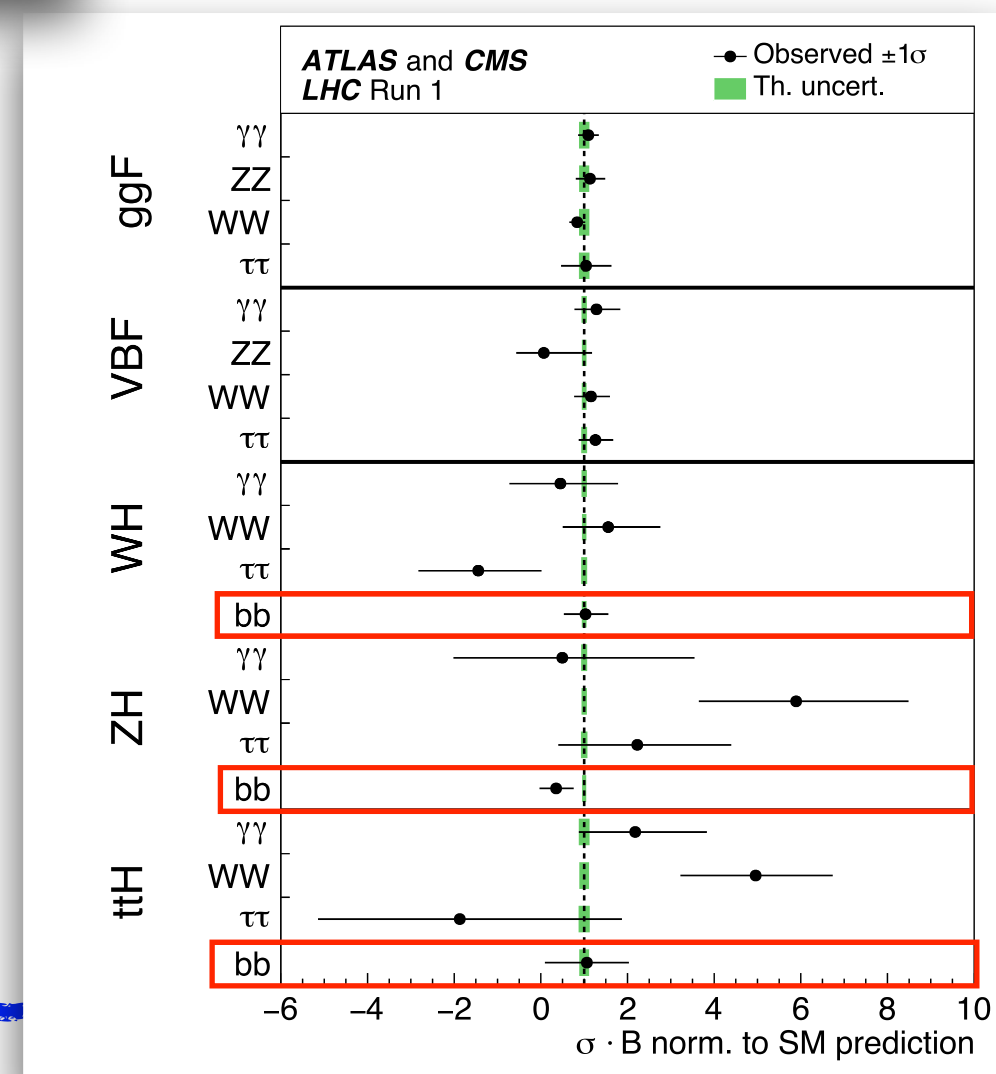
- *July 2012*: Clearly far from the goal to participate to the party

- *Full Run1 result*: ”we could have got evidence …. but we didn’t”

Run1 H->bb significance:

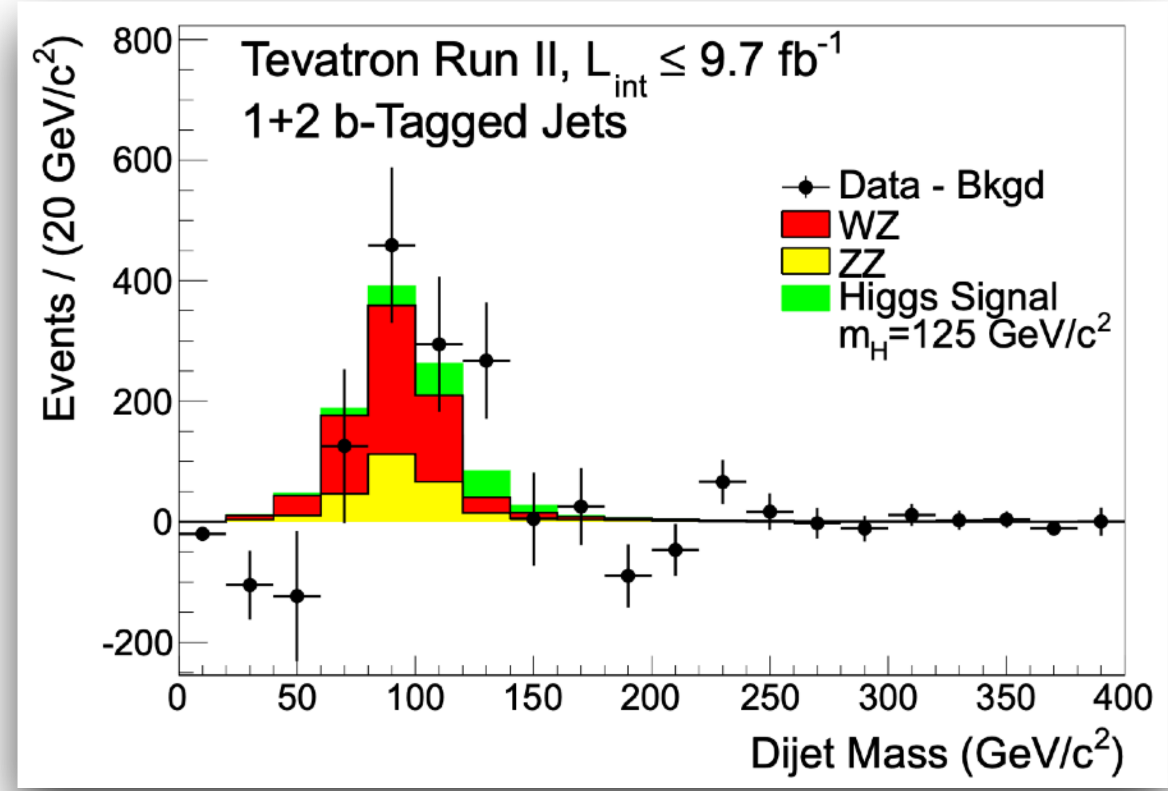
2.6 σ obs. , *3.7 σ exp.*

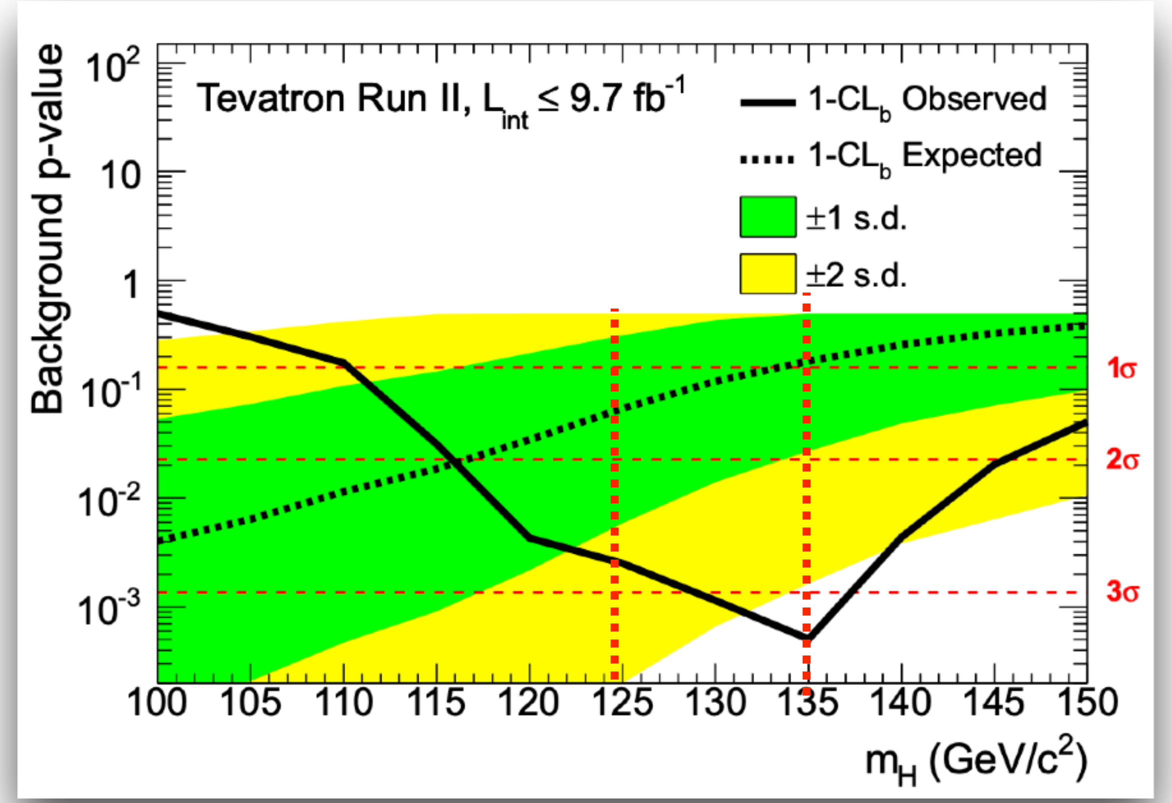
- Already very sophisticated analyses: many SRs/CRs, multiple MVAs



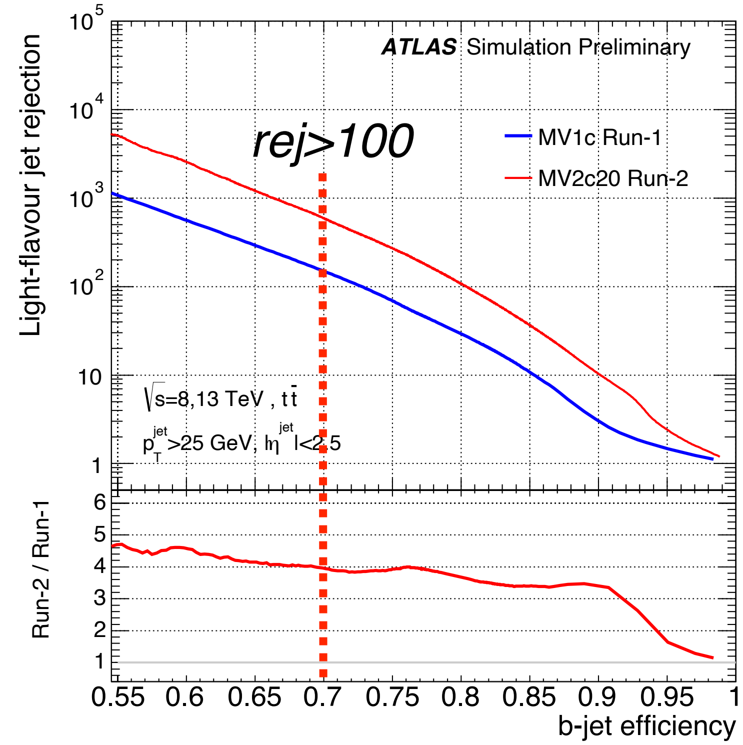
[Phys. Rev. Lett. 109, 071804 (2012)]

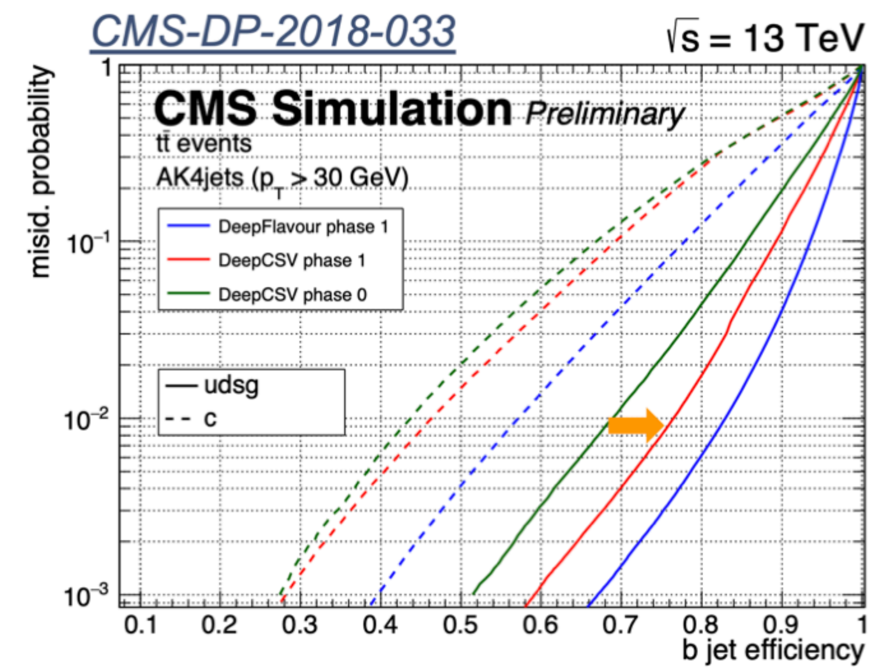
Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron

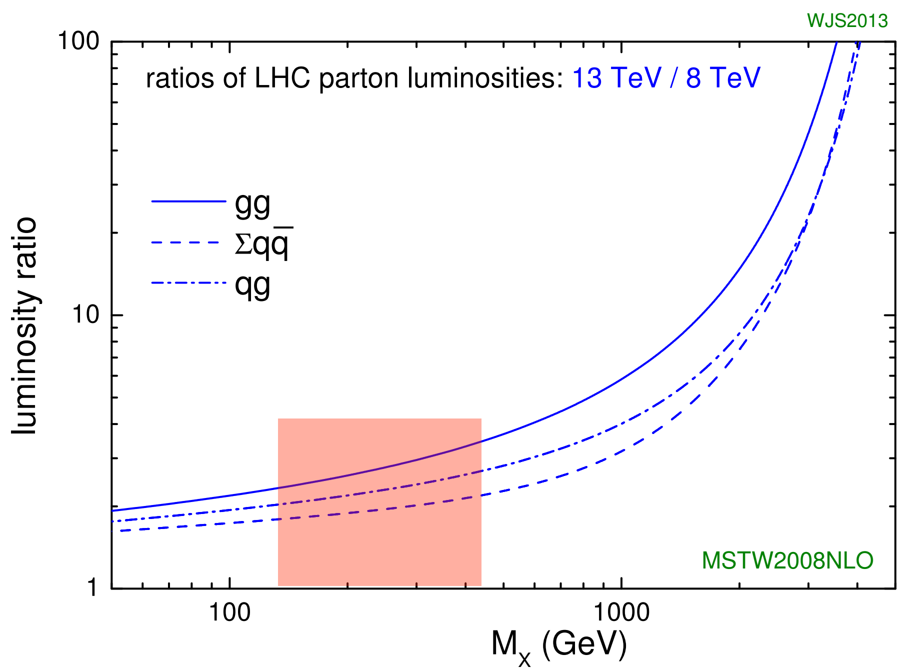




- Improved pixel detector: new layer (new detector) in ATLAS (CMS) from 2015 (2017):
  - massive boost to b-tagging perf.







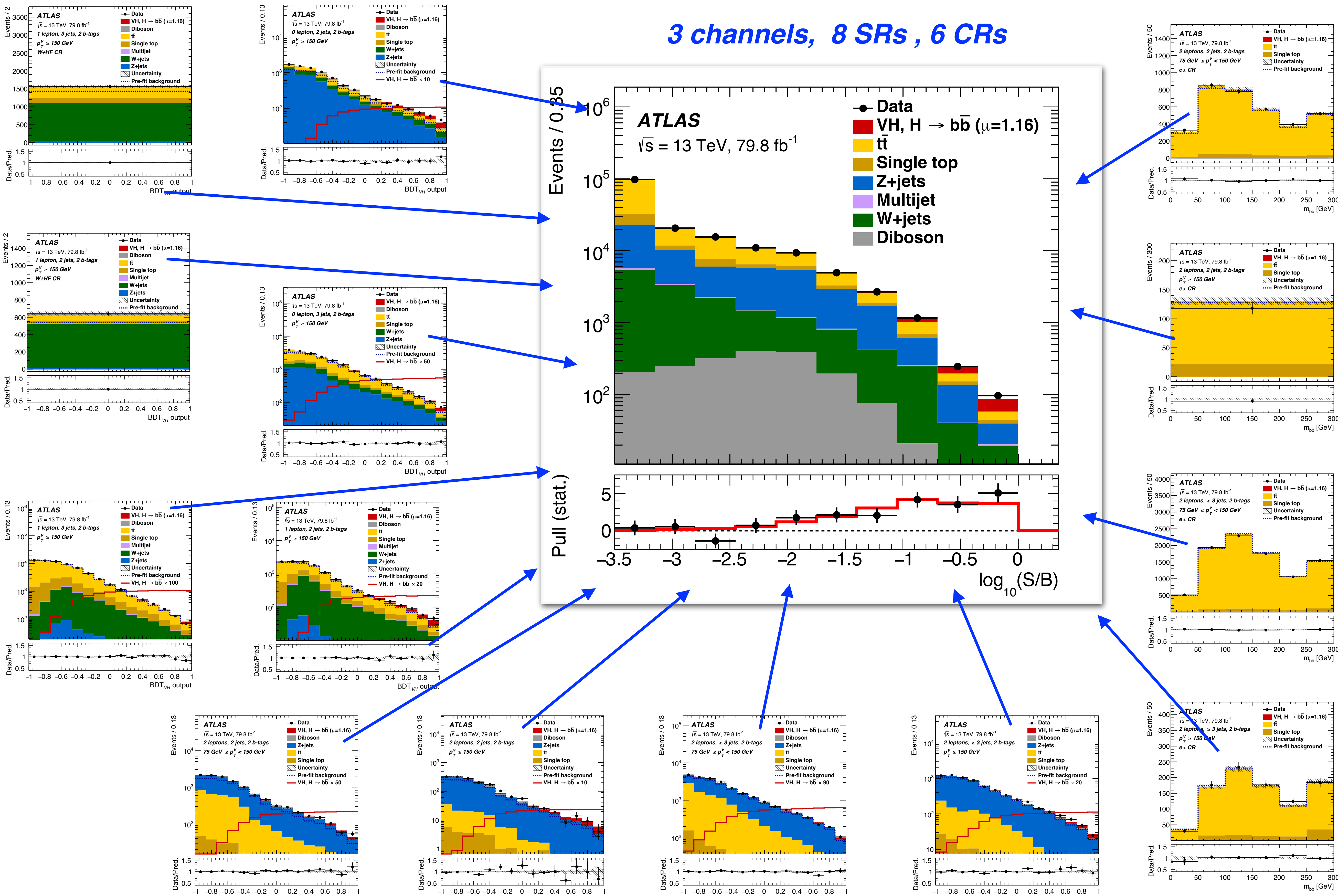
- Increase in centre-of-mass energy (8TeV ->13 TeV) does not always help … but certainly allows to collect events faster:
  - VH/VBF : x2 , ttH: x4
  - ttbar : x3.2

- **Time to re-think analysis strategy**: … with also some simplifications.
  - b-tagging really simplify your background composition
  - no need to include data too far from your SR … ease the understanding of very complicated fits

- ATLAS and CMS have slightly different approach but end sensitivity is very comparable

VHbb: not easy to explain and debug

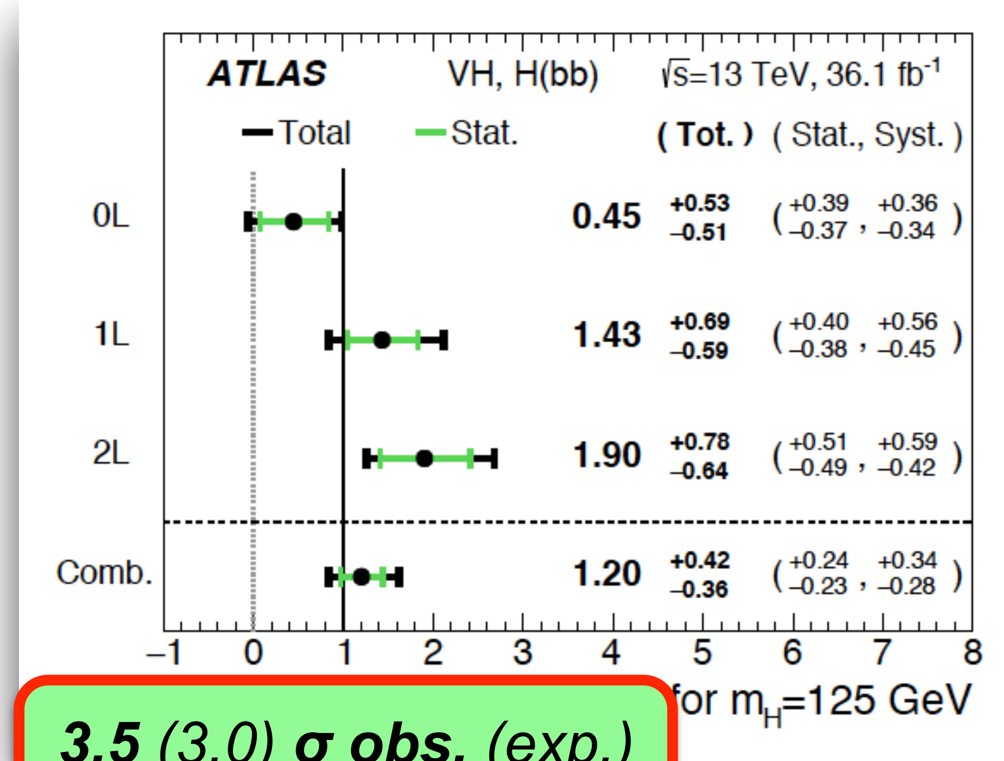
3 channels, 8 SRs , 6 CRs



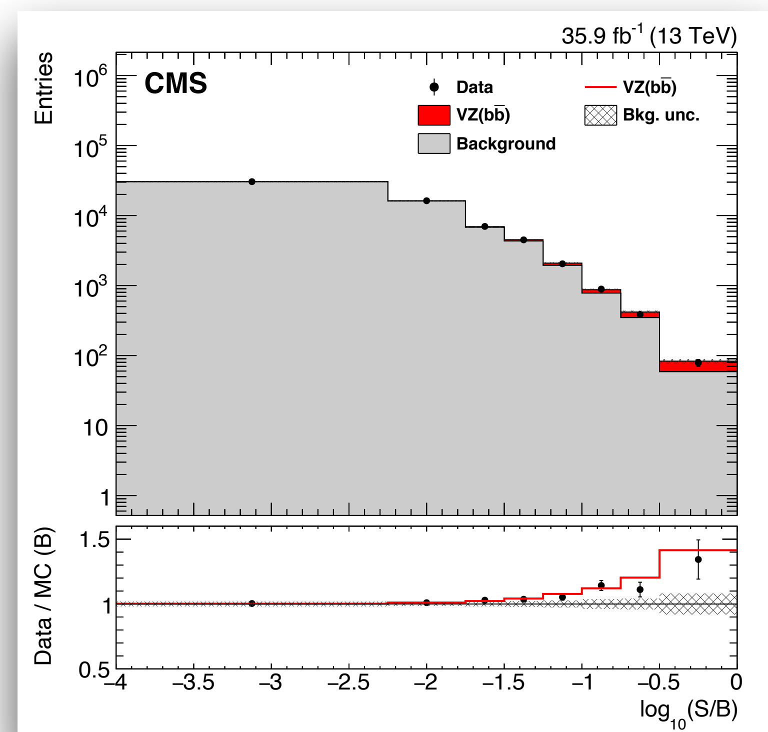
Run2 milestones: July 2017 (EPS)

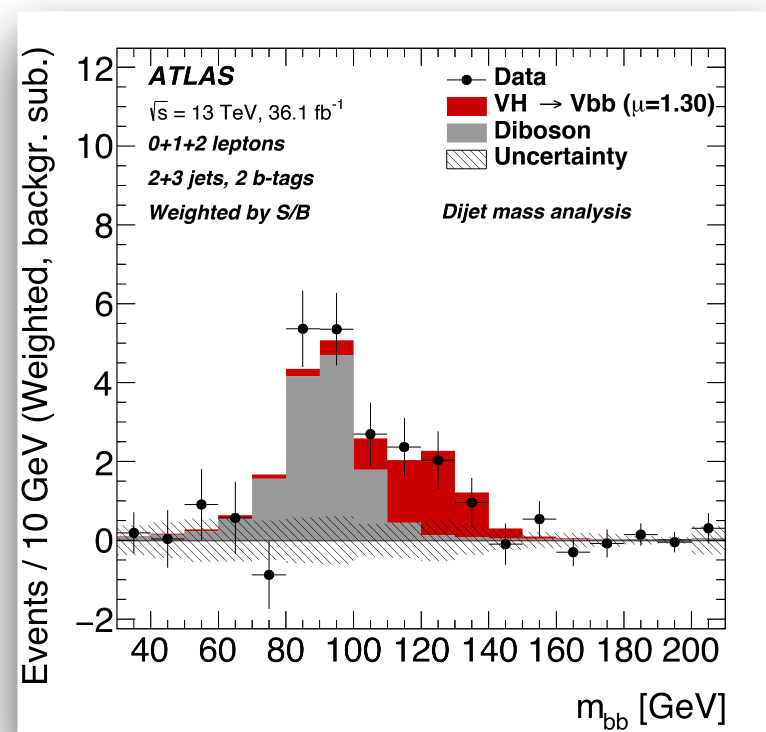
- **Evidence of VHbb** from both experiments using 2016 data

- **Many checks/results to support this claim**:
  - compatibility across channels, WH VS ZH measurements
  - *observation of VZ (Z->bb) process*
  - *mass fit without MVA*: typically ~30% worse sensitivity

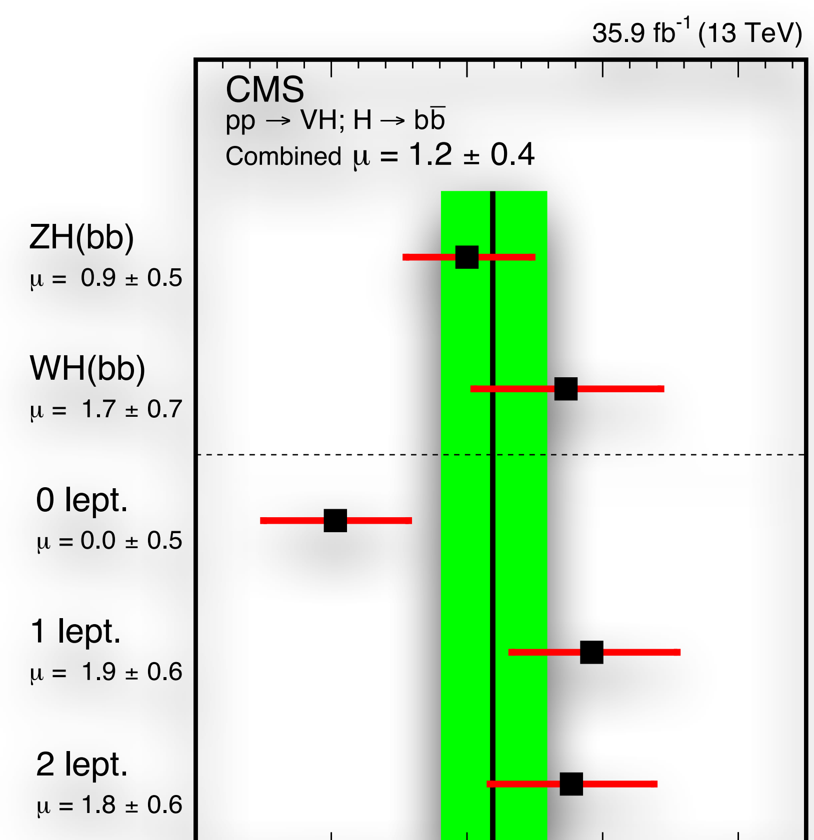


***3.5** (3.0) **σ obs.** (exp.)*



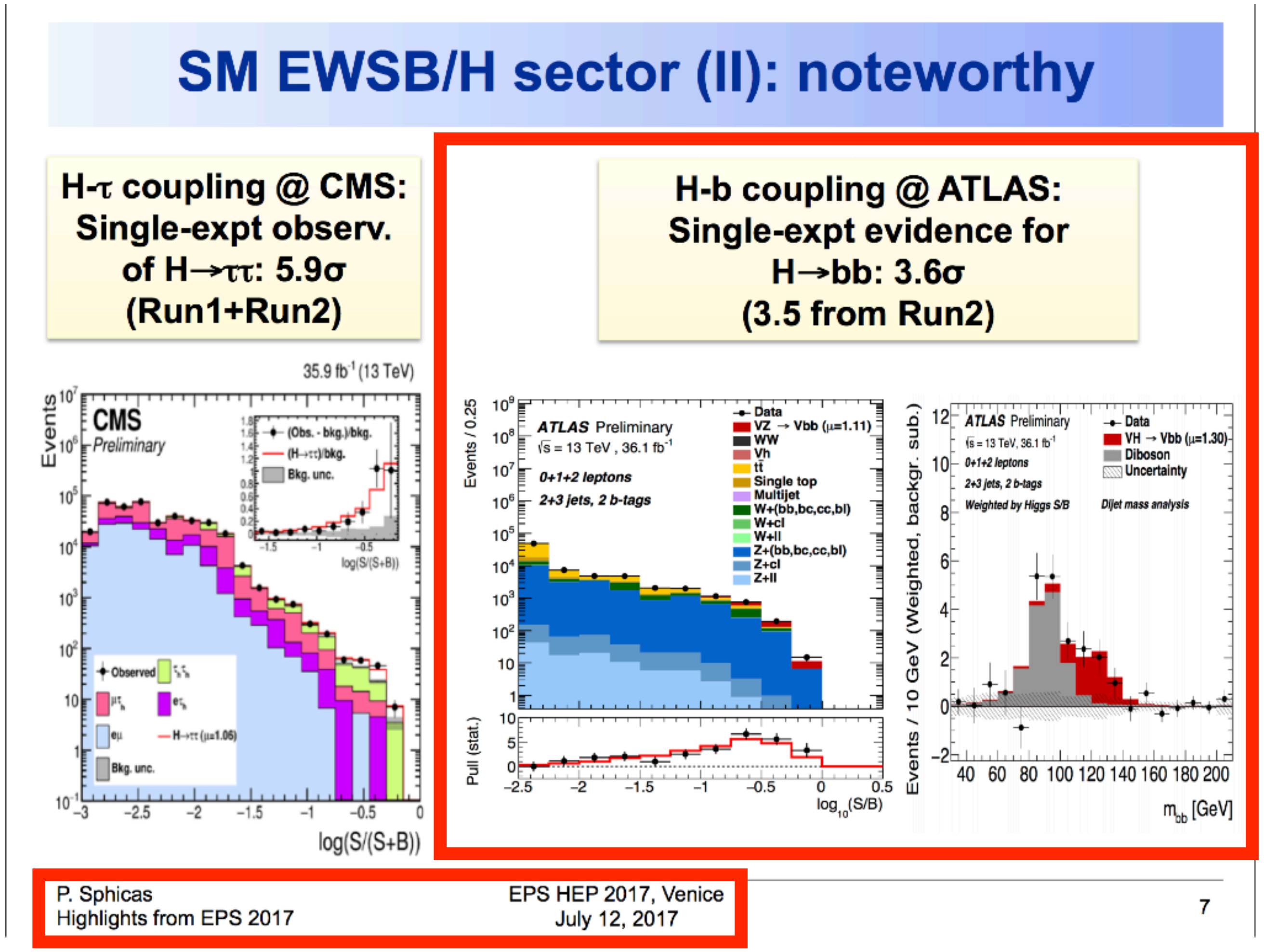


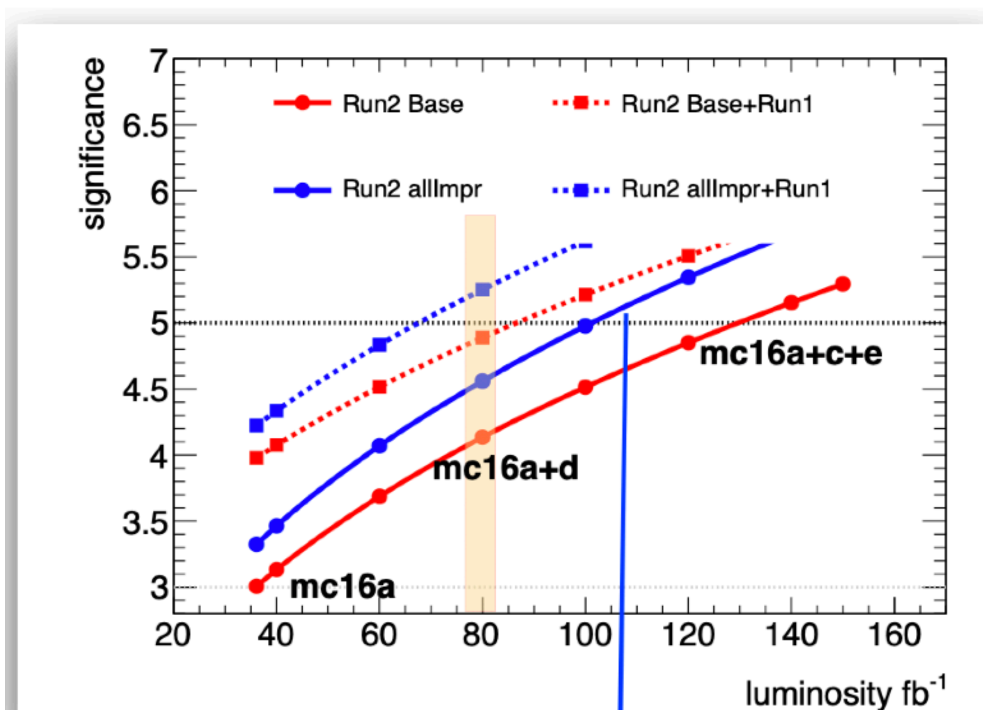
more popular than actual main analysis plot



***3.3** (2.8) **σ obs.** (exp.)*

The only non official plot



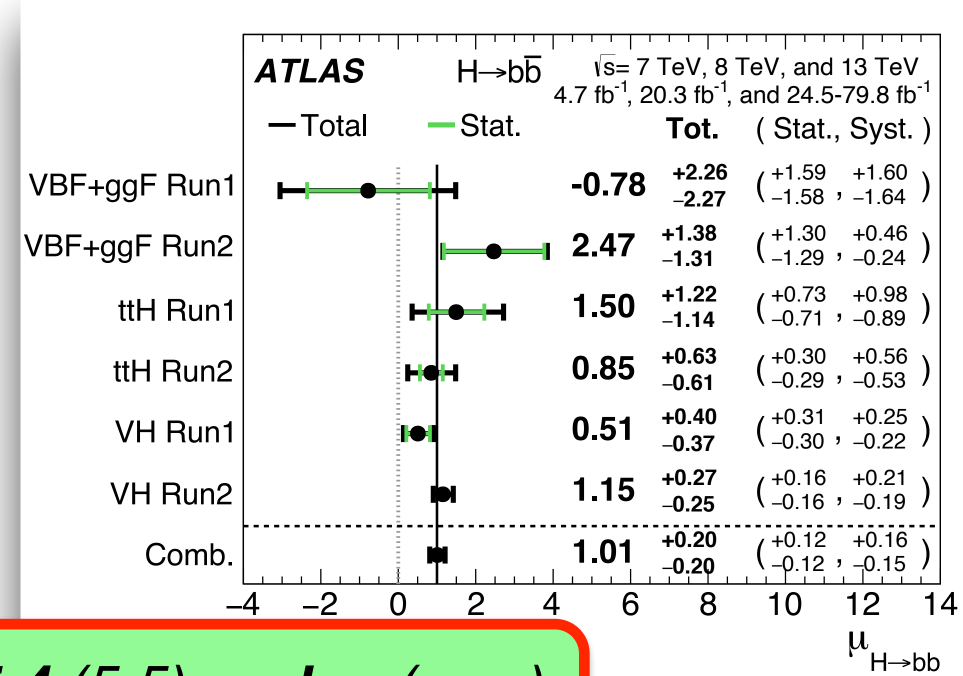


- Observation of VHbb possible with 2017 data if:
  - *keep performing combination with Run1 result (not a big deal)*
  - *improving the Run2 analysis*

- Not simply adding more data: higher pileup, new pixel detector (CMS), new reconstruction SW (ATLAS)

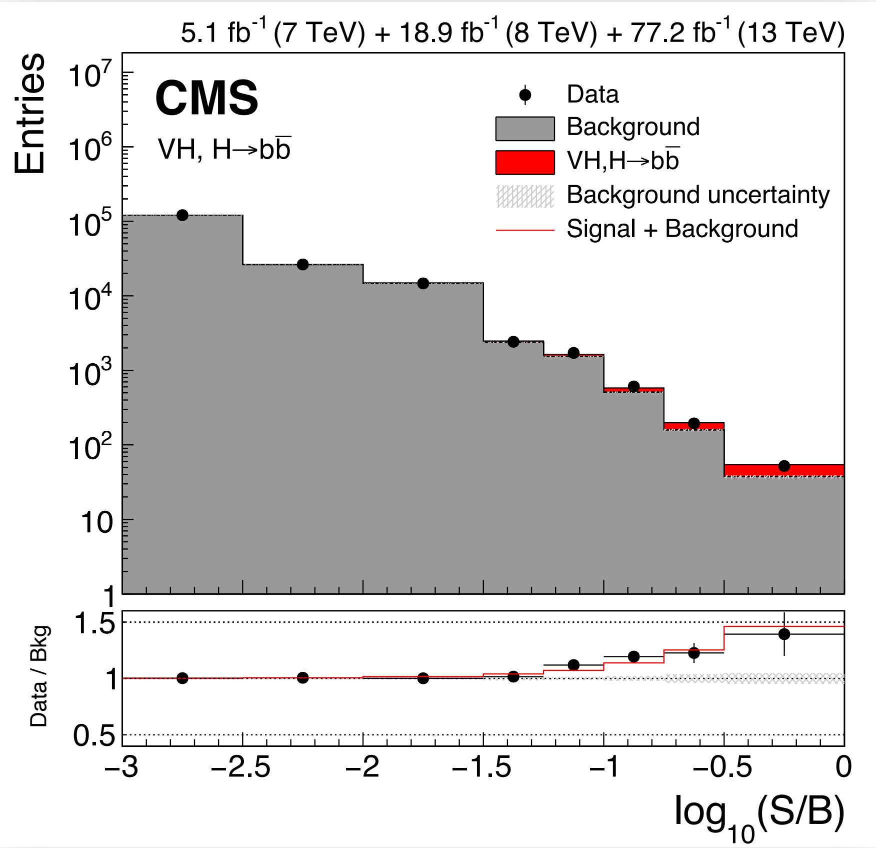
!! 8 months from end of data-taking: incremental improvement, not complete re-design !!

- Full combination:
  - expected improvements from adding 2017 data brought VHbb analysis close to 5-sigma alone
  - prefer not to sweep under the carpet the "down fluctuations" from Run1, contribution from other production mored (mainly ttH) crucial to break the 5sigma ceiling

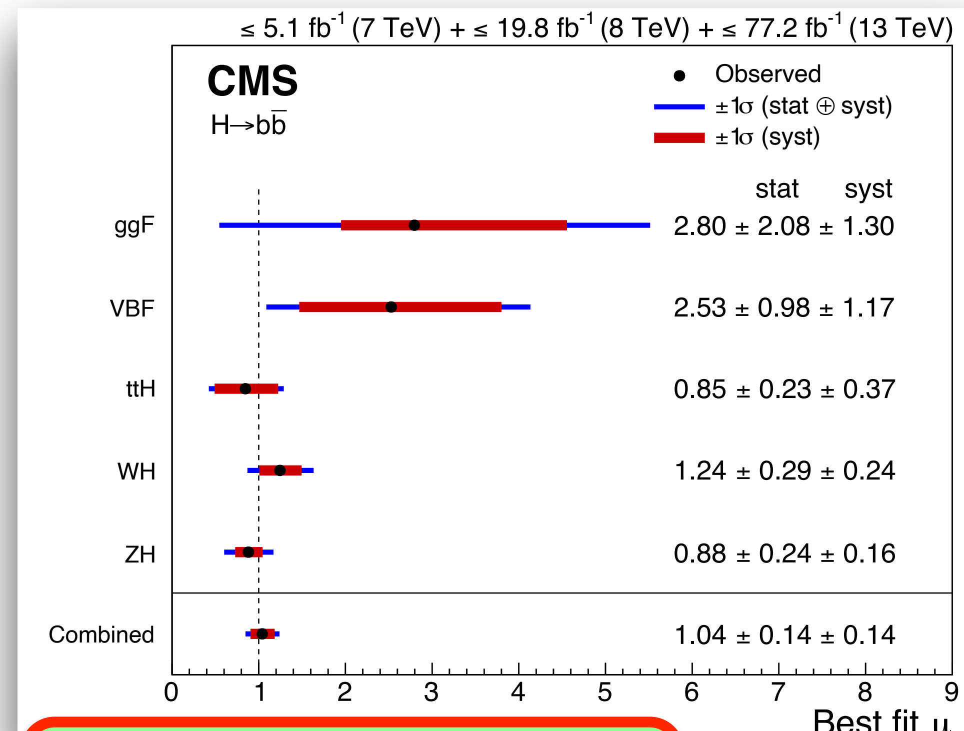


- **Observation of Hbb** from both experiments using Run1 + 2016+2017 Run2 data

***5.4** (5.5) **σ obs.** (exp.)*



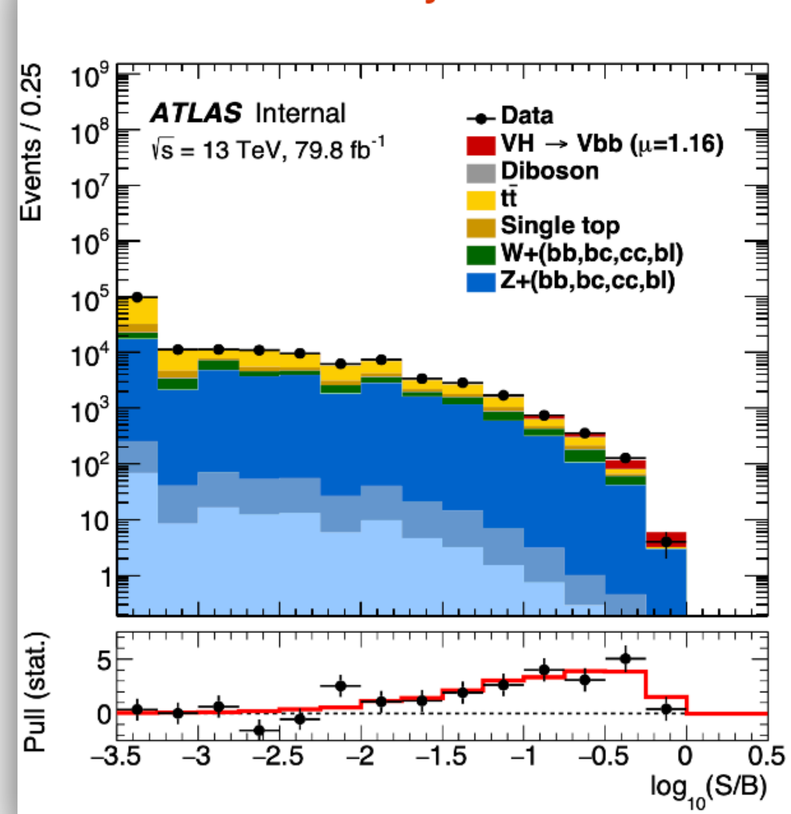
I purposely decided not to discuss exact dates ....



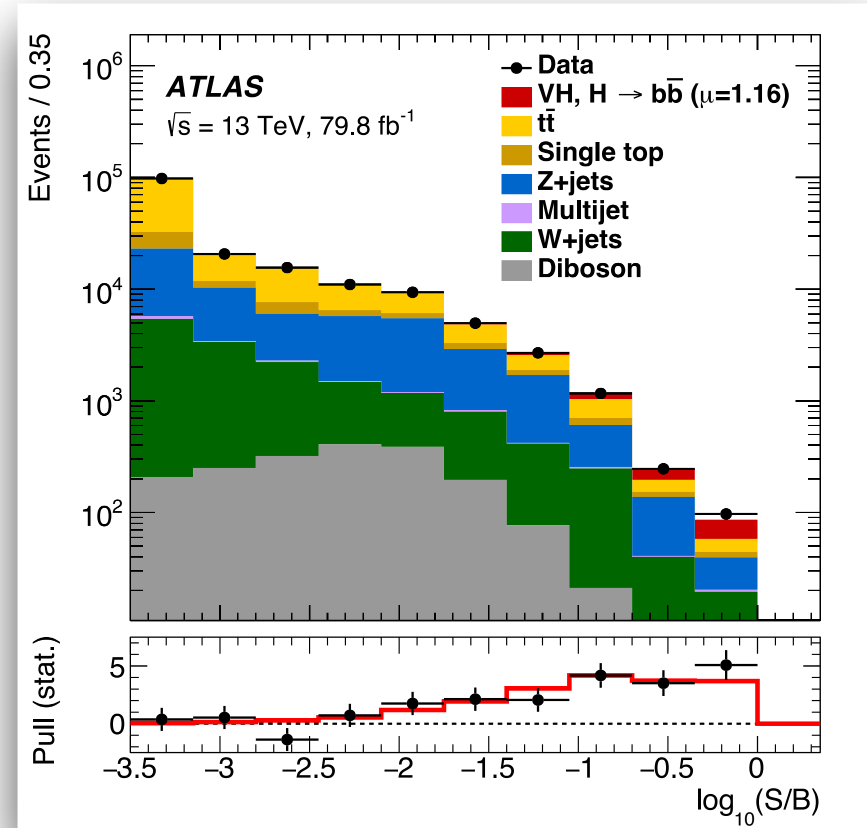
***5.6** (5.5) **σ obs.** (exp.)*

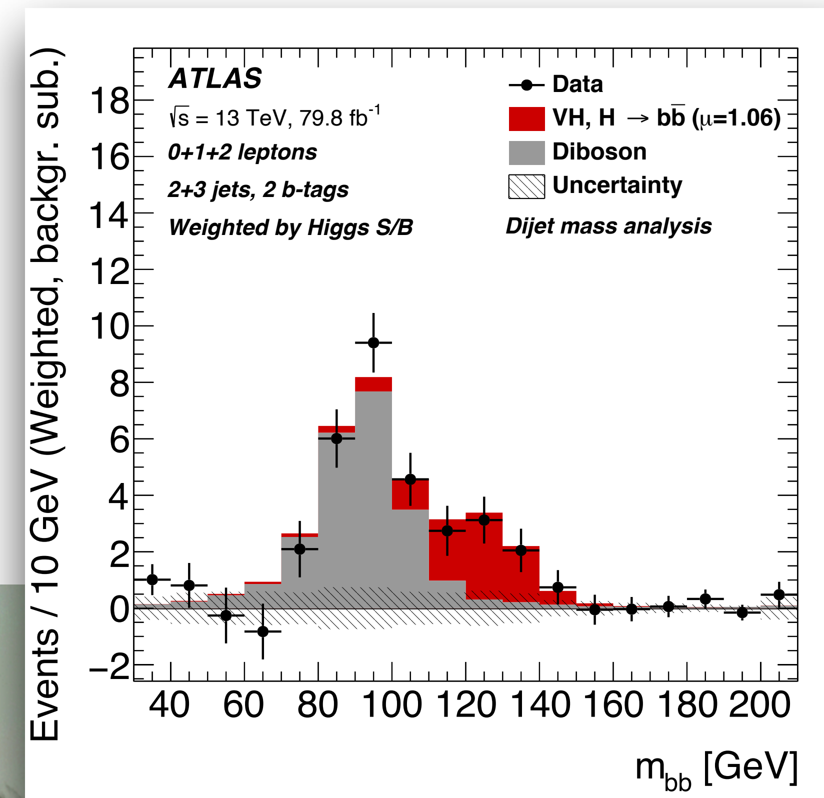
A rewarding moment

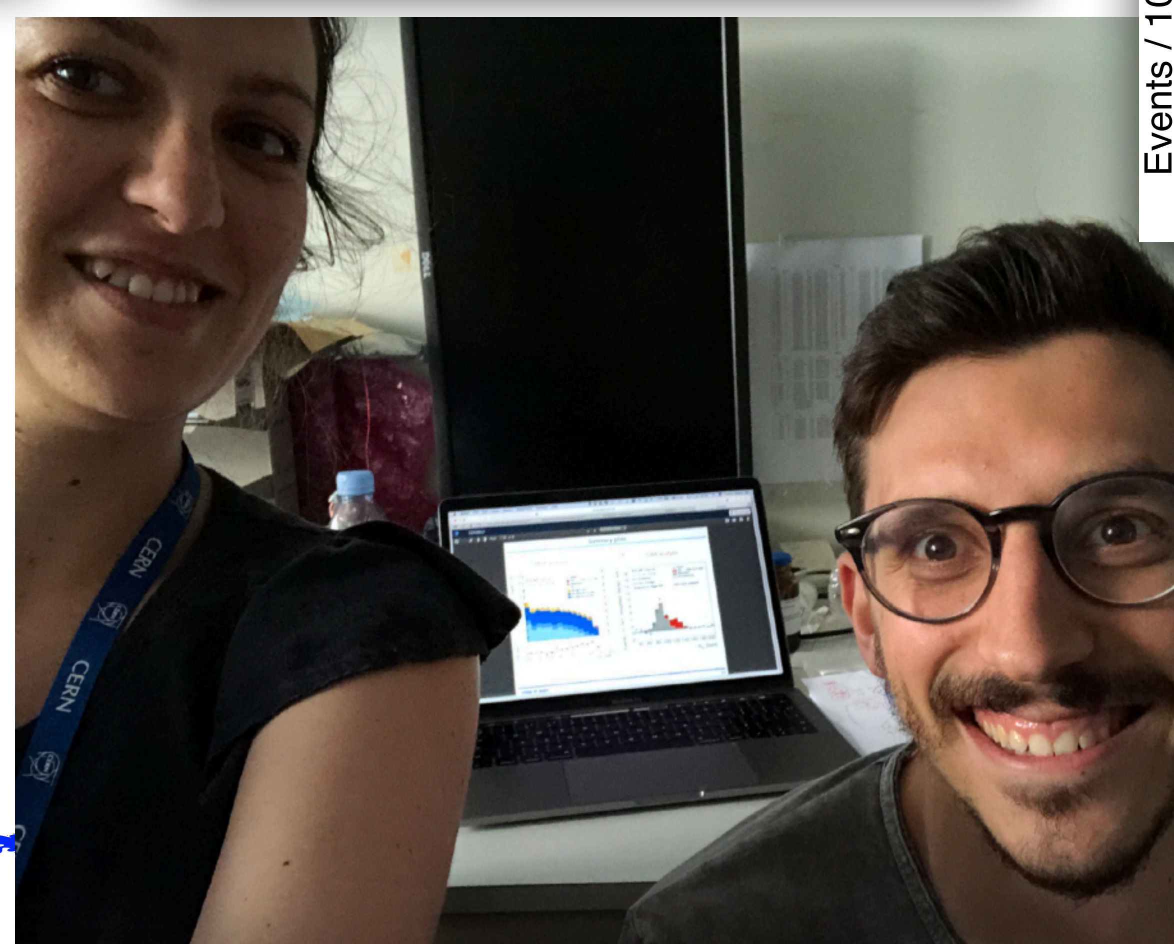
MVA analysis



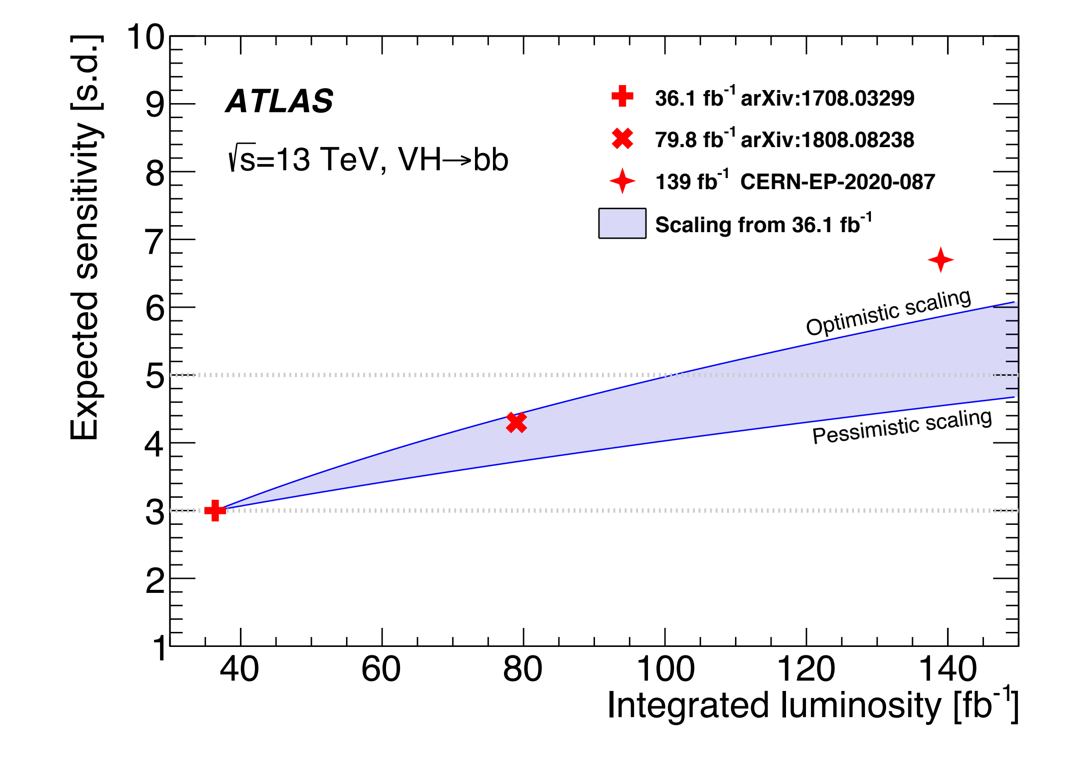
"beautified official version"



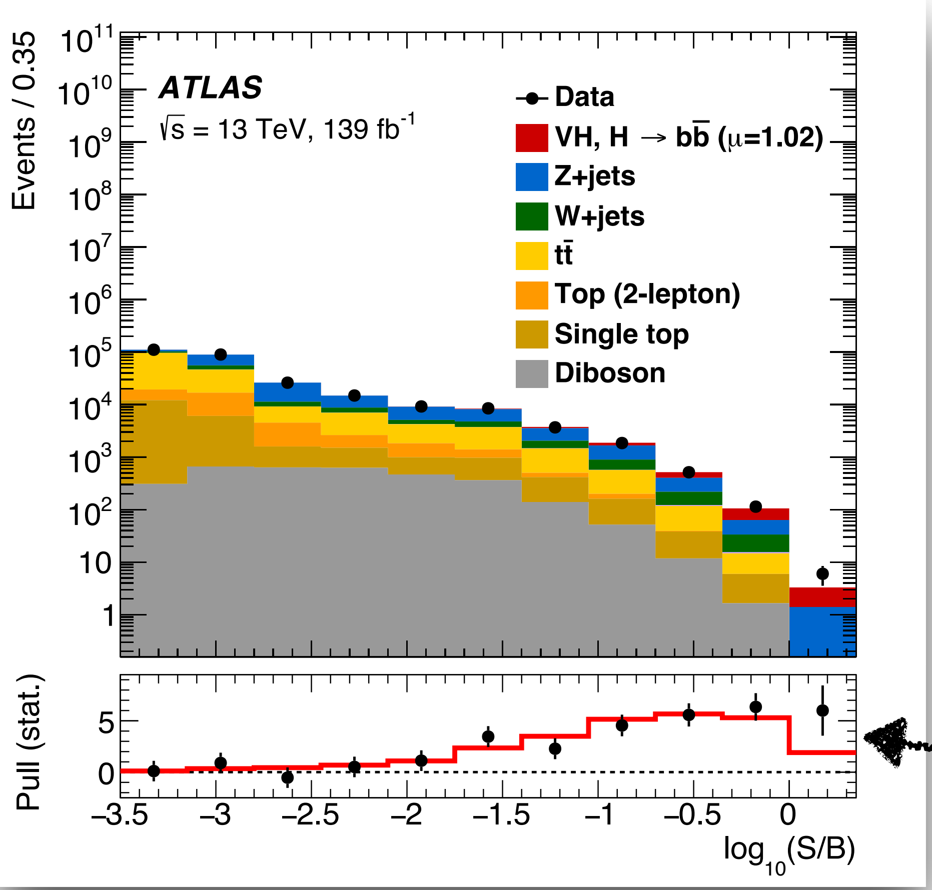




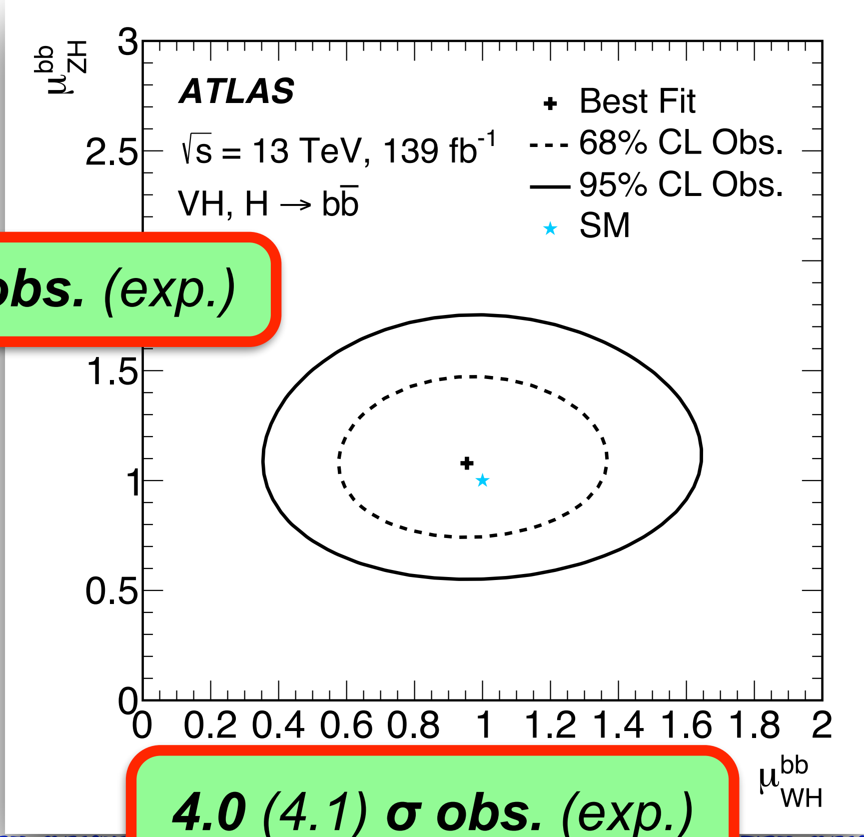
- Analysis **improved more than sqrt(L)** despite being systematically limited:
  - improved physics objects performance / calibration
  - better MC …. but also more MC events
  - further / smarter event classification: isolate high p_T regions with higher S/B
  - more / better control regions to estimate background



- **VHbb dominates our understanding of the Higgs coupling to bottom quarks**



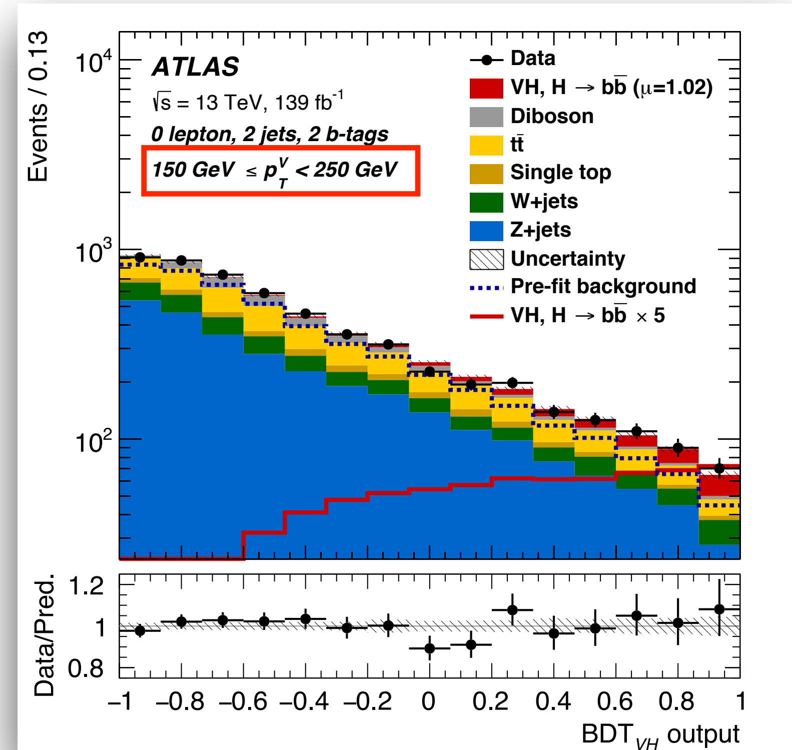
S/B>1

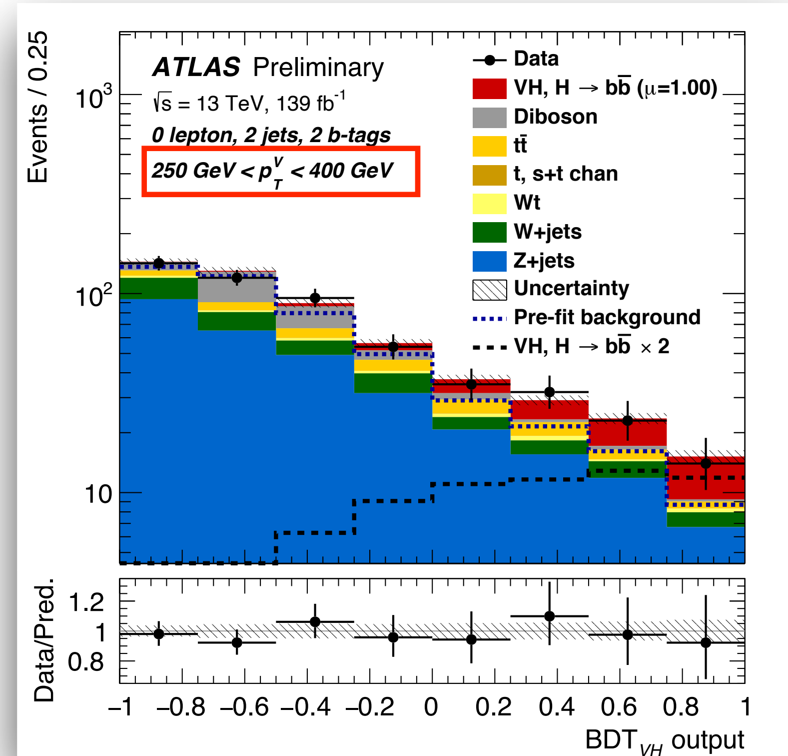


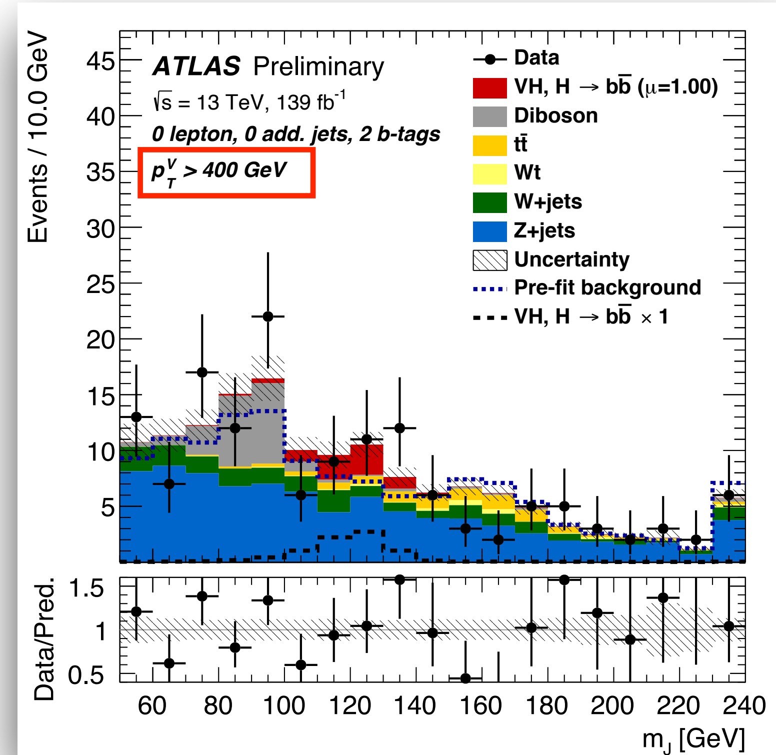
***5.3** (5.1) **σ obs.** (exp.)*

***4.0** (4.1) **σ obs.** (exp.)*

Beyond signal strength







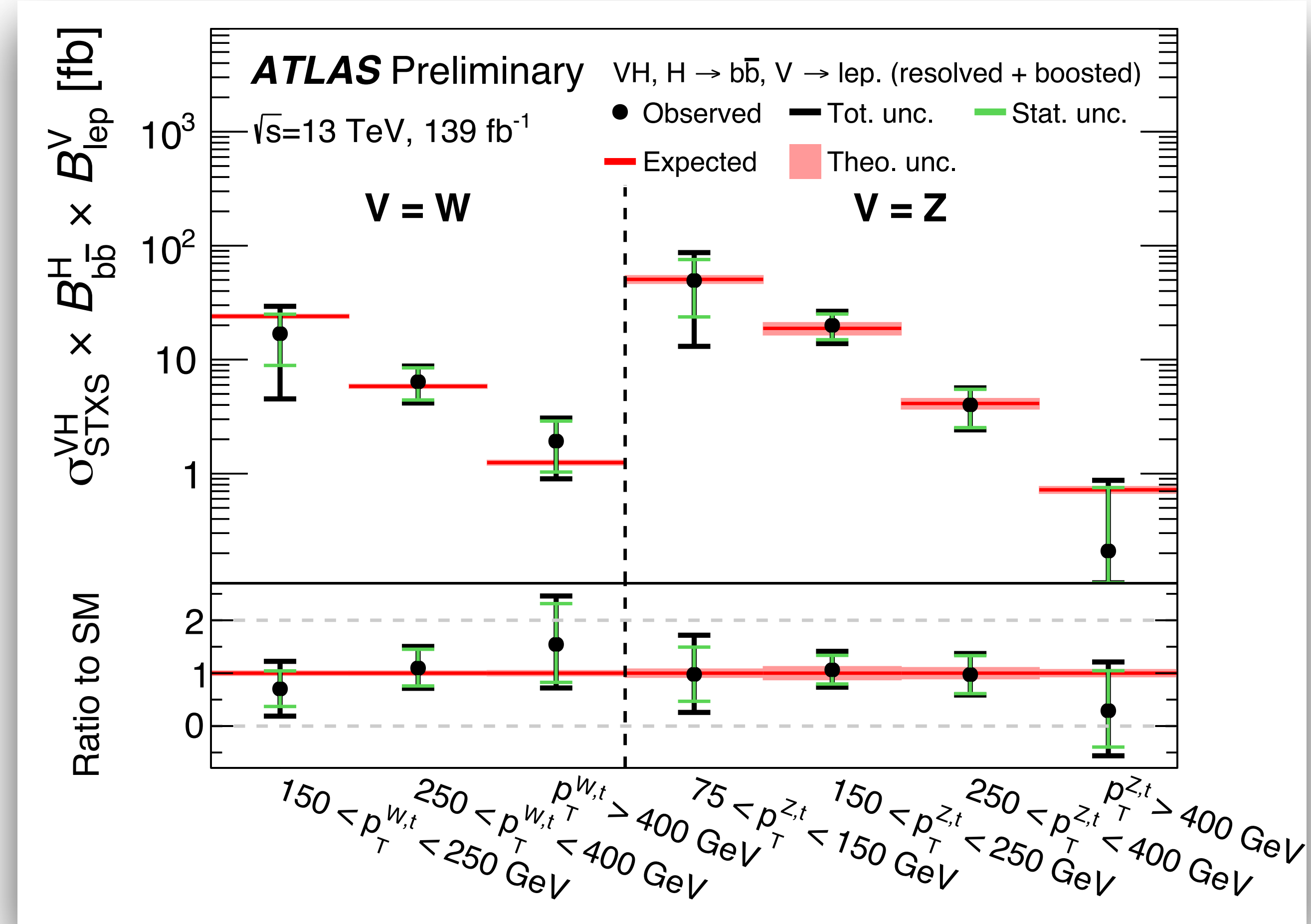
- ✦ V p_T > 400 GeV category exploits reconstruction of Higgs in single largeR (R=1.0) jet [more later]

ATLAS Simulation Preliminary √s = 13 TeV, 139 fb^{-1} (Signal events)

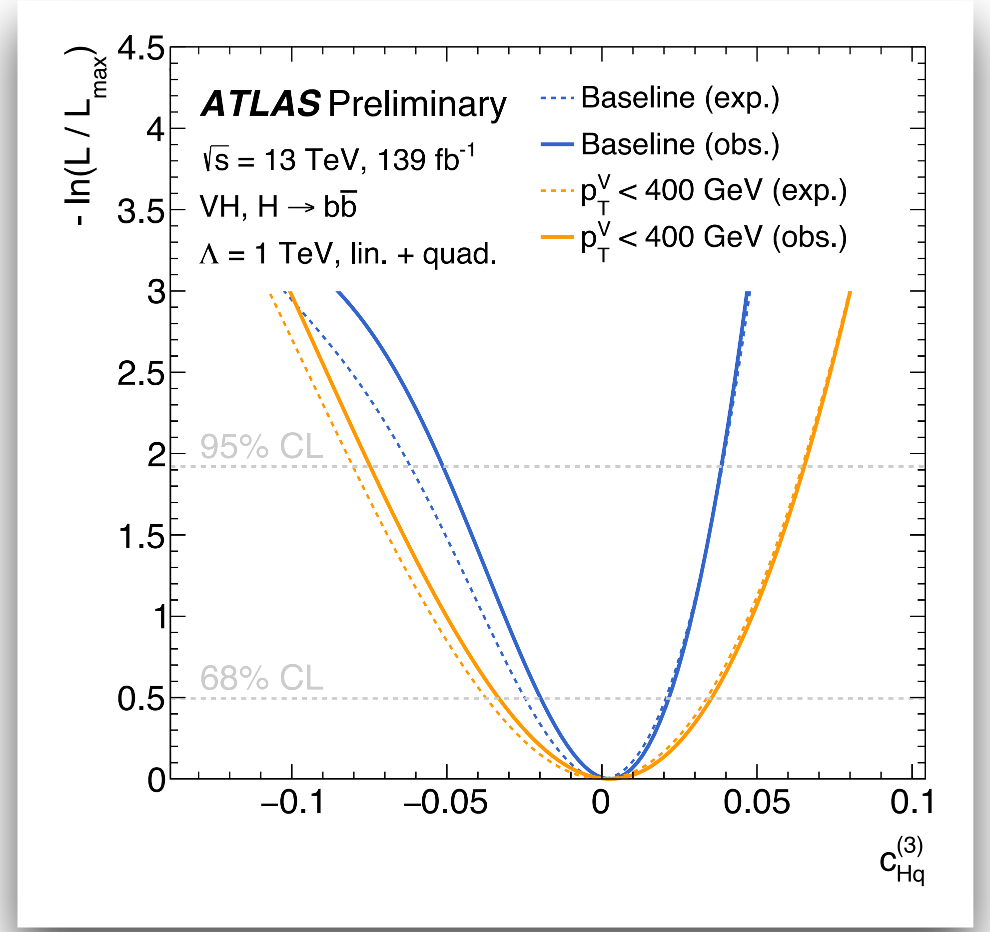
	WH, $p_T^{W,t}$ < 150 GeV	WH, 150 < $p_T^{W,t}$ < 250 GeV	WH, 250 < $p_T^{W,t}$ < 400 GeV	WH, $p_T^{W,t}$ > 400 GeV	ZH, $p_T^{Z,t}$ < 75 GeV	ZH, 75 < $p_T^{Z,t}$ < 150 GeV	ZH, 150 < $p_T^{Z,t}$ < 250 GeV	ZH, 250 < $p_T^{Z,t}$ < 400 GeV	ZH, $p_T^{Z,t}$ > 400 GeV
2 lep, p_T^V > 400 GeV, boosted SR								0.2	3.6
2 lep, ≥ 3 jets, 250 < p_T^V < 400 GeV							0.8	20.7	0.2
2 lep, 2 jets, 250 < p_T^V < 400 GeV							0.3	8.4	0.1
2 lep, ≥ 3 jets, 150 < p_T^V < 250 GeV						1.6	74.0	1.2	
2 lep, 2 jets, 150 < p_T^V < 250 GeV						0.8	33.8	0.5	
2 lep, ≥ 3 jets, 75 < p_T^V < 150 GeV					1.2	124.7	2.9		
2 lep, 2 jets, 75 < p_T^V < 150 GeV					0.9	78.4	1.4		
0 lep, p_T^V > 400 GeV, boosted LP SR			0.1	1.5				0.6	5.4
0 lep, p_T^V > 400 GeV, boosted HP SR			0.1	1.3				0.6	6.0
0 lep, 3 jets, 250 < p_T^V < 400 GeV		0.4	4.4	1.1			3.6	23.3	1.5
0 lep, 2 jets, 250 < p_T^V < 400 GeV		0.3	4.8	1.4			2.7	24.2	1.8
0 lep, 3 jets, 150 < p_T^V < 250 GeV	2.8	22.6	7.0	0.3		12.5	93.0	9.8	0.1
0 lep, 2 jets, 150 < p_T^V < 250 GeV	2.1	22.6	7.5	0.5		11.0	95.0	10.9	0.1
1 lep, p_T^V > 400 GeV, boosted LP SR			0.8	8.8					0.1
1 lep, p_T^V > 400 GeV, boosted HP SR			0.9	9.8					0.1
1 lep, 3 jets, 250 < p_T^V < 400 GeV		4.8	35.4	2.8			0.1	0.6	0.2
1 lep, 2 jets, 250 < p_T^V < 400 GeV		3.8	37.2	3.2				0.4	0.1
1 lep, 3 jets, 150 < p_T^V < 250 GeV	16.2	110.5	13.4	0.1		0.7	3.5	1.0	0.1
1 lep, 2 jets, 150 < p_T^V < 250 GeV	12.7	118.6	15.1	0.2		0.4	2.4	0.6	

- ✦ Exploit analysis categorisation to provide signal strength in kinematic bins

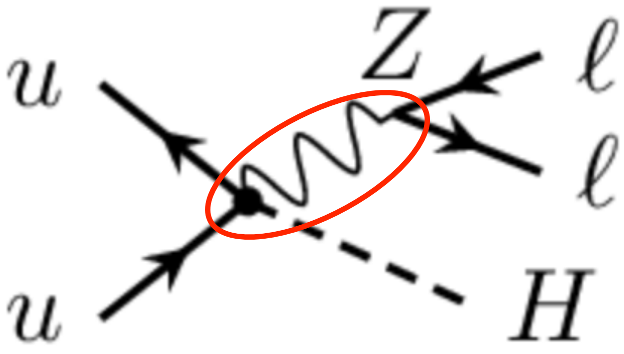
Beyond signal strength



- ✦ **VH differential information** in 7 STXS categories as a function of vector boson p_T
  - ✦ uncertainties ranging from 20% to 100%

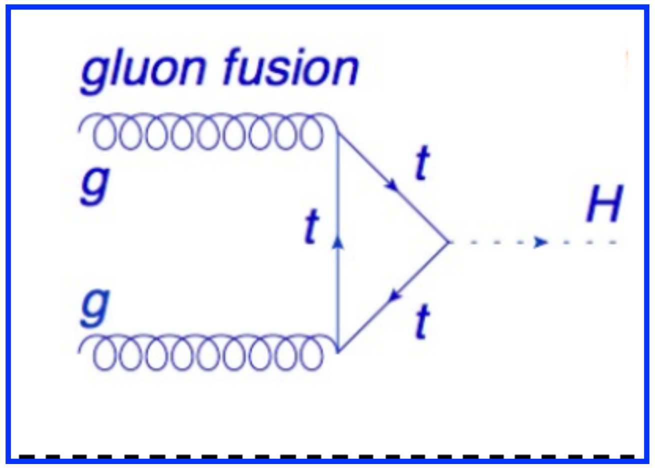


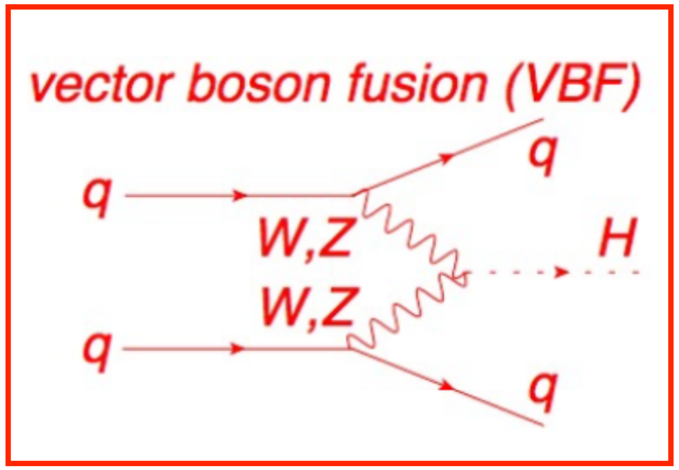
- ✦ Important information to constrain effects of **energy-dependent EFT operators** [enhanced effects at high p_T]
- ✦ Less precise info than H->VV decay …. but expected effect much larger.

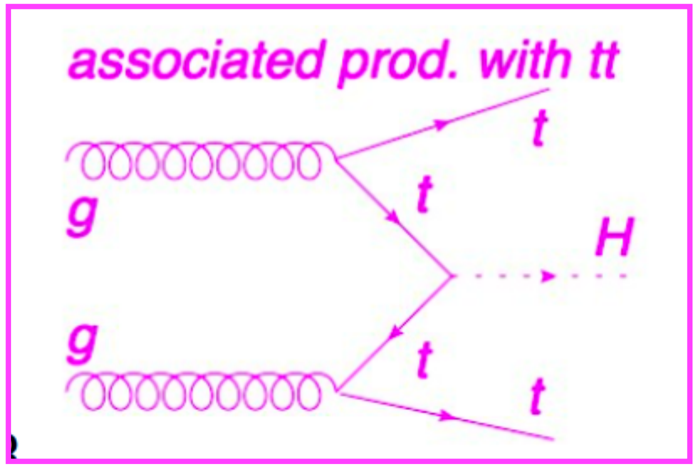


Beyond VH

- Other production mode are still crucial to test consistency of the model
- Despite low S/B … H->bb decay provides the largest statistics ==> important measuring/search tool



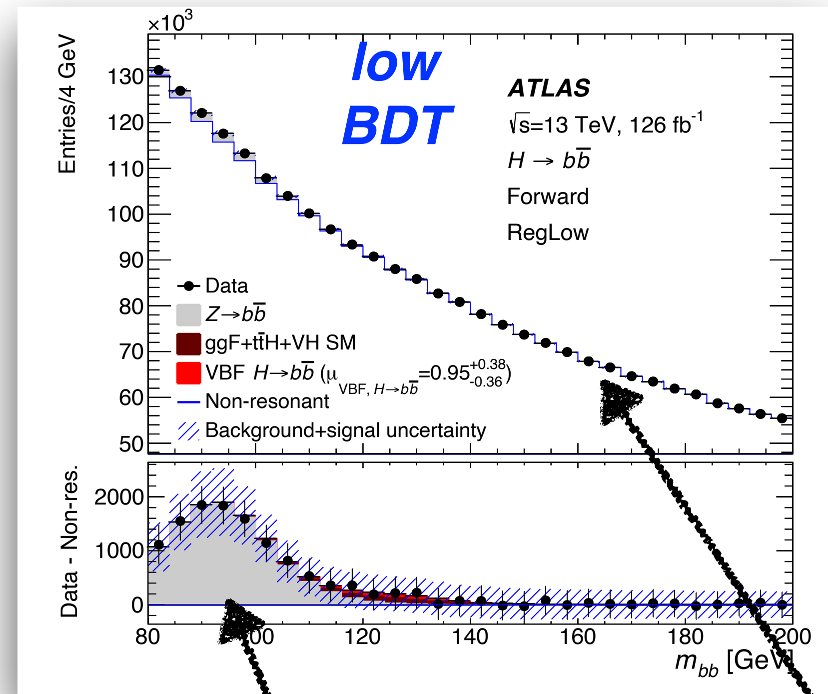


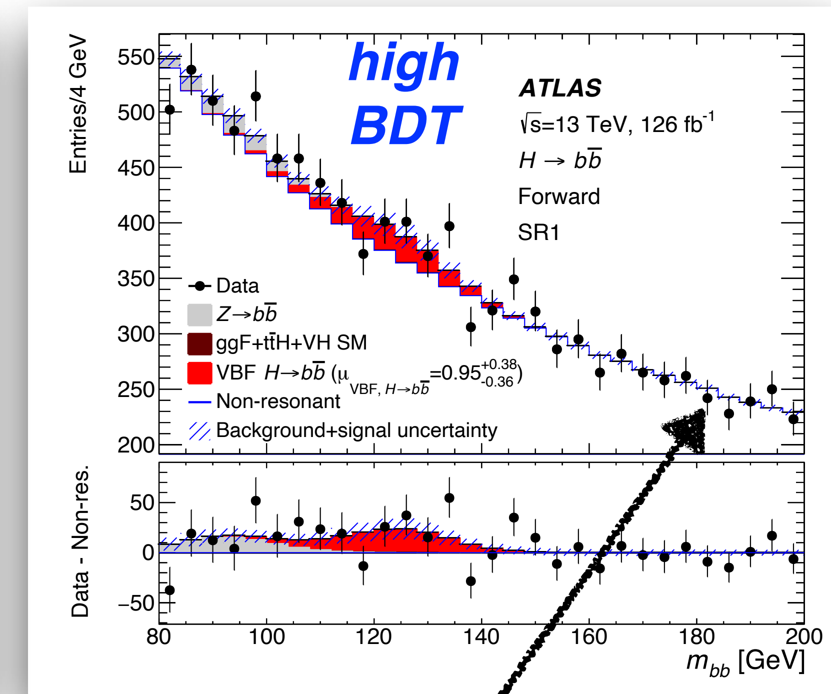


see Tamara's talk

- **Main hurdles of not relying on leptons**:
  - how to select events effectively? Solution: **many objects or raise pt threshold**
  - how to estimate the large QCD multi-jet background? Poor description from MC: **need to rely on data-driven techniques.**

VBF H->bb





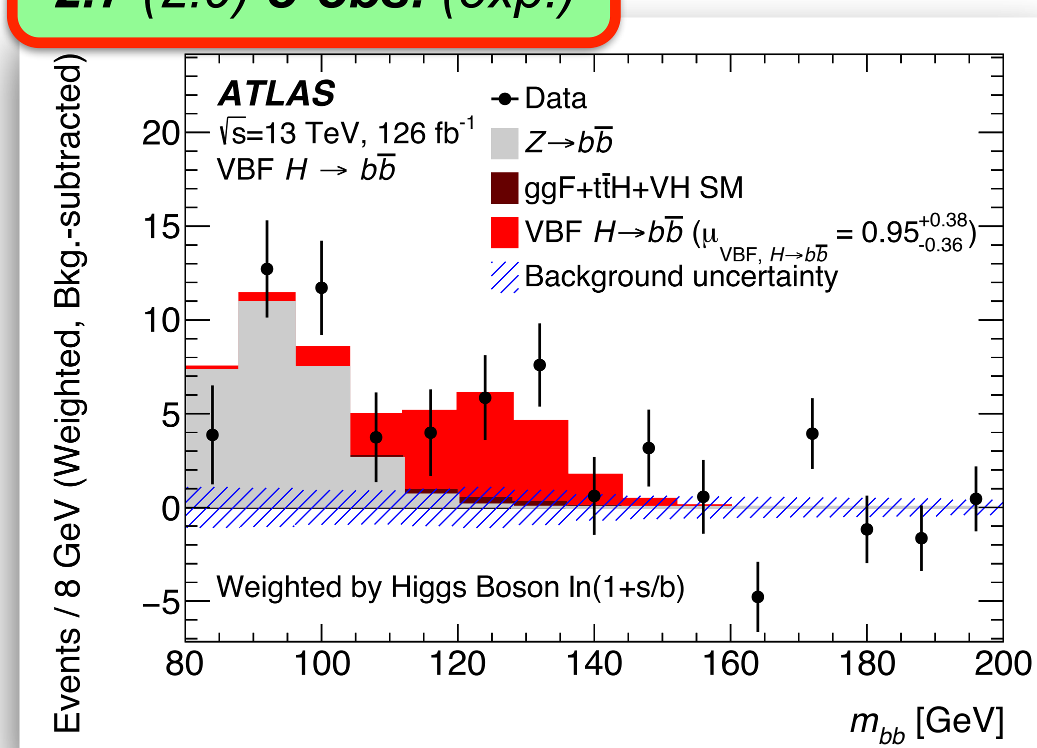
- 2 b-jets + 2 fwd (or 1 cen-1 fwd jet)
- 5 event categories defined with neural network exploiting object kinematics
- **adversarial training** to ensure that shape of multi jet background is identical across the various regions

- Zbb normalisation estimated from clean Z->mumu events
- QCD shape in SR taken from high stat. CR

Fantastic example of how to solve non conventional problems and bring fully hadronic final states in the spotlight

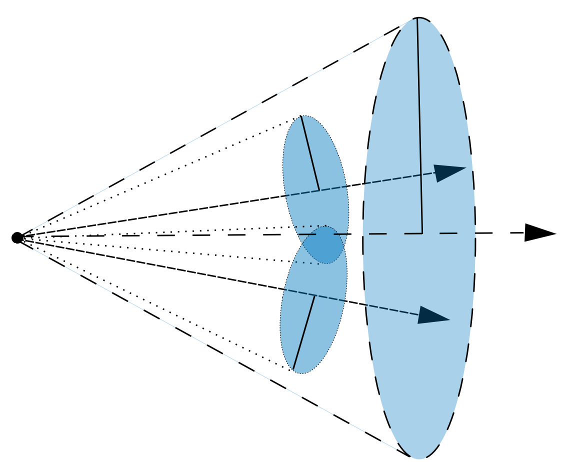
- **Very close to evidence for H->bb VBF production**: x6 improvement in performance w.r.t. early Run2 analysis. (also exploiting VBF+H+photon selection)

***2.7** (2.9) **σ obs.** (exp.)*

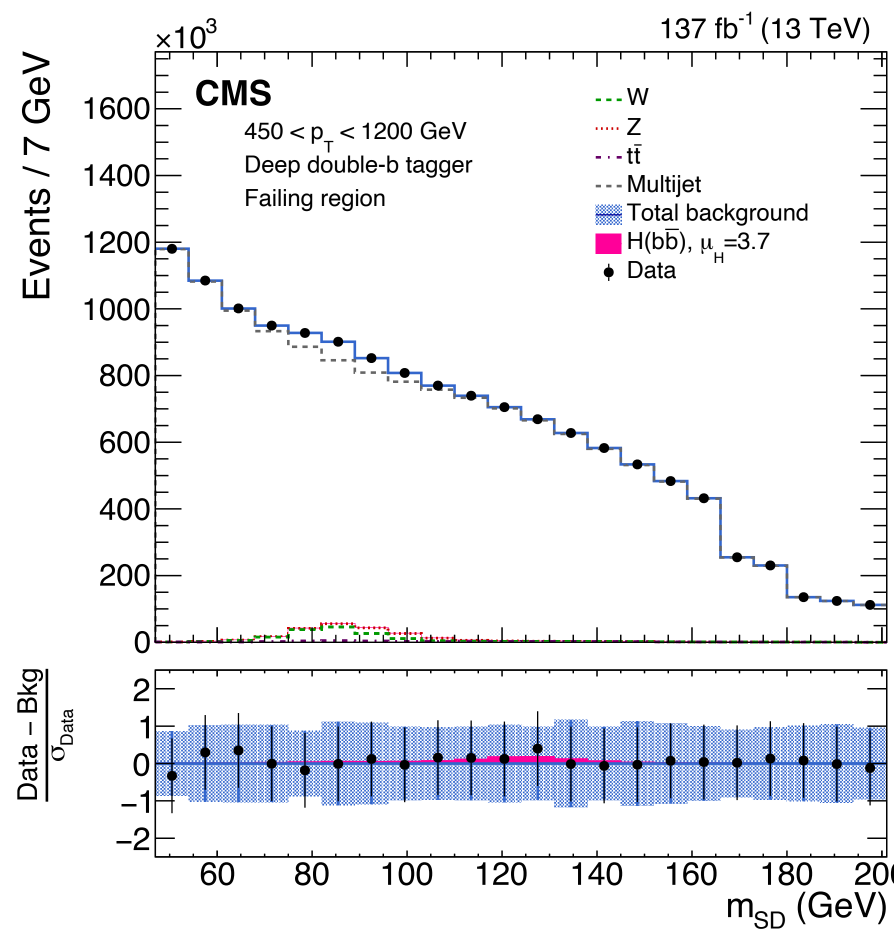


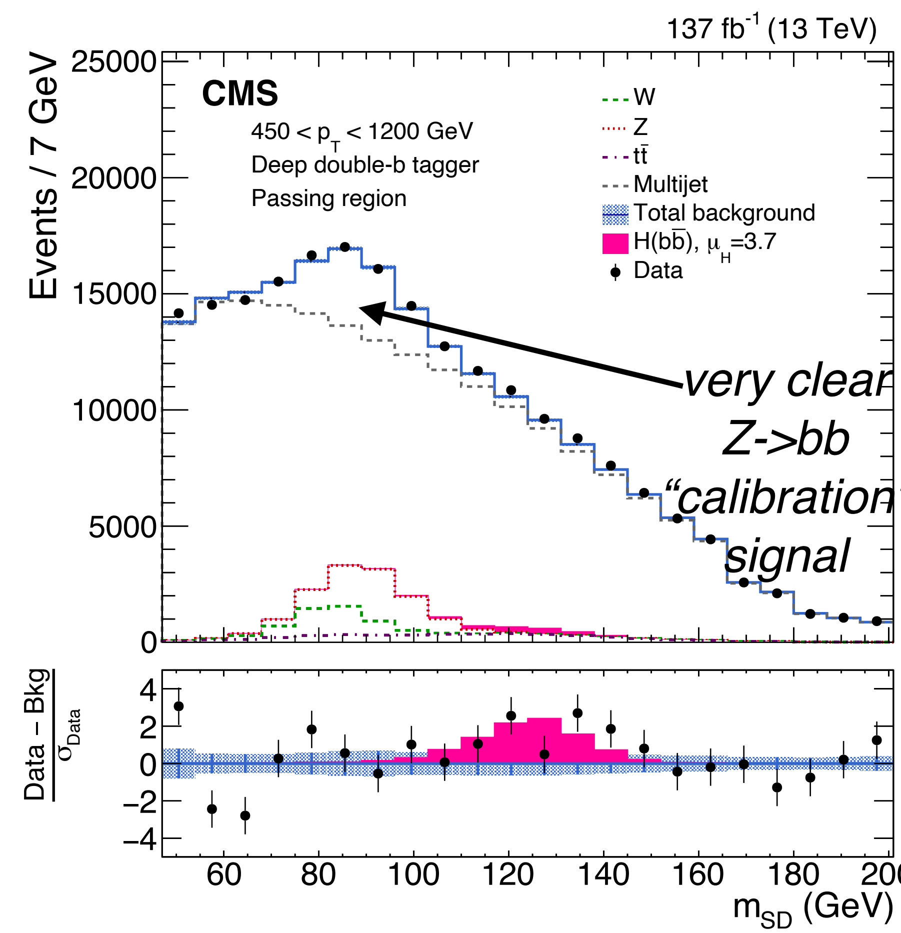
- Capturing Higgs decay in single large radius jet makes the impossible ggH final state possible …

CMS		ATLAS
0.8	*jet radius*	1.0
soft drop	*jet mass*	jet trimming
leading p_T>450 GeV	*jet p_T*	leading (sub) p_T>450 (250)
mass-decorrelated N12	***substructure***	**2 sub jets (track based)**
dedicated NN for H->bb	***tagger***	**sub-jet tagging**
from failed region	*background*	functional form



- ATLAS and CMS are a lot less aligned …. and the difference matters



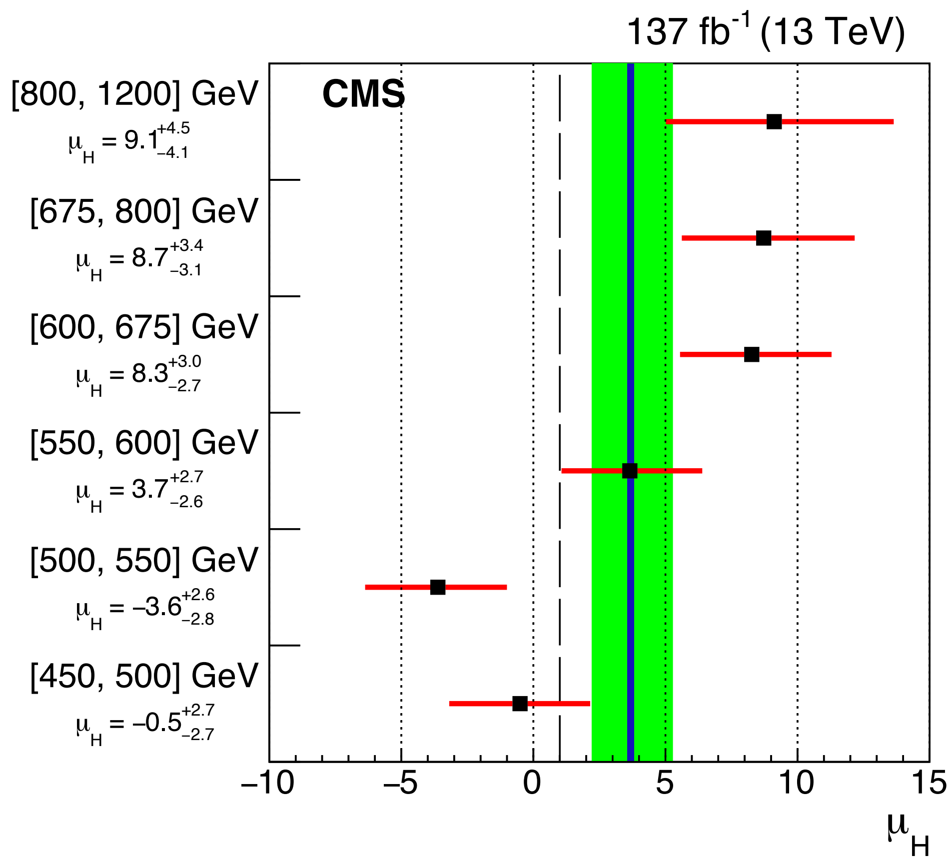


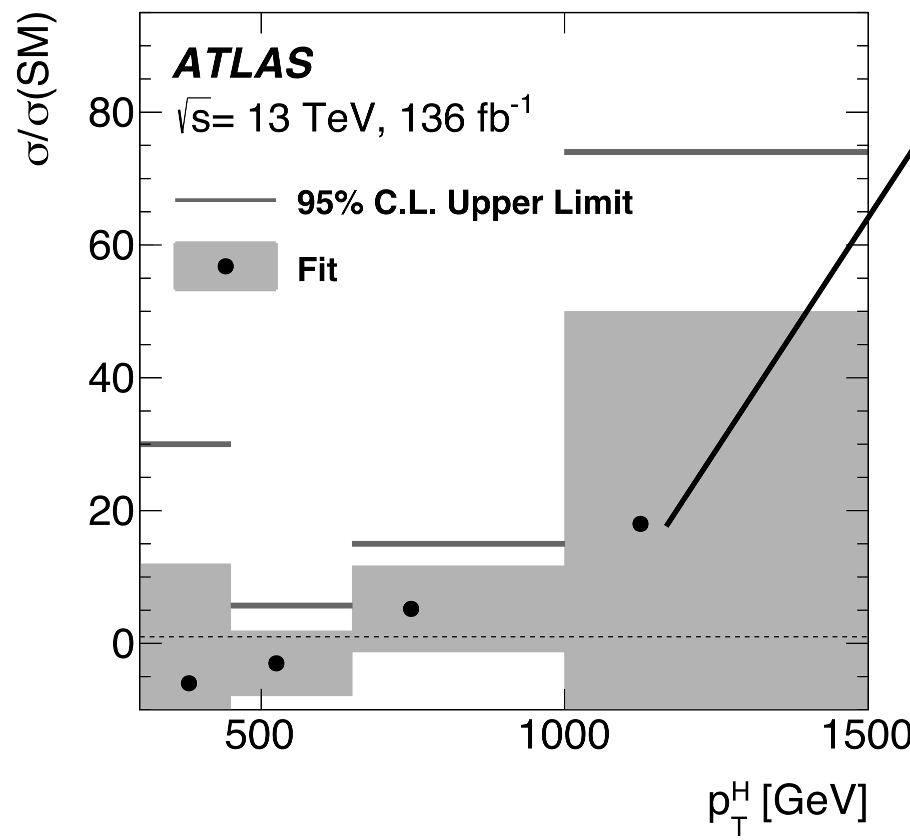
- Results extracted in several (not aligned) bins of jet p_T.

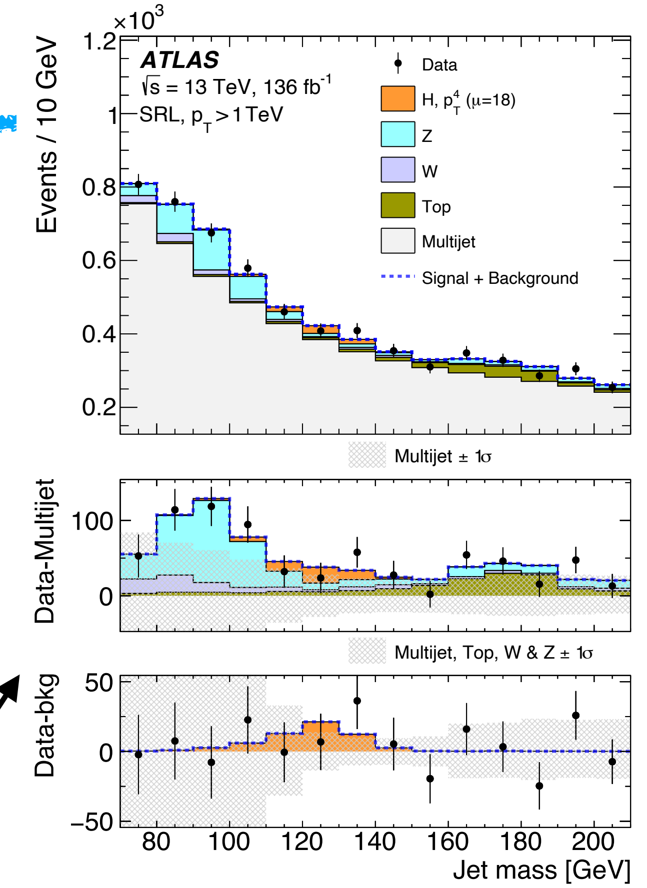
using HJ-MINLO as a reference

$\mu = 3.7 \pm 1.5$

$\mu = 0.8 \pm 3.2$







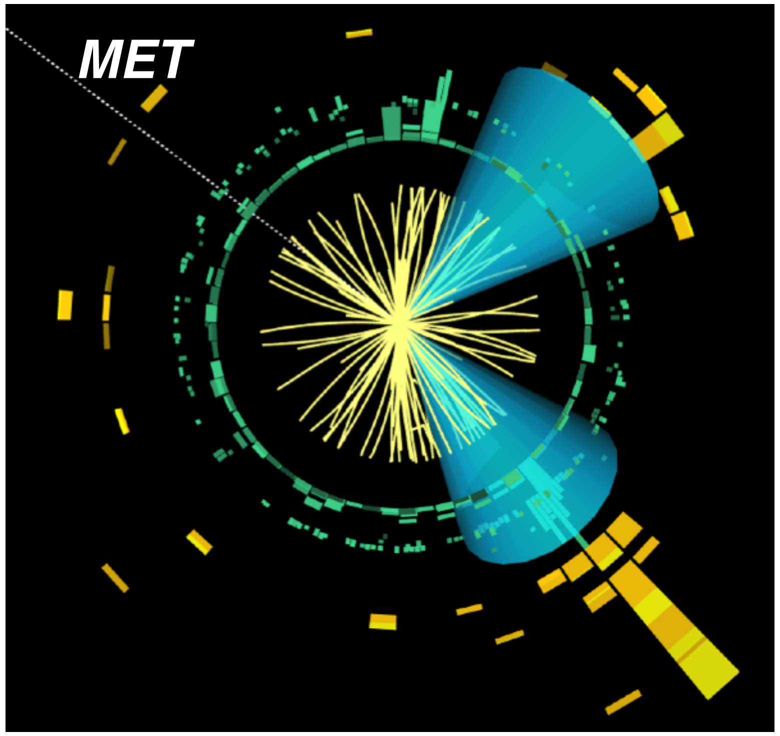
- Unique regime only reached by H->bb.

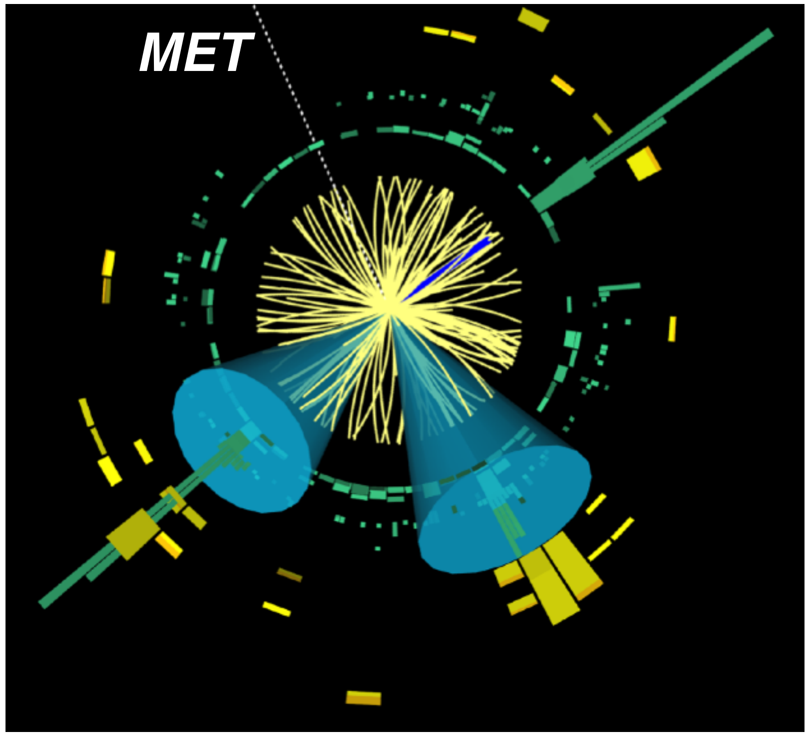
- Partial Run2 results from CMS sparked a lot of attention on high p_T theoretical calculations: high p_T disagreement increases when considering proper top-mass effects.

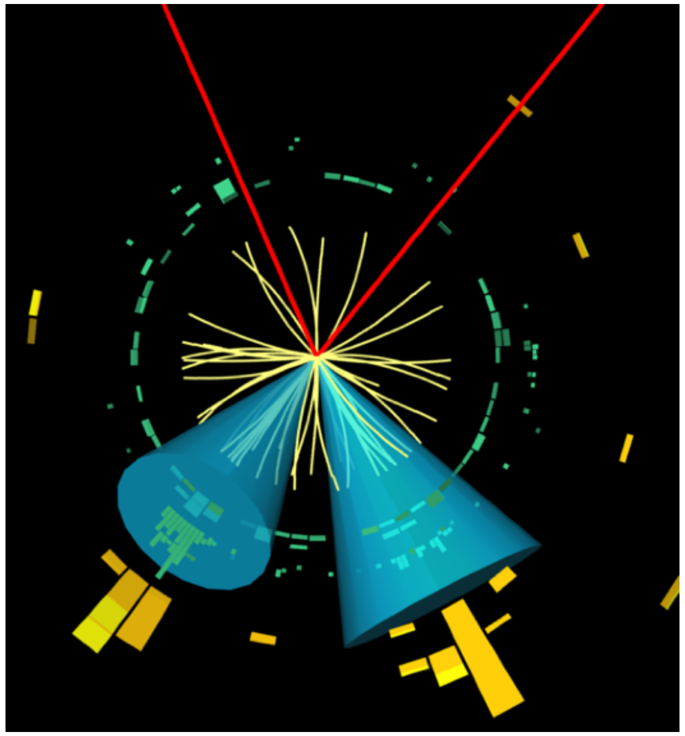
Very intriguing patterns …….

H->cc

…. like H->bb but harder …

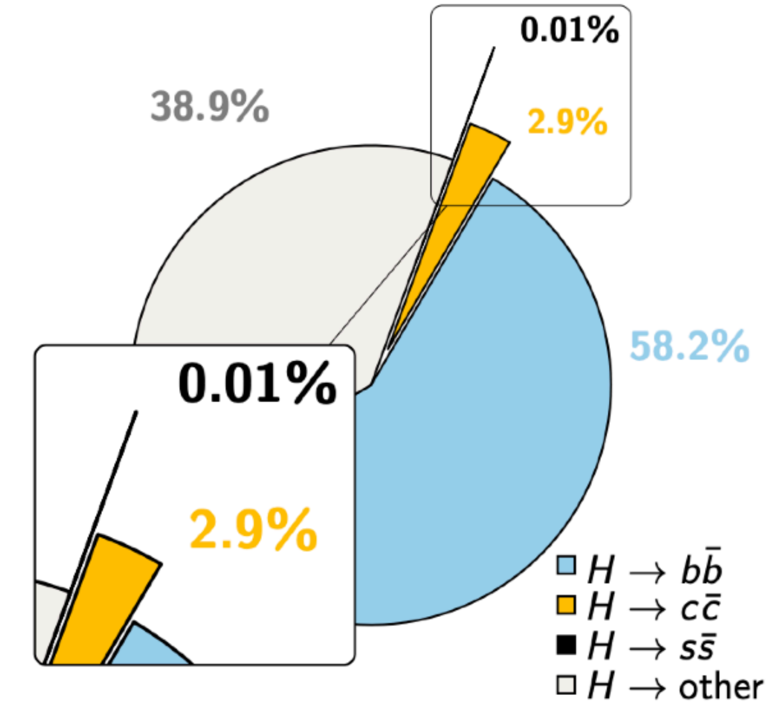




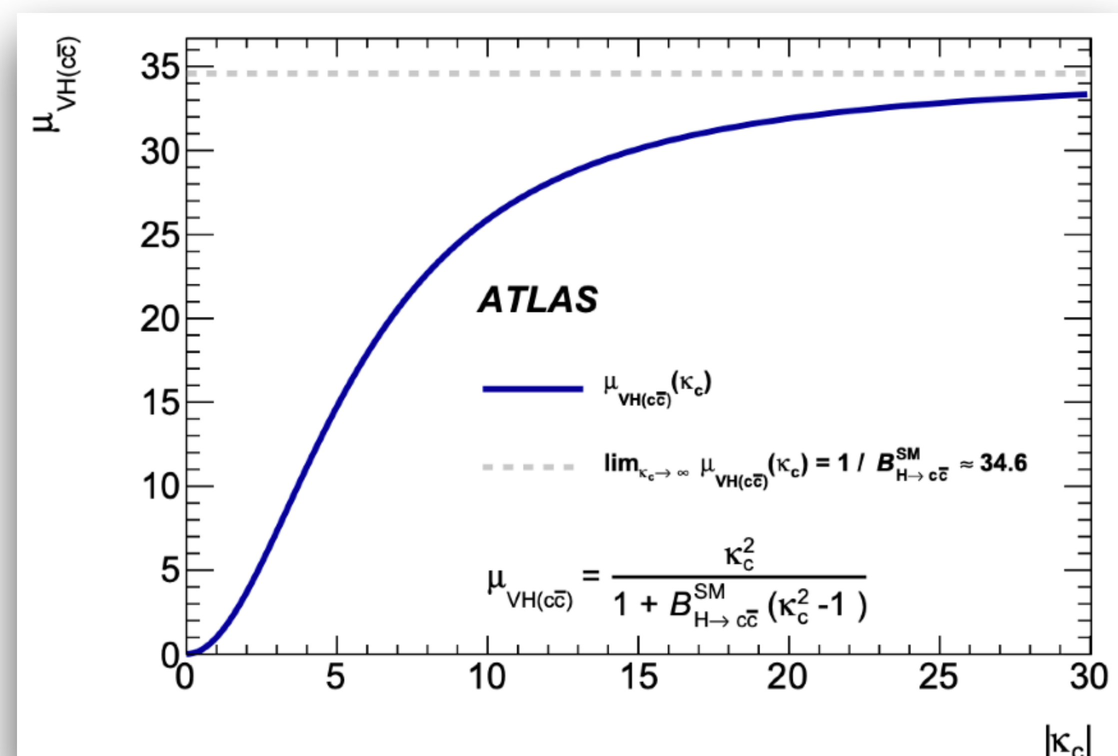


- **H->cc will shed light on Higgs coupling to second generation quarks:** *important to verify that the Higgs mechanism works in similar ways for all the fermions*

- In comparison to H->bb:
  - **x20 smaller Br**: H->bb is non negligible background for H->cc
  - **c-tagging is much harder than b-tagging** (smaller mass, smaller lifetime): can achieve %-level fake rate only with 20-30% signal efficiency … need to work with looser WP
  - **g->cc background is larger than g->bb one** (but fewer c than b in ttbar)



SM $m_H = 125$ GeV $H \to q\bar{q}$ branching fractions, $H \to u\bar{u}/d\bar{d}$ too small to show!



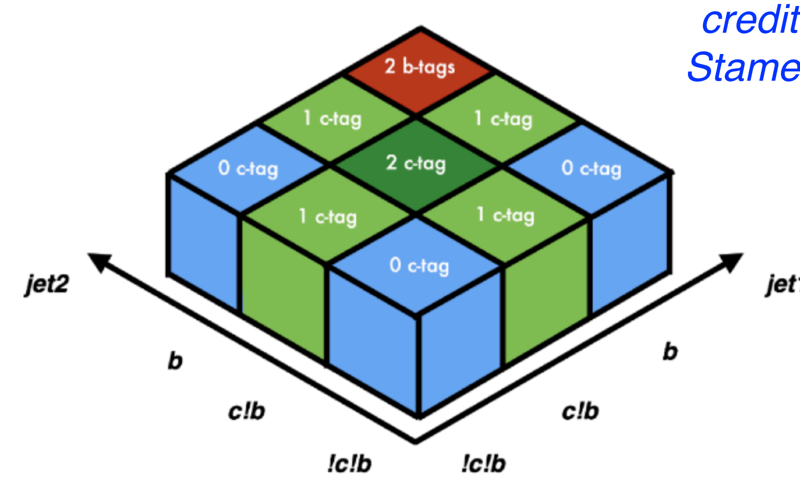
- **H->cc analyses shares many aspects with much more mature H->bb analyses** (VH is still the golden channel) **….. but sometimes thinking outside the box can lead to large gains**

- Br dependency on k_c saturates:
  - can only set limit on k_c if Br signal strength is below 34

VH H->cc: ATLAS (2021)

credit to M. Stamenkovic

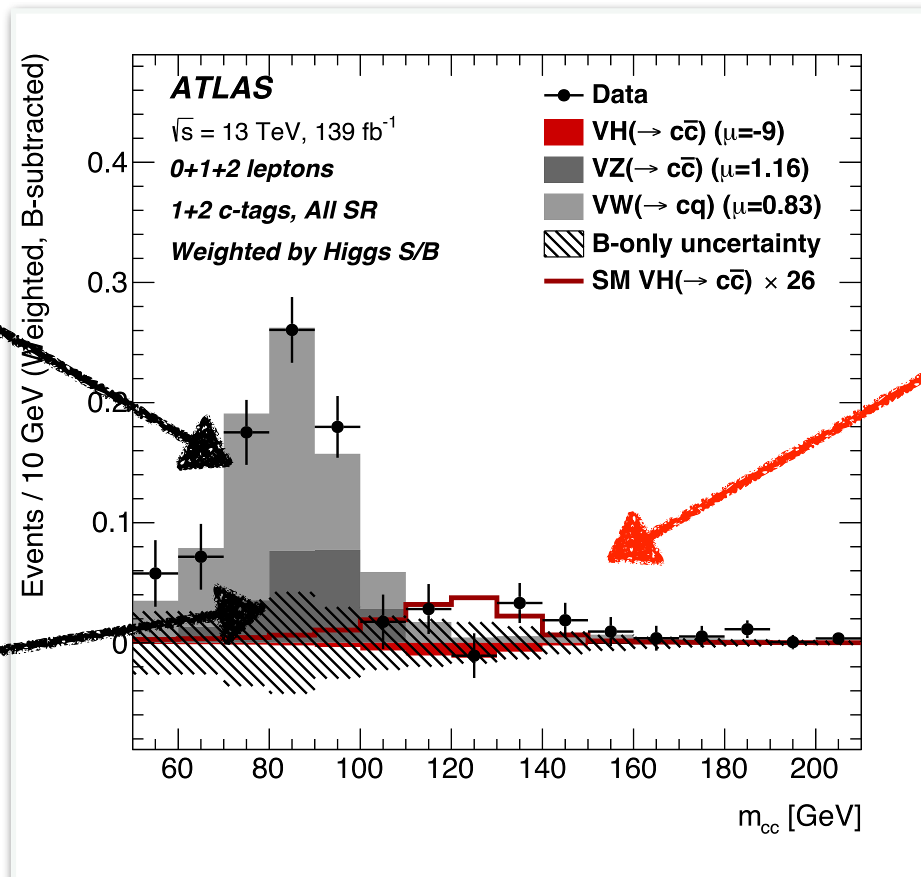
- **Single c-tagging working point used in the analysis:**
  - *c-jet eff. = 27% [b-jet eff. = 8% , light jet eff. = 1.6%]*
  - *include veto of algorithm used for VH H->bb analysis ==> designed for combination*
  - *exploiting both 1-ctag and 2-ctag regions*



- **Combined fit to m_{jj} distribution in 44 regions** (no MVA discriminants)

VW, W->cs
3.8 (4.6) σ obs. (exp.)

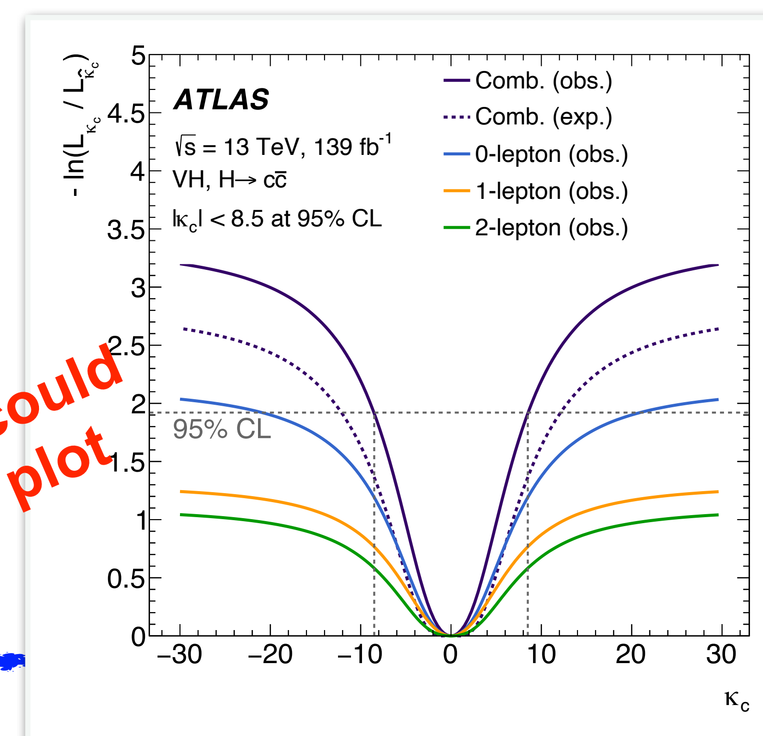
VZ, Z->cc
2.6 (2.2) σ obs. (exp.)

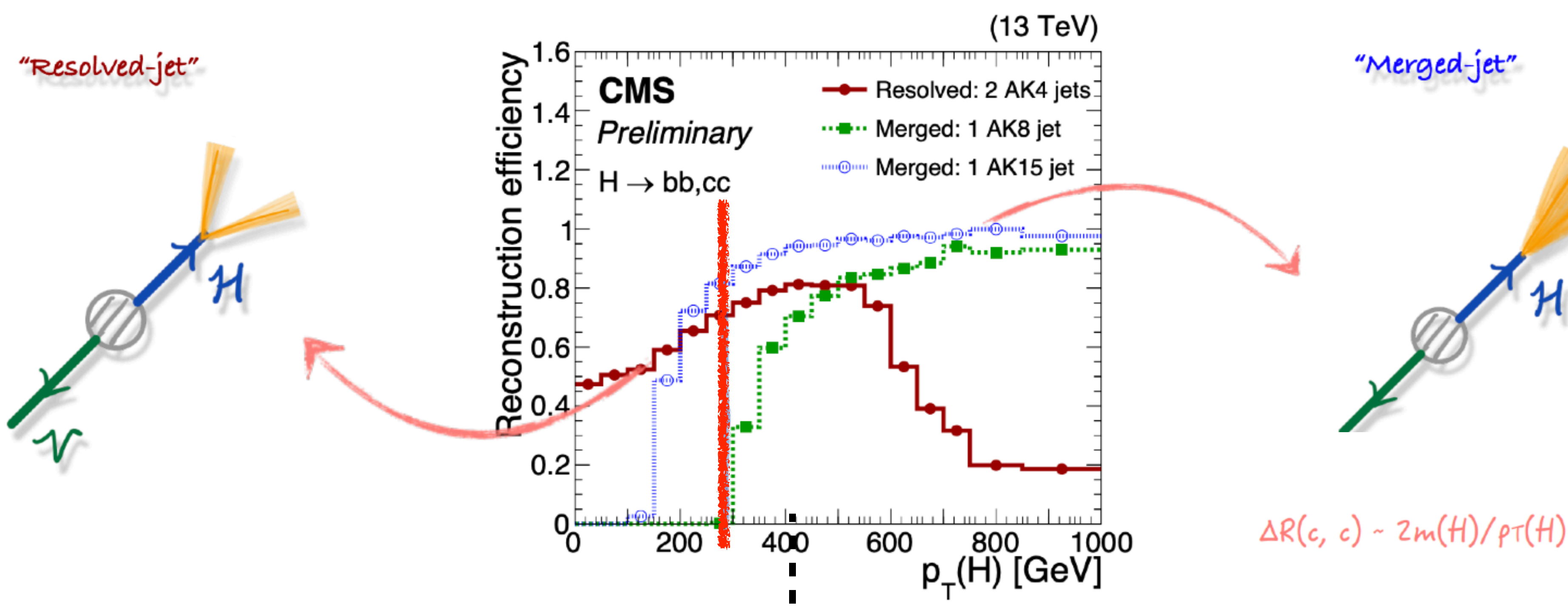


VH H->cc
$\mu = -9 \pm 10(stat.) \pm 12(sys.)$

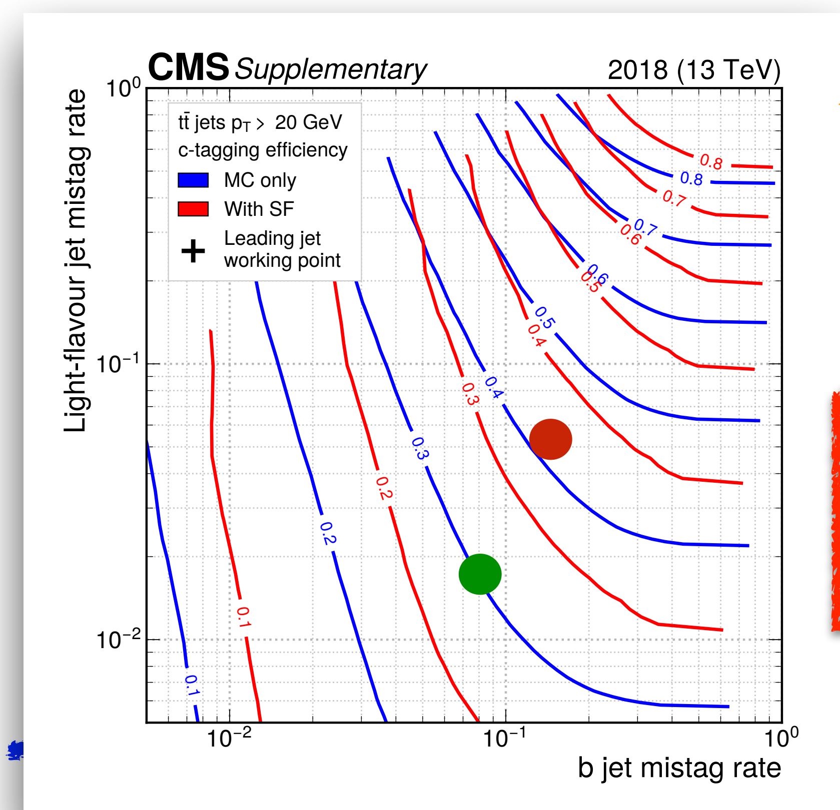
- 95% CL limits:
  - ***observed*** *(exp.)* ***σ*BR/(σ*BR)$_{SM}$ < 26*** *(31)*
  - ***observed $|k_c|$< 8.5*** *exp.: $|k_c|$< 12.4*

first time we could make such plot



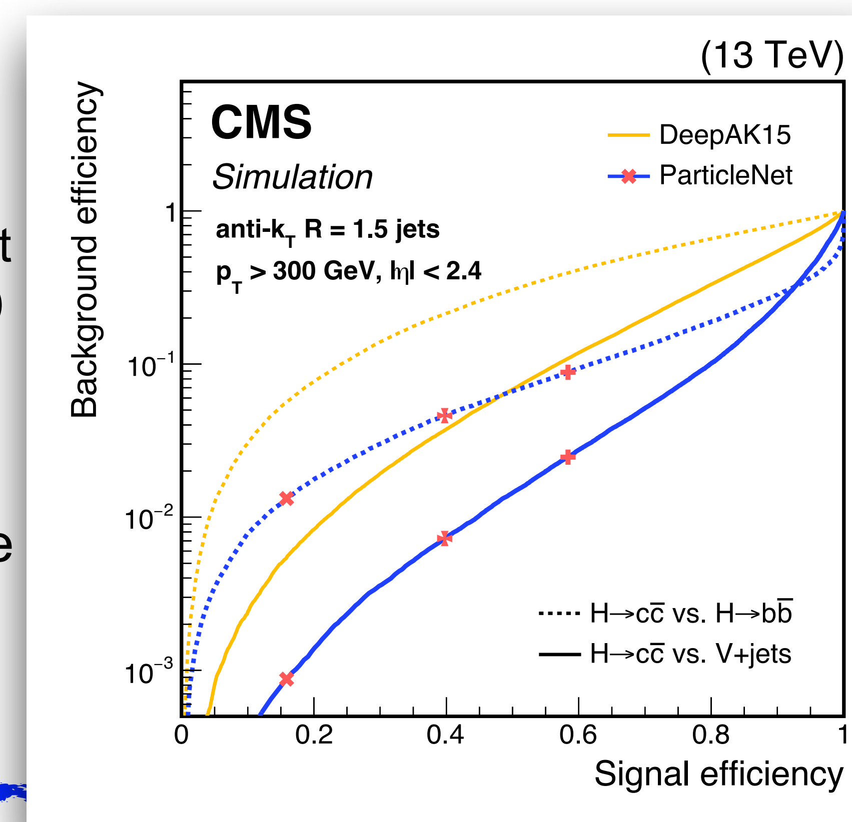


- exploit 2D tagger discriminant in continuous way:
  - *cutting on leading jet quantities to define SR and CR (for V+light, V+bb, ttbar)*



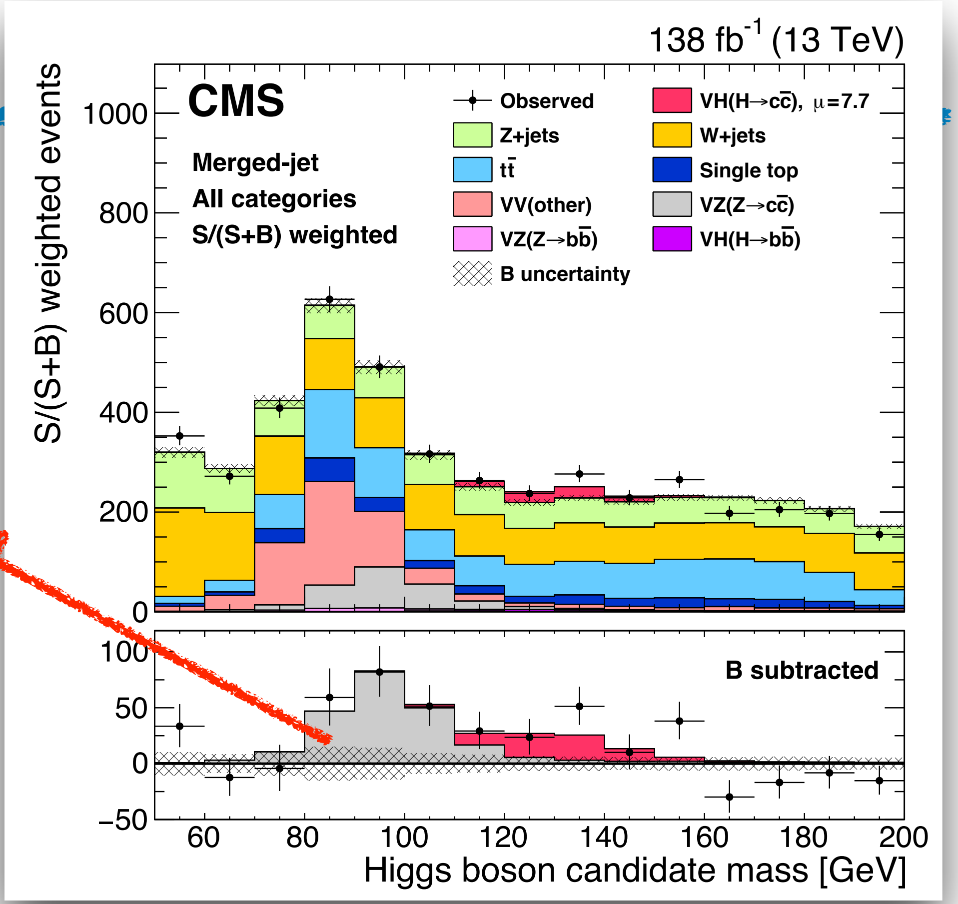
- *feeding rest of information in analysis-level MVA*
- *much looser selection w.r.t. **ATLAS***

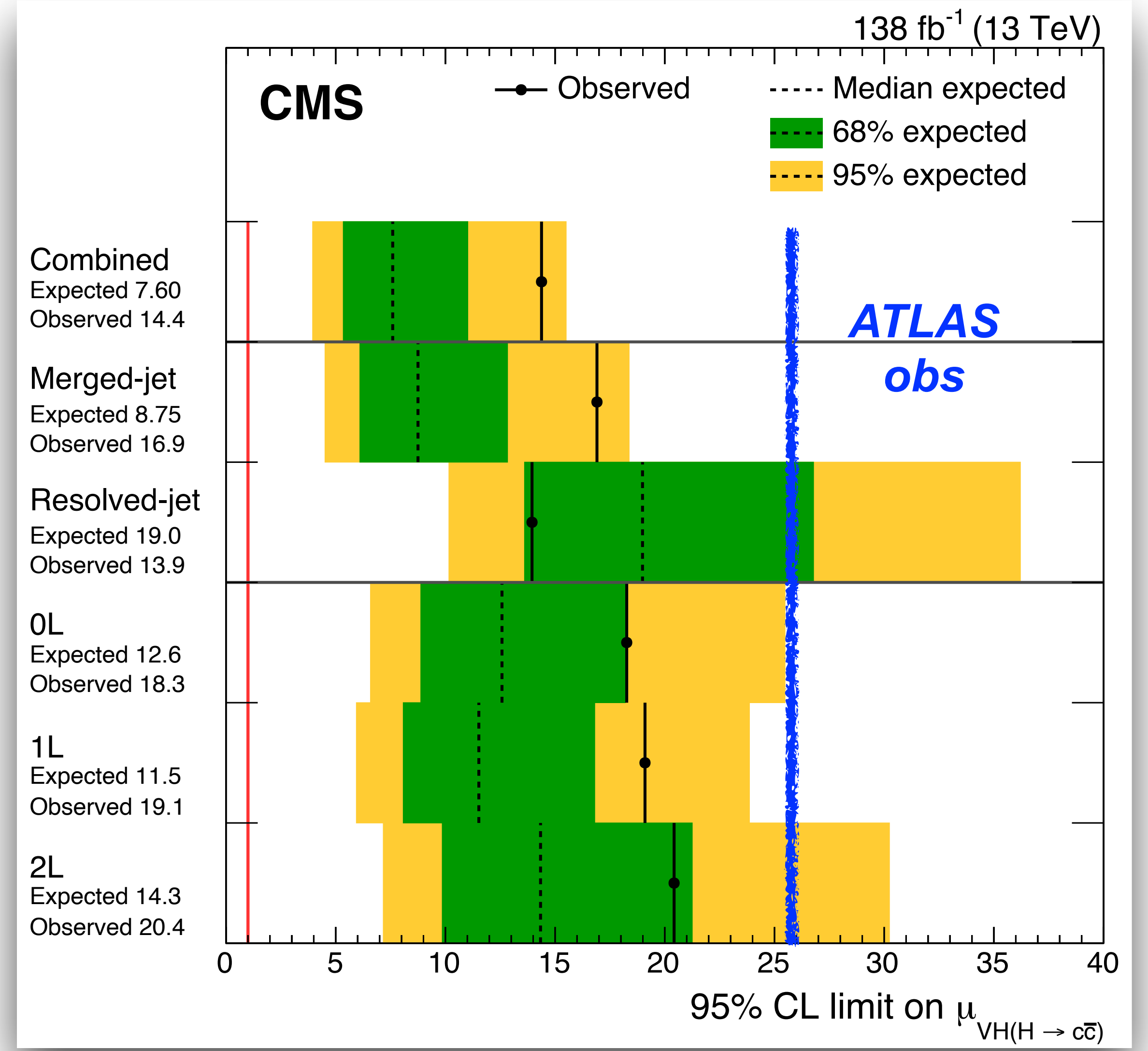
- ParticleNet tagger exploiting low level inputs (particle flow objects) of the entire largeR jet
- trained to avoid mass sculpting of the QCD background
- **impressive performance** (especially against QCD background) by exploiting combination of flavour tagging and substructure techniques



- **Resolved analysis:** fit MVA discriminant in 7 SRs and several CRs
- **Boosted analysis:**
  - *fit largeR jet mass mass*
  - *5 SR x 3 different tagger operating points*
  - *dedicated CR determined with auxiliary MVAs*

VZ, Z->cc
***5.7** (5.9) **σ obs.** (exp.)*



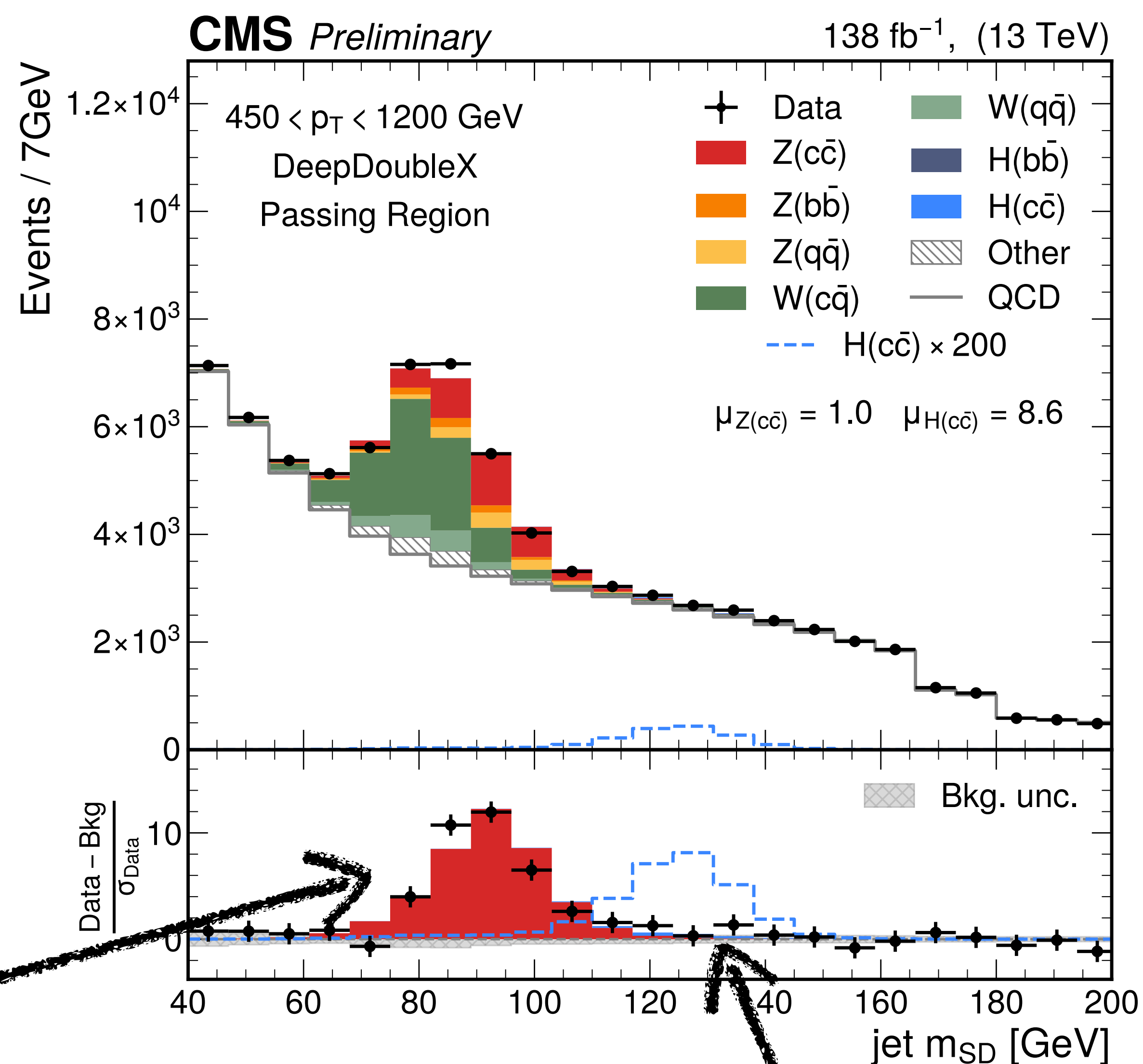


VH H->cc μ = 7.7 ± 3.7

- Large excess in boosted categories, deficit in resolved
- Boosted topology dominates the sensitivity

- 95% CL limits:
  - ***observed** (exp.) **σ*BR/(σ*BR)$_{SM}$ < 14.4** (7.6)*
  - ***observed: 1.1 < |k$_c$|< 5.5***
  - *expected: |k$_c$|< 3.4*

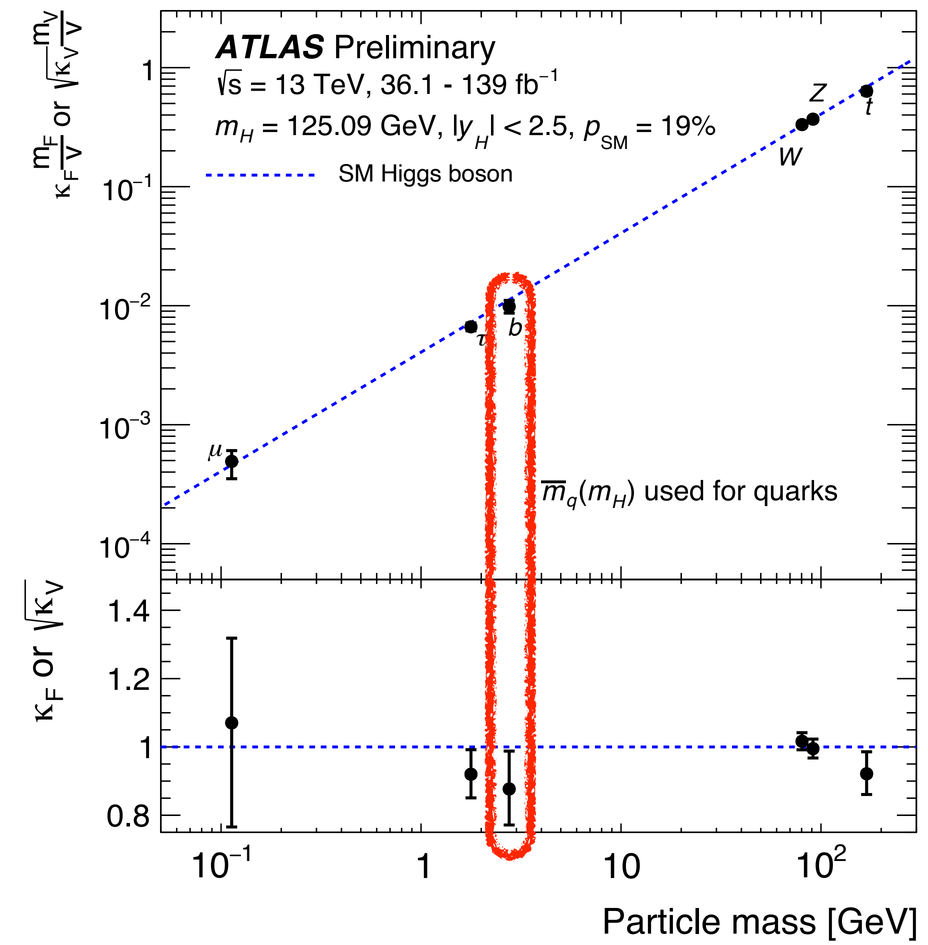
- Another example of the power of largeR jet tagger:



- Z->cc:
  - ***observed at >5 s.d.***
  - ***0.91 ± 0.17 (exp.) ± 0.07 (th.) ± 0.05 (stat.)***
- 95% CL limits for H->cc:
  - ***observed*** *(exp.)* ***σ*BR/(σ*BR)$_{SM}$ < 45*** *(38)*

Summary and Conclusions

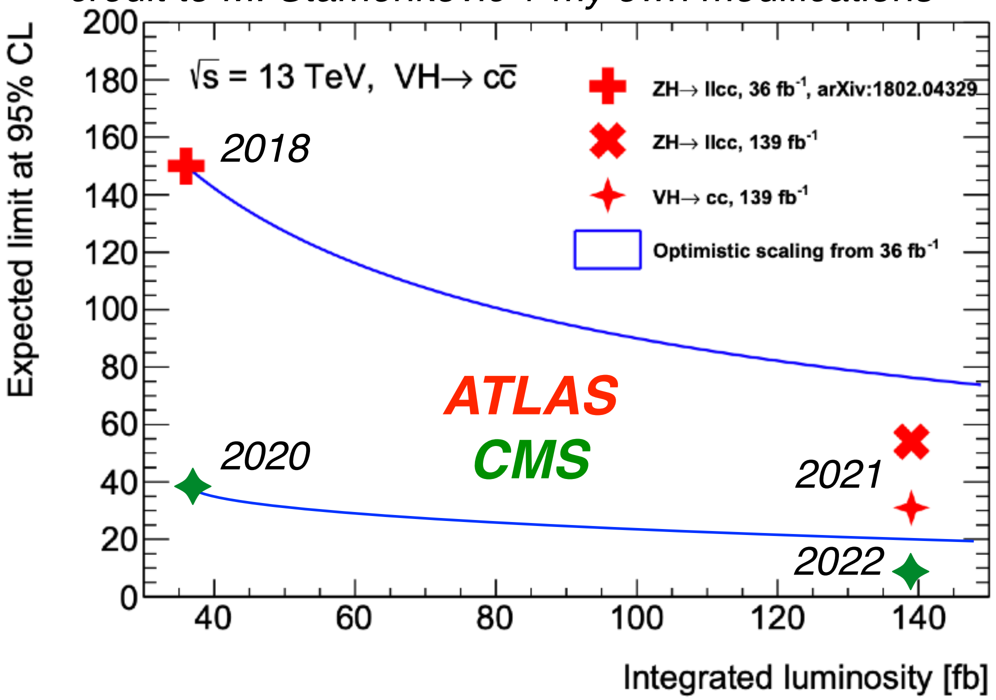
✦ ***H->bb / H->cc analyses have made impressive progresses since the start of LHC***



✦ **H->bb:**

- ✦ *observation achieved (!)*
- ✦ *transition from search to precision measurement in VH*
- ✦ *providing differential information on Higgs Section in the main production modes*
- ✦ *refined boosted techniques to improve high p_T precision*

credit to M. Stamenkovic + my own modifications



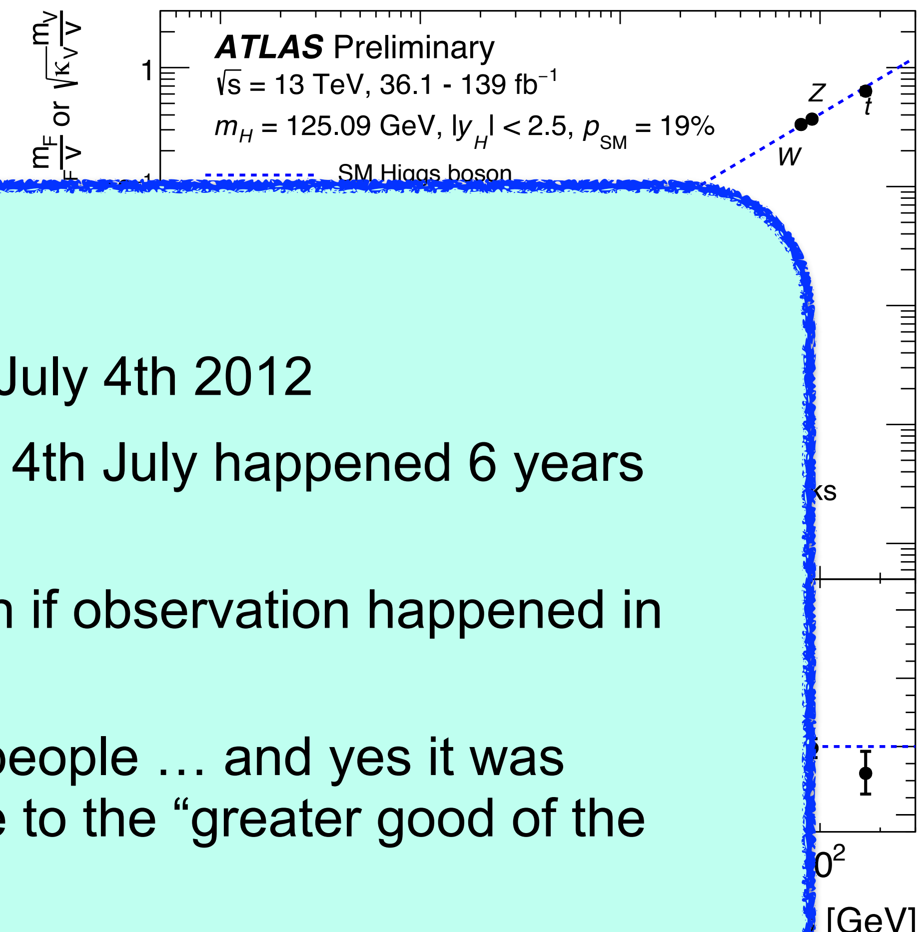
✦ **H->cc:**

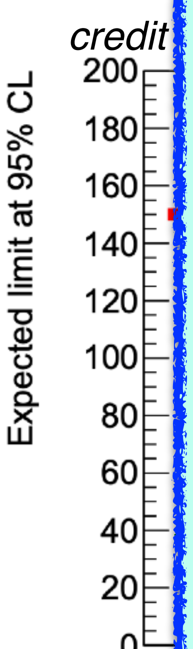
- ✦ *previously unthinkable precision achieved thanks to a mixture of sophisticated algorithms, analysis techniques and available data*
- ✦ *direct constraints on charm coupling are now possible and they are more than competitive with complementary indirect constraints from Higgs pT differential distribution*
- ✦ *charm entry soon to be added to the mass plot….*

!!!! LHC will restart next week: more & better results to come !!!!

- H->bb / H->cc analyses have made impressive progresses since the start of LHC

- **H->bb:**
  - observation achieved (!)



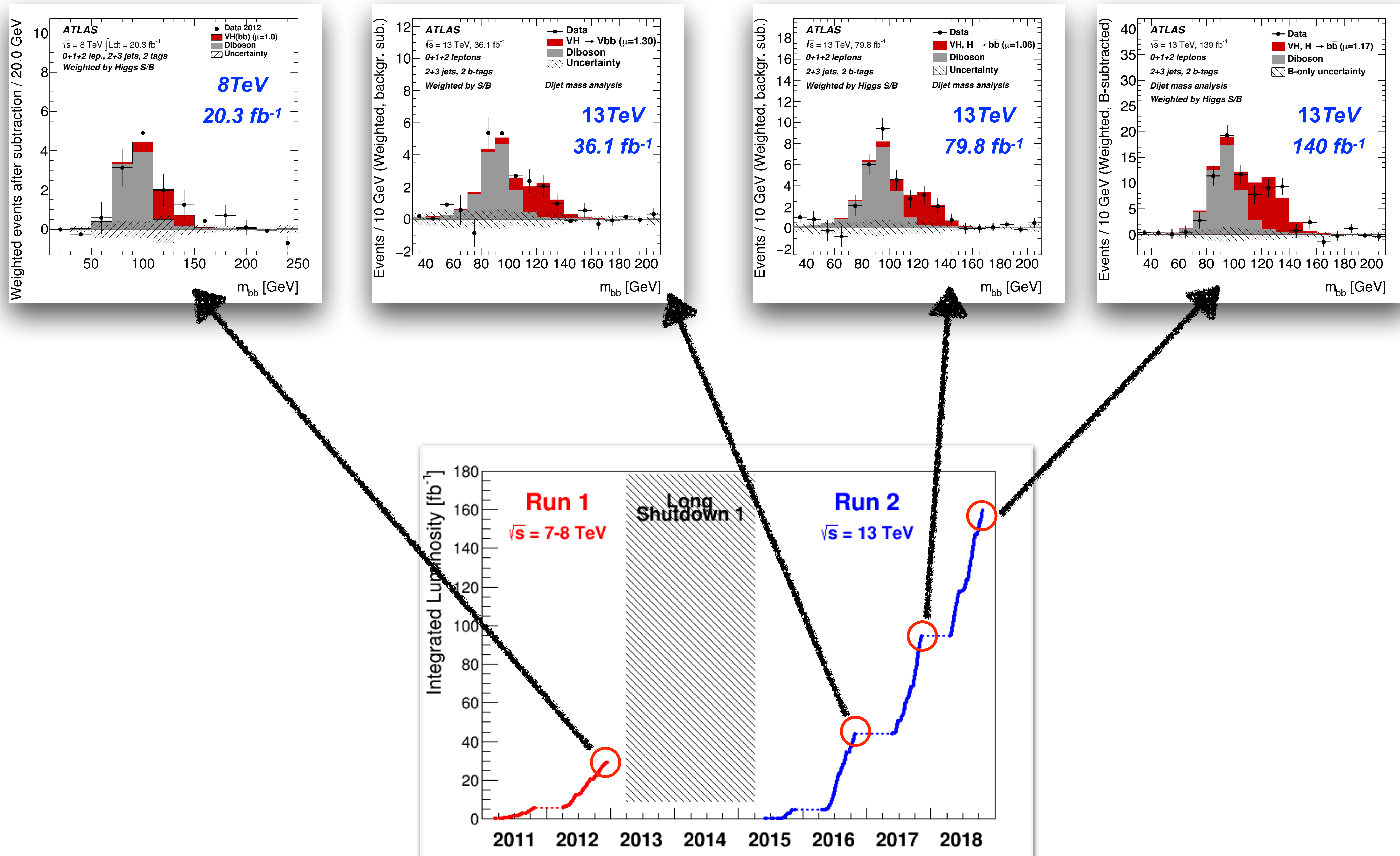


- **Purely personal and ultra-biased considerations:**
  - in the view of the public nothing will be comparable to July 4th 2012
  - …. but likely for many people working on H->bb … the 4th July happened 6 years later ….
  - very long effort with many many contributions …. even if observation happened in Run2, the analysis didn't start from scratch
  - it was a pleasure to collaborate with so many brilliant people … and yes it was really a collaboration with minimal clashes (maybe due to the "greater good of the observation"?)
  - careful balance between revolutionary ideas and incremental improvement: the closer you are to your goal the more attention needs to be put
  - VHbb successfully demonstrated transition from search to measurement …..
  - …….
  - …. other H->bb analyses / VHcc are a bit behind in sensitivity but are catching up fast with ground breaking ideas and methods …. will the focus of the analysis switch to consolidation as they are approaching their target?

!!!! LHC will restart next week: more better results to come !!!!

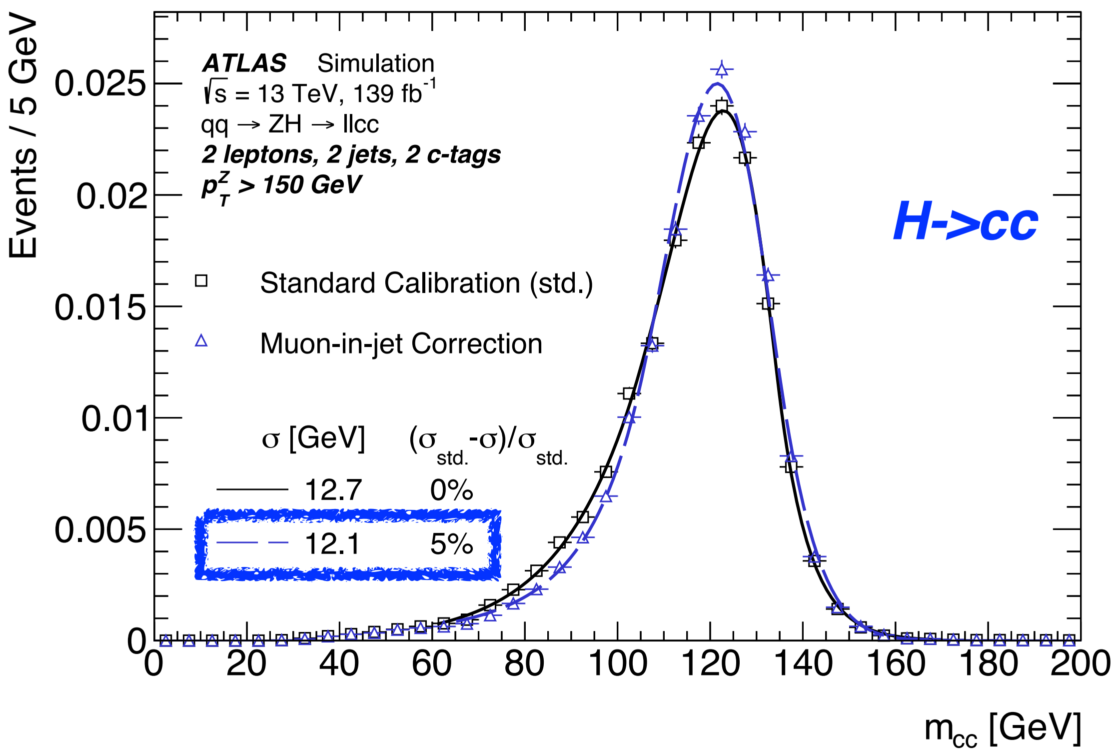
BackUp

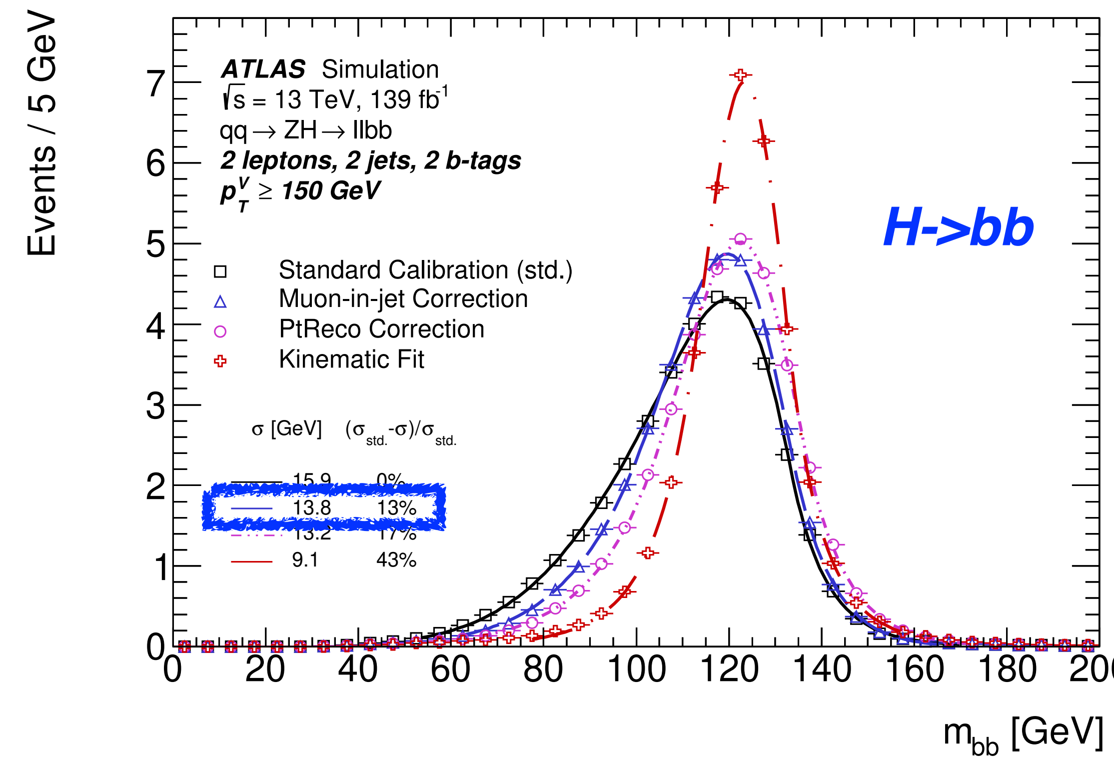
History plot



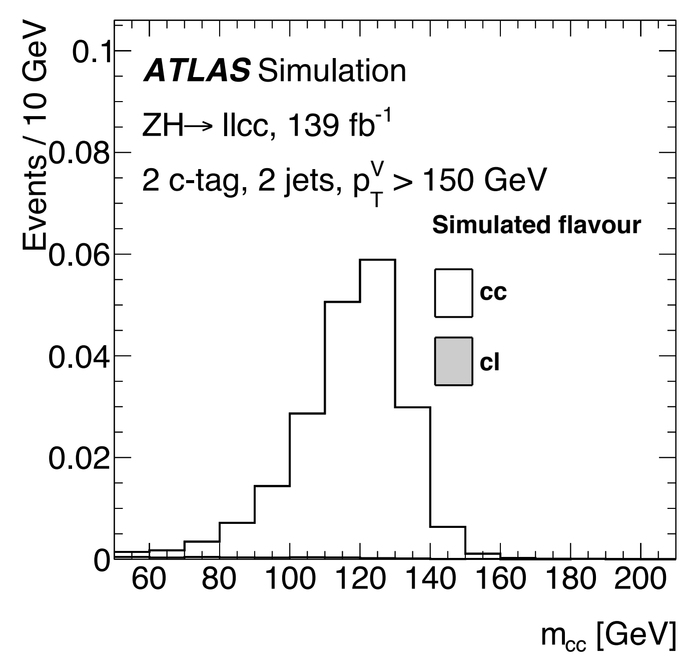
VHcc mass resolution

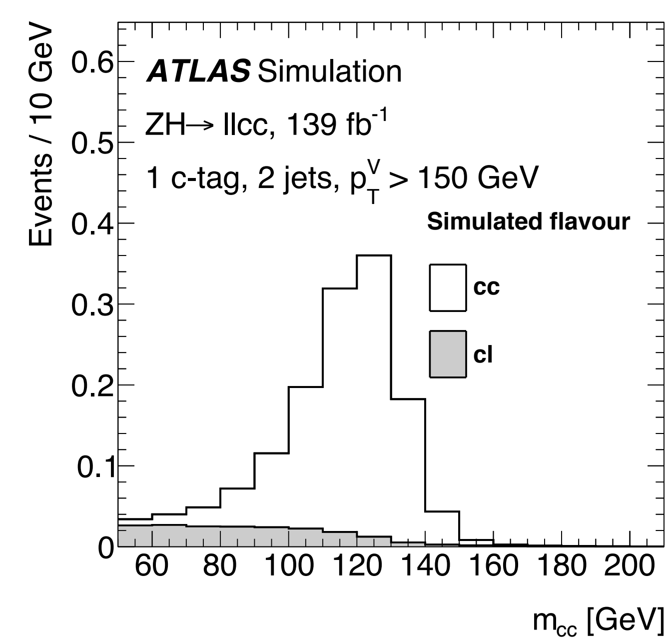
- **Excellent performance of calorimeter-based jet reconstruction and calibration**
  - 10% resolution after re-summing muon founds in jets
  - even better performance than H->bb due to lower rate of neutrinos in inclusive c-hadron decays vs. b-hadrons





- **signal mass resolution depends on the event topology:**
  - worsening from 2->3 jet and 2->1 tag primarily due to combinatorics



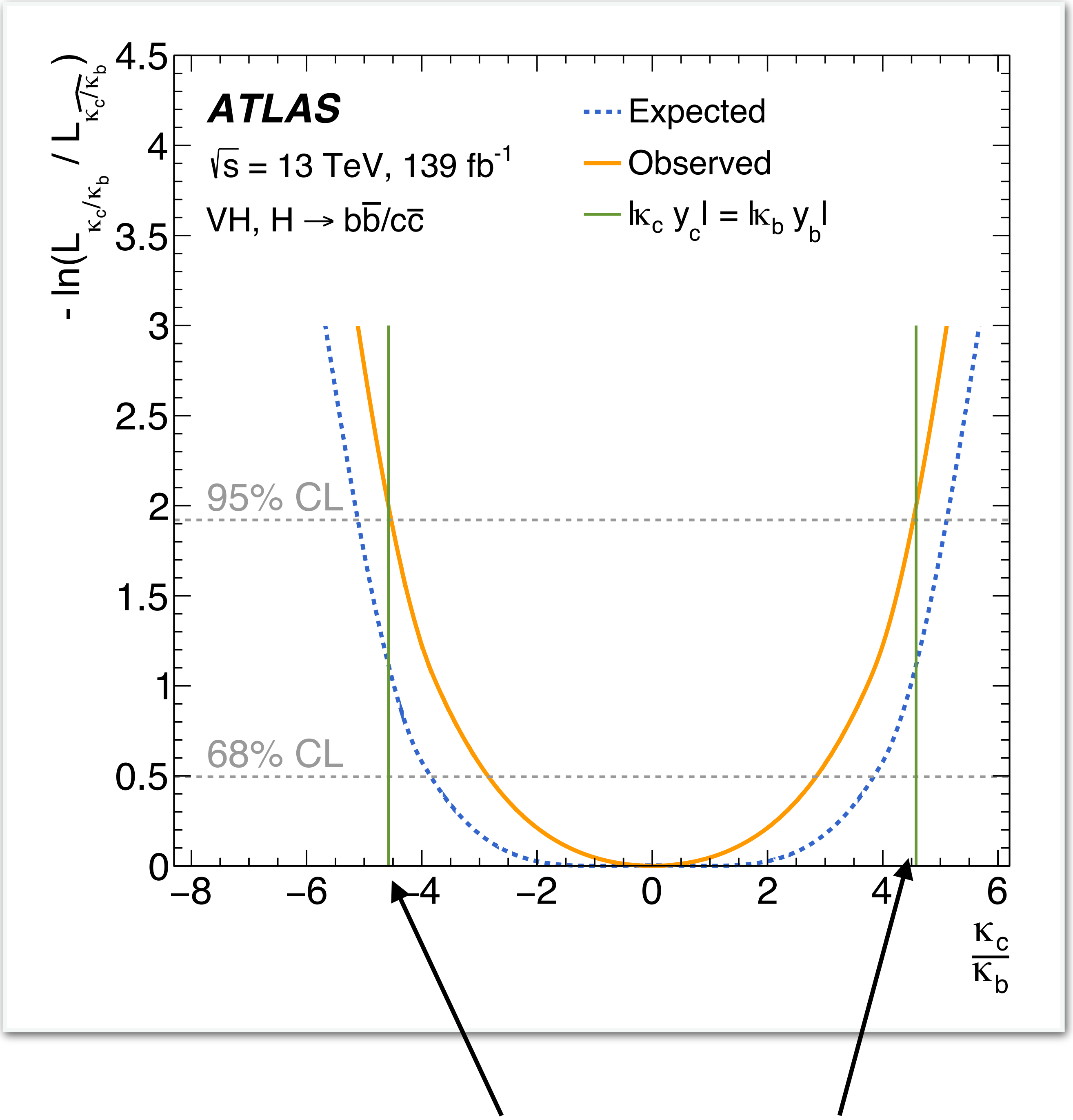


k_c / k_b scan

- **Ratio of coupling modifiers is more model independent reduce the effect of Higgs width** (k_b is also profiled in the fit)

$|k_c/k_b| < 4.5$ @ 95% CL

- **Higgs boson coupling to charm quarks is smaller than Higgs boson coupling to b-quarks**



$m_b/m_c = 4.578 \pm 0.008$

Phys. Rev. D 98 (2018) 054517

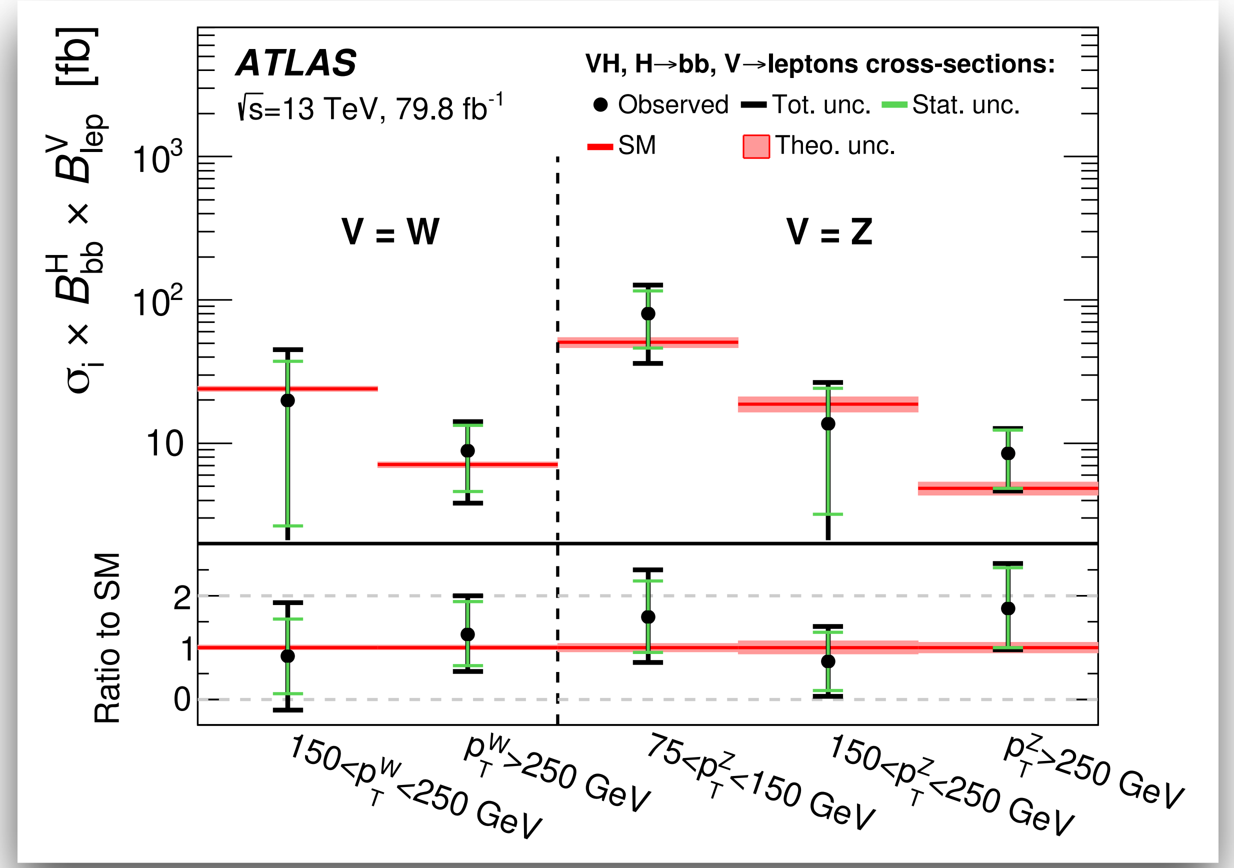
Lattice QCD results

Sys comparison

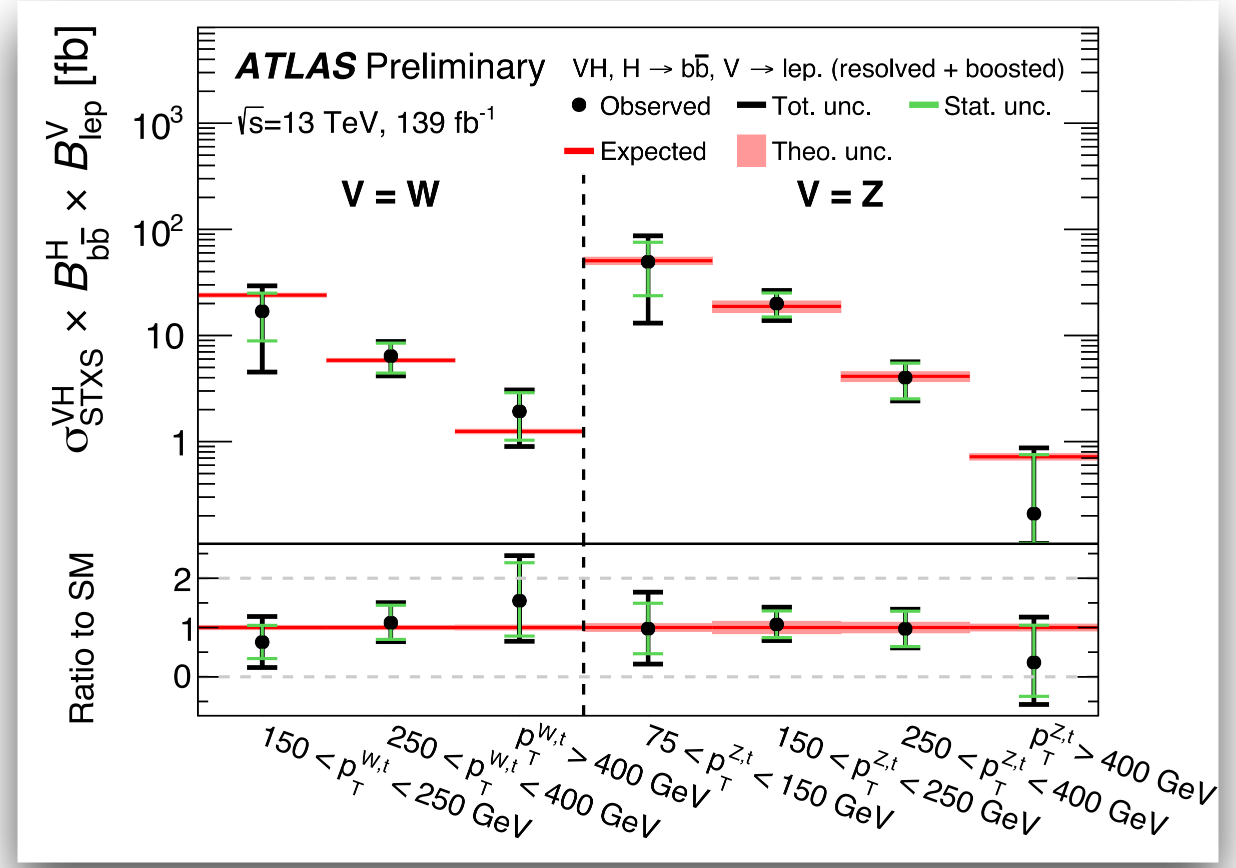
Source of uncertainty		σ_μ
Total		0.259
Statistical		0.161
Systematic		0.203
Experimental uncertainties		
Jets		0.035
$E_{\mathrm{T}}^{\mathrm{miss}}$		0.014
Leptons		0.009
b-tagging	b-jets	0.061
	c-jets	0.042
	light-flavour jets	0.009
	extrapolation	0.008
Pile-up		0.007
Luminosity		0.023
Theoretical 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

Source of uncertainty		σ_μ VH	σ_μ WH	σ_μ ZH
Total		0.177	0.260	0.240
Statistical		0.115	0.182	0.171
Systematic		0.134	0.186	0.168
Statistical uncertainties				
Data statistical		0.108	0.171	0.157
$t\bar{t}$ $e\mu$ control region		0.014	0.003	0.026
Floating normalisations		0.034	0.061	0.045
Experimental uncertainties				
Jets		0.043	0.050	0.057
$E_{\mathrm{T}}^{\mathrm{miss}}$		0.015	0.045	0.013
Leptons		0.004	0.015	0.005
b-tagging	b-jets	0.045	0.025	0.064
	c-jets	0.035	0.068	0.010
	light-flavour jets	0.009	0.004	0.014
Pile-up		0.003	0.002	0.007
Luminosity		0.016	0.016	0.016
Theoretical and modelling uncertainties				
Signal		0.072	0.060	0.107
Z + jets		0.032	0.013	0.059
W + jets		0.040	0.079	0.009
$t\bar{t}$		0.021	0.046	0.029
Single top quark		0.019	0.048	0.015
Diboson		0.033	0.033	0.039
Multi-jet		0.005	0.017	0.005
MC statistical		0.031	0.055	0.038

A Huge leap

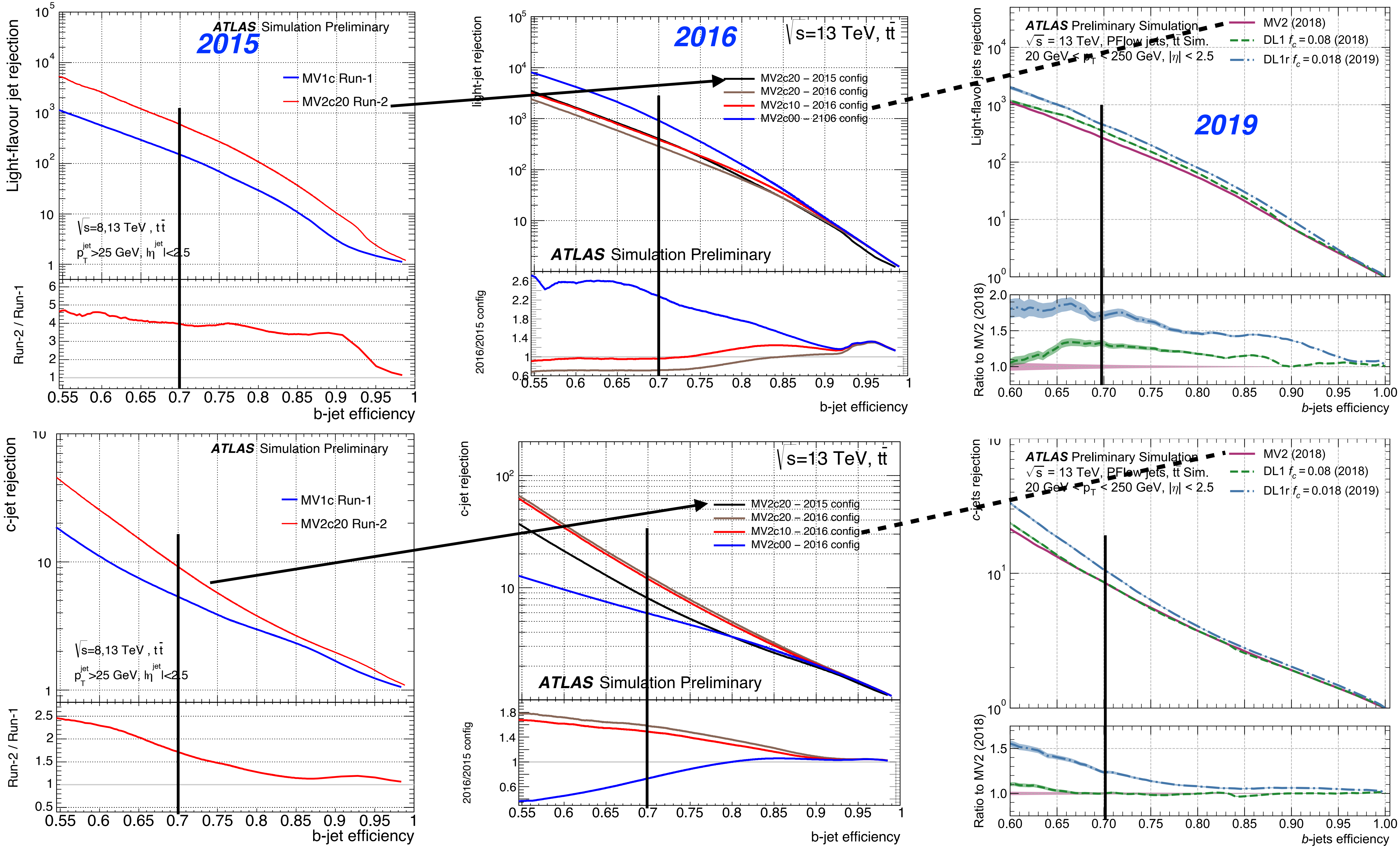


ATLAS
√s=13 TeV, 79.8 fb-1
VH, H→bb, V→leptons cross-sections:
Observed
Tot. unc.
Stat. unc.
SM
Theo. unc.
V = W
V = Z
Ratio to SM
150<pT W<250 GeV
pT W>250 GeV
75<pT Z<150 GeV
150<pT Z<250 GeV
pT Z>250 GeV



ATLAS Preliminary
√s=13 TeV, 139 fb-1
VH, H → bb̄, V → lep. (resolved + boosted)
Observed
Tot. unc.
Stat. unc.
Expected
Theo. unc.
V = W
V = Z
Ratio to SM

constant FTAG improvements



- unfortunately many differences across these plots (jet pt cut, pileup, JVT)
- in general trying to increase c-rejection for similar light rejection