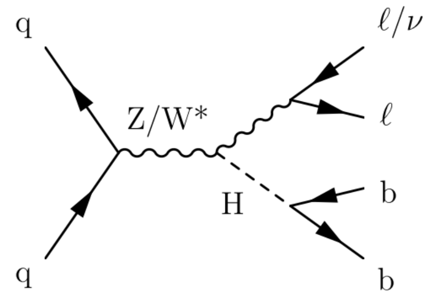


CMS VH generators studies

Luca Perrozzi and Philipp Eller (ETH Zurich)
on behalf of the CMS collaboration

WG1 - VH/VBF subgroup
CERN, June 24rd 2015

Outline

- Summary of Run 1 MC tools and results
- MC studies for Run 2
- Impact of negative weights
- PDF uncertainties
- Points for discussions

VH in CMS: the Hbb case

http://arxiv.org/abs/1310.3687

3 signal channels:

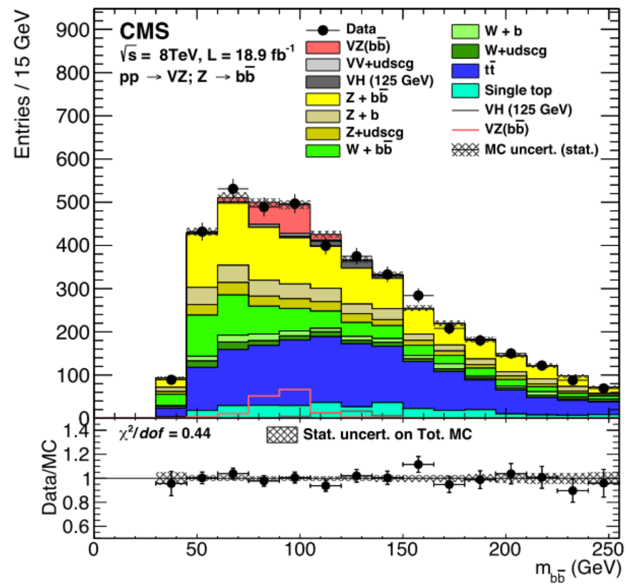
- $Z \to ll$
- $Z \to \nu\nu$
- $W \to l\nu$

3 control regions:

- Z+light jets
- Z+b jets
- Ttbar

Several relevant variables

- $p_{T,V}$
- $p_{T,bb}$
- m_{bb}
- Number of additional jets
- MET
- ΔR_{jj}
- $\Delta\phi_{(V,H)}$



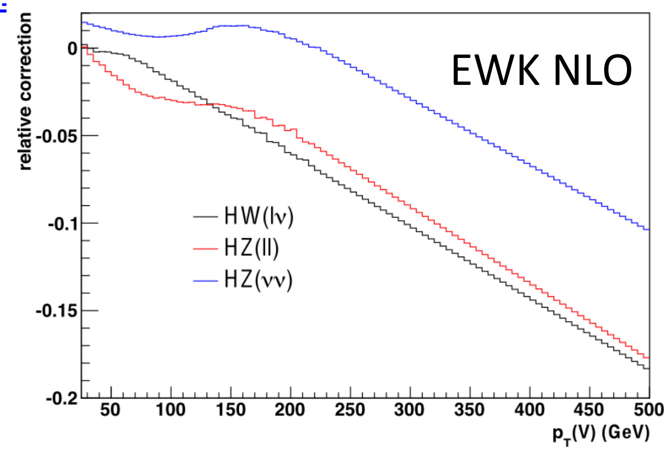
Source of uncertainty	uncertainty (%)	μ_{VH}(%)
Luminosity	2.2-2.6	<2
Lepton efficiency and trigger	3	<2
Jet energy scale	2-3	5.0
Jet energy scale	dist	
Jet energy resolution	3-5	5.9
Jet energy resolution	dist	
b tagging	3-15	10.2
b tagging efficiency	dist	
b tagging mis-identification	dist	
Monte Carlo statistics	dist	13.3
Scale Factors	4.5-13.1	15.9
theory MC normalization	15	5
signal model	11.2-14.7	3.9

Run 1 MC tools

- MC events generated with (N)LO ME+PS

sample	matrix elements	parton shower	PDF
VH	Powheg [71]	Herwig++ [37]	MSTW2008 [72]
VV	Madgraph [27]	Pythia6 [34]	CTEQ6L1 [21]
V+jets	Madgraph	Pythia6	CTEQ6L1
$t\bar{t}$	Madgraph	Pythia6	CTEQ6L1
single-t	Powheg [73, 74]	Pythia6	MSTW2008

- No MC events were available for ggZH
  – Reweighting applied with NLO calculation (JHEP 02 (2013) 078) as function of $p_{T,H}$ and number of jets
- Higher order corrections applied through reweighting
  – QCD NNLO corrections with VHNNLO (PRL 107 152003, 2011 - HEP 1404 (2014) 039 - PLB 740 (2015) 51) provided by authors as function of $p_{T,H}$ for 0j and >0j cases
  – EWK NLO corrections with HAWK (JHEP 03 (2012) 075) provided by authors as function of $p_{T,V}$



Run 2 MC studies

- Several tools being investigated for both signal and backgrounds
  - Different combination of ME, PS, PDF, QCD order, number of matched jets

- Examples for the ZH

ME generators	Parton shower (tune)	PDF	flavour scheme	LO jets	NLO jets	matching/merging (scale in GeV)
Powheg HZ (GoSam)	Pythia 8 (4C)	CT10 (NLO)	5F	-	0	Powheg
Powheg HZJ (GoSam)	Pythia 8 (4C)	CT10 (NLO)	5F	-	0,1	MiNLO
Madgraph/ aMC@NLO	Pythia 8 (CUETP8)	NNPDF (NLO)	5F	-	0,1	FxFx (50)
Madgraph/ aMC@NLO	Herwig 6	MSTW2008 (NLO)	5F	-	0,1,2	FxFx (50)
Amegic/Comix, OpenLoops	Sherpa	CT10 (NLO)	4F	0,1,2	0,1	MEPS@NLO (30)

BDT variables in the signal region

|η|<2.5 and pT > 20 GeV for leptons and jets, pT(Z) > 100 GeV

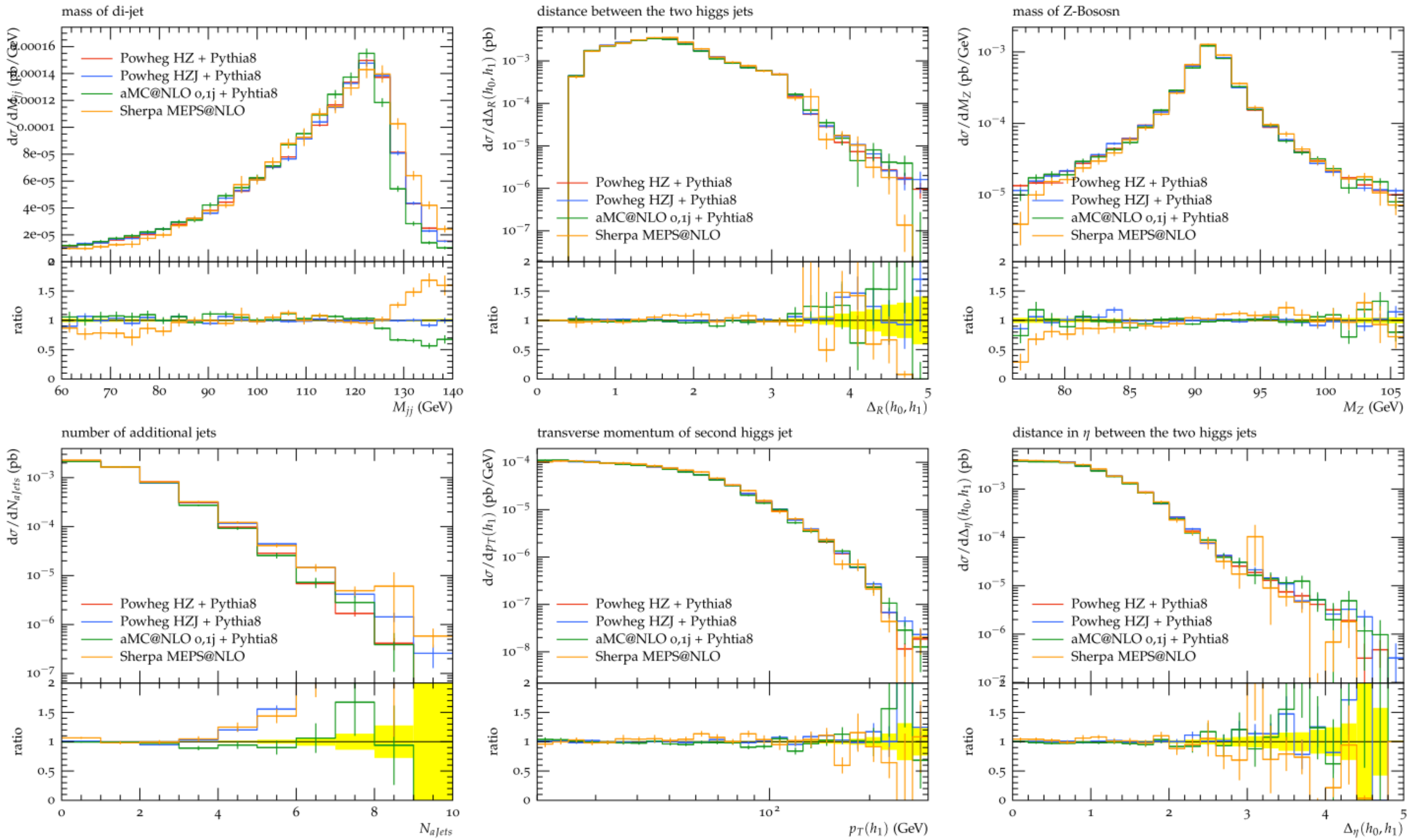


Figure 11.4: $VH(b\bar{b})$ kinematical BDT variables, plus the hardest additional jet and the H-Z p_T distributions in the $VH(b\bar{b})$ selection and their ratio over Powheg HZ, errors include statistical uncertainty only.

BDT variables in the signal region

|η|<2.5 and pT > 20 GeV for leptons and jets, pT(Z) > 100 GeV

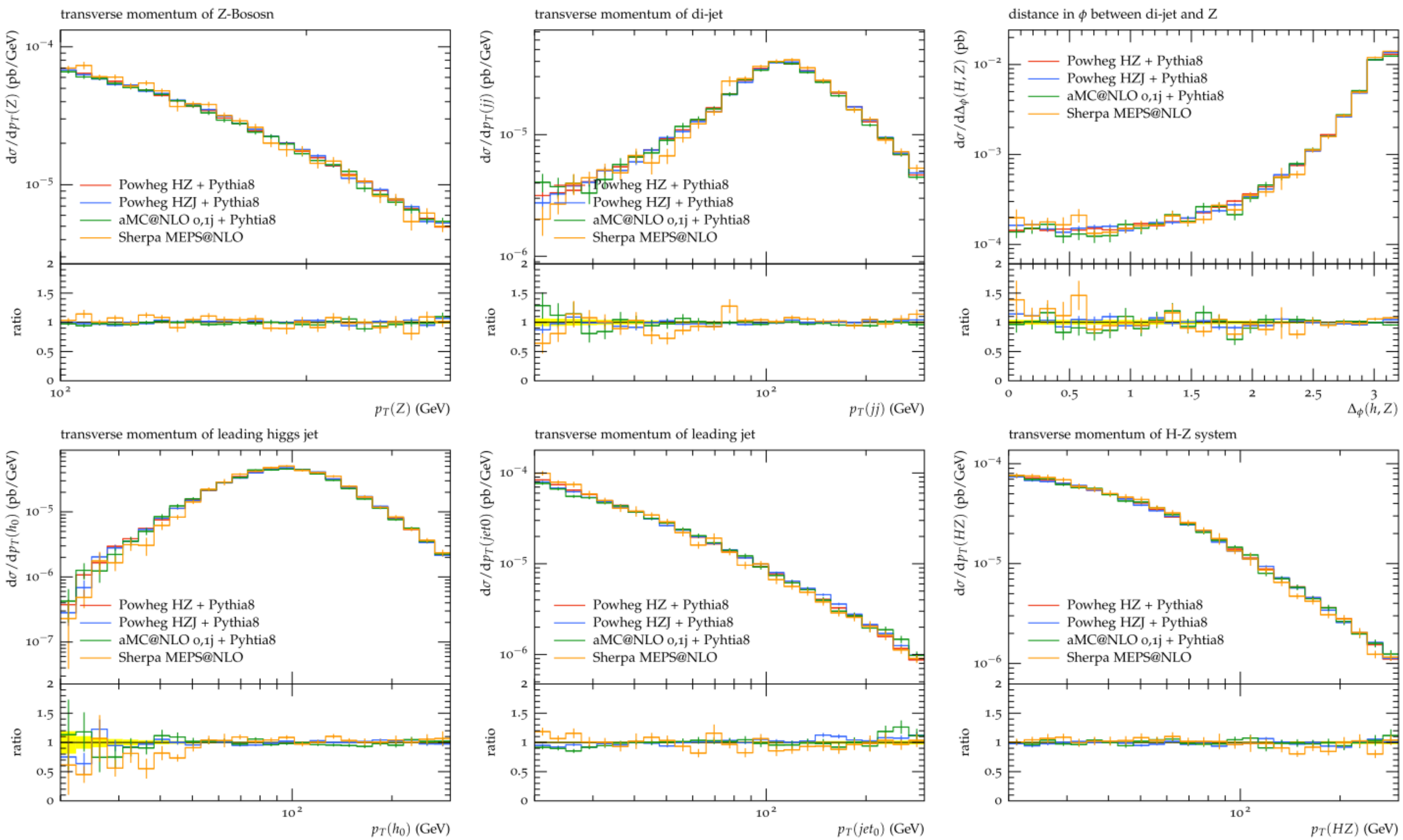
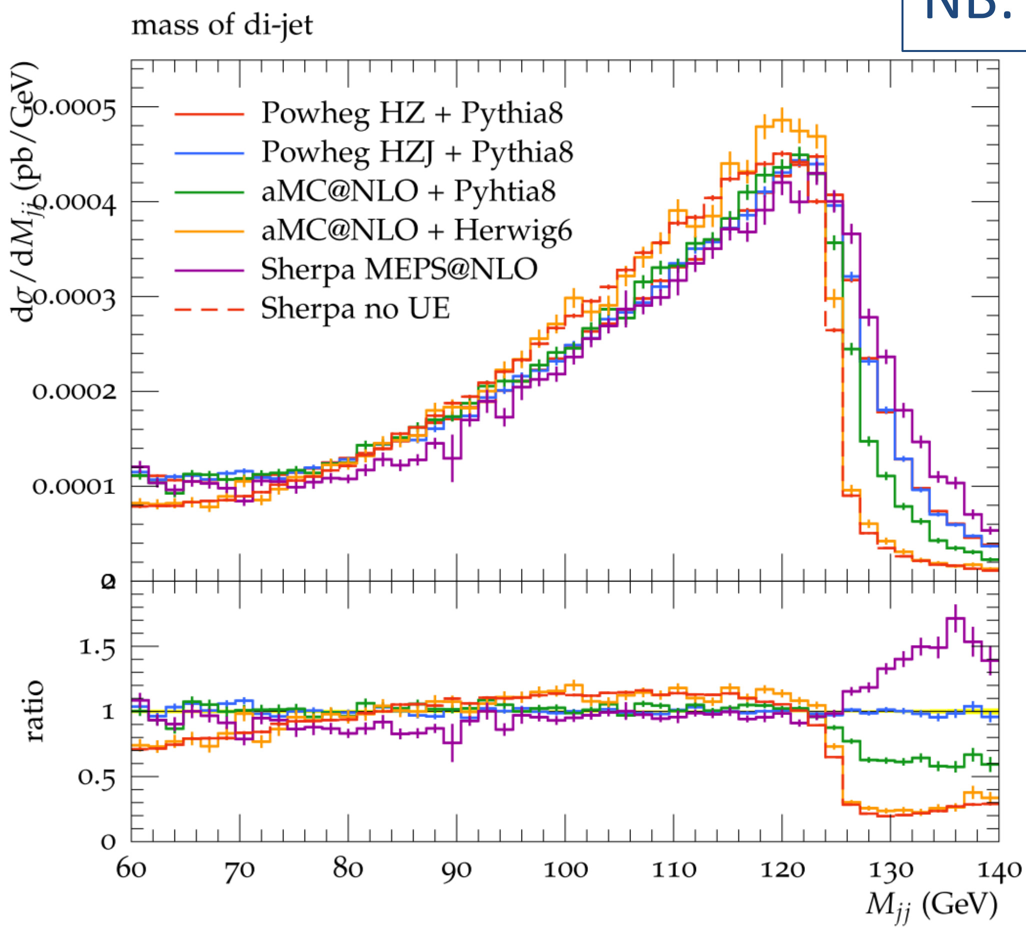


Figure 11.4: $VH(b\bar{b})$ kinematical BDT variables, plus the hardest additional jet and the H-Z p_T distributions in the $VH(b\bar{b})$ selection and their ratio over Powheg HZ, errors include statistical uncertainty only.

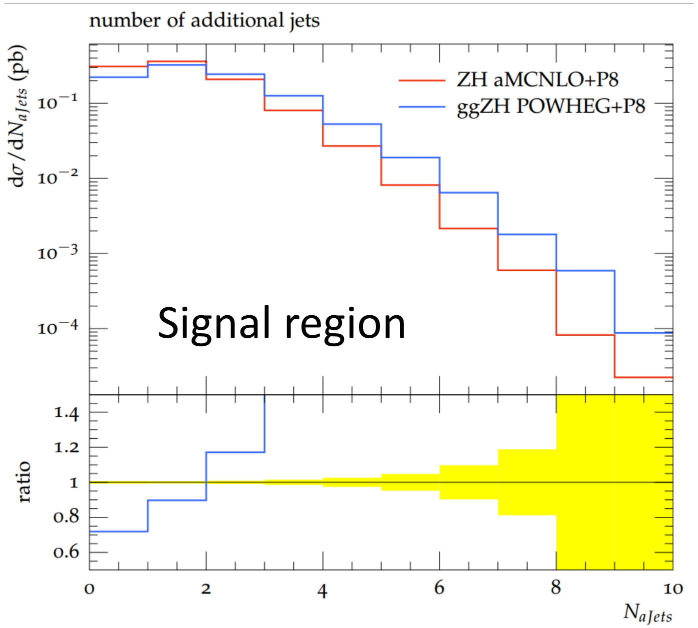
Closer look at the invariant mass

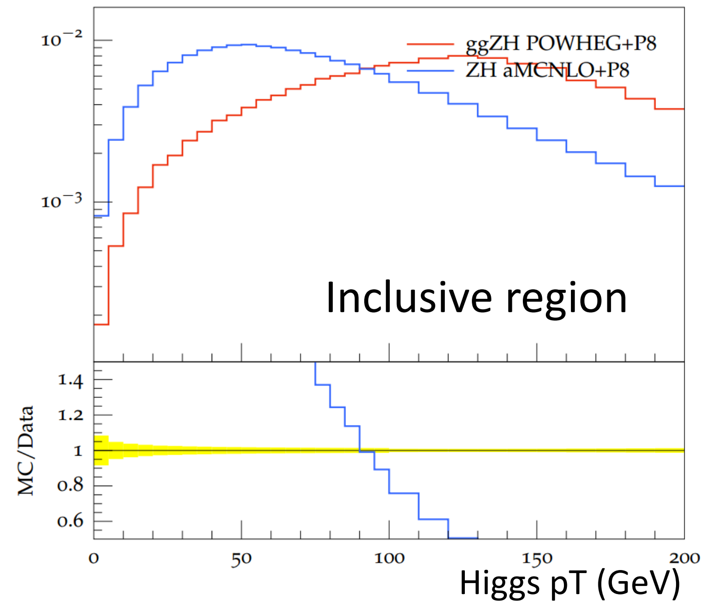
NB: inclusive region



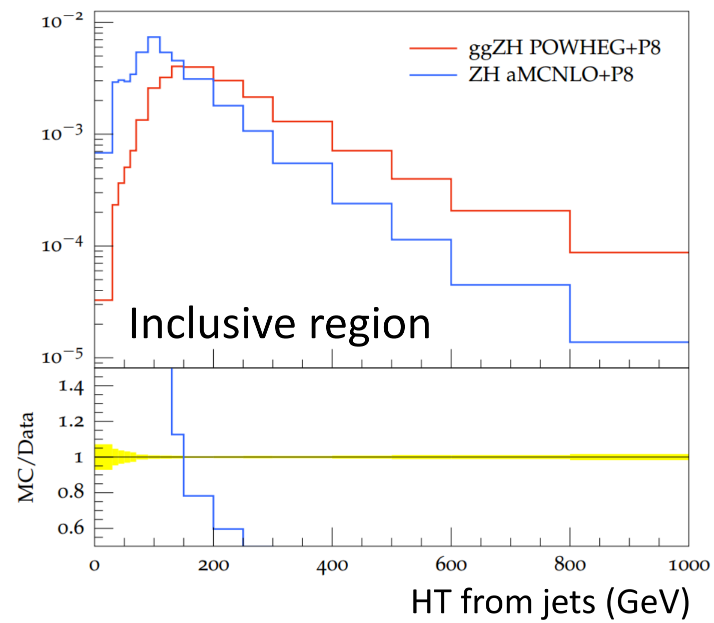
- Differences in the di-jet invariant mass seem to come mainly from the UE settings in the parton shower

ZH and ggZH comparison





- Activity much higher in ggHZ, as previously observed/reported



Default settings for CMS Run 2 MC

- Parton shower: pythia8
  - UE tune CUEP8M1 based on Monash (available as Tune:pp 18)
  - For POWHEG we use the built-in emission veto with power showers
  - For aMC@NLO (NLO) we use wimpy showers along with some other required settings for the mc@NLO matching
  - for LO madgraph with MLM matching, or aMC@NLO with FXFX merging there are additional dedicated configurations of the jet matching/merging
  - Also available/used for pythia8: 4C, Monash, "MBR", CUETP8M1, CUETP8S1-CTEQ6L1, CUETP8S1-HERAPDF
  - Also available/used for pythia6: Z2*
  - Also available/used for herwig++: Tune EE5C, CUETHS1

- PDF: NNPDF 3.0 NLO with α_s=0.118
  - Variations available as replicas and different α_s values (0.117, 0.119)
  - Also available: CT10 NLO + α_s=0.117 + α_s=0.119
  - Also available: MMHT2014 nlo 68% α_s=0.118 + α_s variations

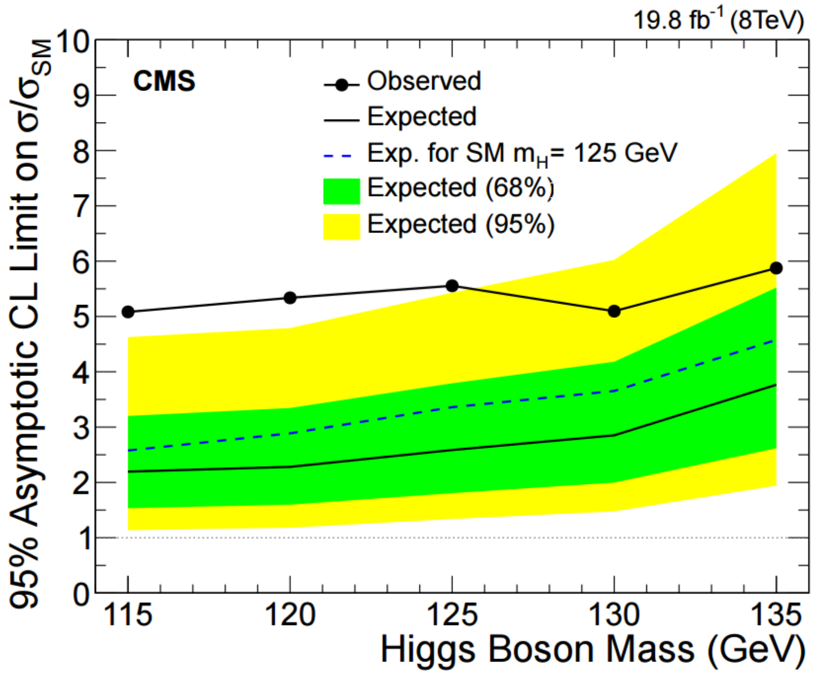
CMS prospects for MC in Run 2

- Pursuing both POWHEG and aMC@NLO for both VH and VBF
  - Preference might fall on the former due to lower negative weight fraction
  - ggZH finally available (also in aMC@NLO!)

- Higher order corrections (reweighting, systematics)
  - Shall we improve/change our QCD and EWK reweightings?

CMS VBF and Hbb grand combination

arxiv:1506.01010



- New analysis just released!

Table 5: Observed and expected 95% CL limits, best fit values on the signal strength parameter $\mu = \sigma/\sigma_{SM}$ and signal significances for $m_H = 125\,\text{GeV}$, for each $H \to b\bar{b}$ channel and their combination.

$H \to b\bar{b}$ Channel	Best fit (68% CL) Observed	Upper limits (95% CL) Observed	Upper limits (95% CL) Expected	Signal significance Observed	Signal significance Expected
VH	0.89 ± 0.43	1.68	0.85	2.08	2.52
$t\bar{t}H$	0.7 ± 1.8	4.1	3.5	0.37	0.58
VBF	$2.8^{+1.6}_{-1.4}$	5.5	2.5	2.20	0.83
Combined	$1.03^{+0.44}_{-0.42}$	1.77	0.78	2.56	2.70

Further topics for discussion

Mostly taken from Les Houches
Run 1 summary talks

http://phystev.cnrs.fr/wiki/2015:programme

Statistical power of a weighted sample (i.e. watch out for negative weights)

4.3 Monte Carlo reweighting (from M. Mulder's PhD thesis)

Another useful analysis technique is the Monte Carlo reweighting approach. The Monte Carlo events that are normally used all have equal weight (=1). By reweighting every generated event with an a posteriori weight w_i, a Monte Carlo sample can be made to 'mimic' a Monte Carlo sample with different generated properties, e.g. a different W mass, provided the statistics are sufficient. In the W mass measurements of the other LEP experiments Monte Carlo reweighting plays a central role [64, 65] (see section 5.1). For the results presented in this thesis it is used for some cross-checks and the final calibration of the W boson width measurement.

The penalty that has to be paid for reweighting is that by giving different weights to events, statistical fluctuations are amplified, and statistical information is not optimally distributed for the newly obtained distributions. It is convenient to express this loss of statistical significance in terms of an *equivalent number* of 'non-weighted' events n_{eqv}:

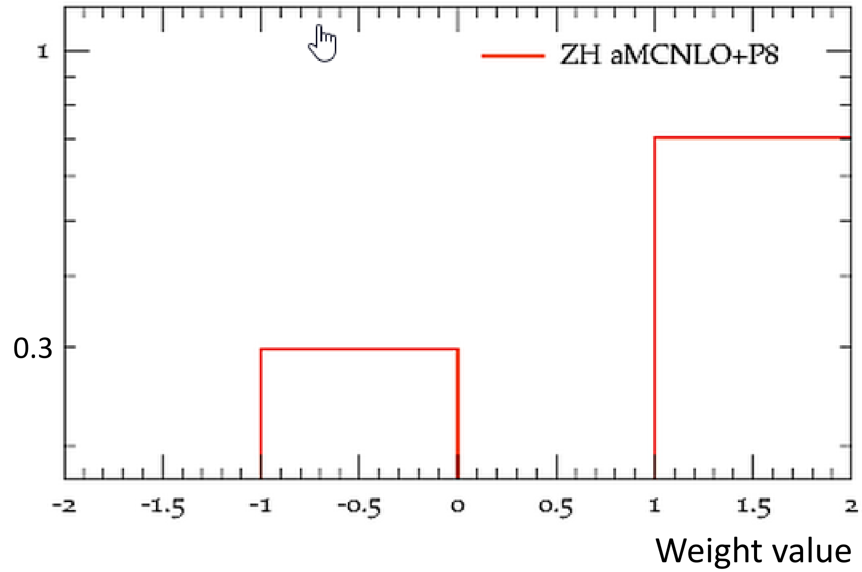
$$n_{\text{eqv}} \equiv \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} w_i^2} = \frac{(n \cdot \langle w \rangle)^2}{n \cdot (\text{var}(w) + \langle w \rangle^2)} = n \cdot \frac{\langle w \rangle^2}{\text{var}(w) + \langle w \rangle^2} \tag{4.15}$$

where $\langle w \rangle$ is the average event weight, and $\text{var}(w)$ the variance. When all weights are equal (no reweighting), the number of equivalent events n_{eqv} is equal to n. Otherwise, the statistical

Statistical power of a weighted sample (i.e. watch out for negative weights)

- According to the formula, the propagation of statistical uncertainty is $N_w = N/(1-2f)^2$
  - N_w is the number of events to be produced with fraction of negative weights f to have a statistical power N

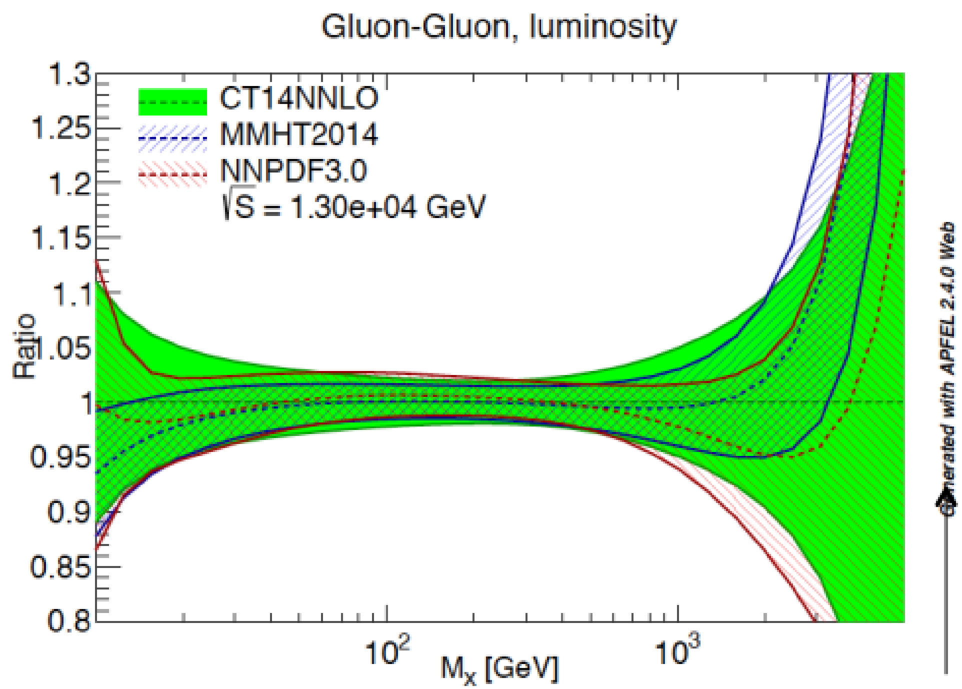
- This translates into:
  - 10% negative weights ---> factor 1.6 more stat
  - 15% ---> factor 2
  - 20% ---> factor 2.8
  - 25% ---> factor 4
  - 30% ---> factor 6.25
  - 35% ---> factor 11
  - 40% ---> factor 25



POWHEG can lower the fraction of negative weights thanks to the so called phase space "folding parameters"
… What about aMC@NLO?

PDFs: the next generation

- NNPDF3.0 (arXiv:1410.8849)
- MMHT14 (arXiv:1412.3989)
- CT14 (on LHAPDF, archive soon)
- HERAPDF2.0 (soon)
- The gg PDF luminosities for the first three PDFs are in good agreement with each other in the Higgs mass range
- PDF uncertainty using the CT14, MMHT14, CT14 PDFs would be 2-2.5%, comparable to new scale dependence at NNNLO, and comparable to the as uncertainty



A very useful tool

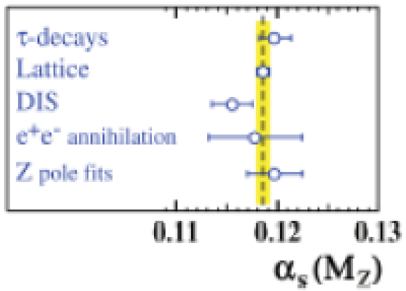
THE VALUE OF α_s

PDG VALUE (AUGUST 2014): $\alpha_s(M_Z) = 0.1185 \pm 0.0006$

to account for perturbative truncation errors

COMMENTS (S.F.)

- LATTICE UNCERTAINTY CURRENTLY ESTIMATED BY FLAG (arXiv:1310.8555) TO BE TWICE THE PDG VALUE (± 0.0012)
- IT IS AN AN AVERAGE OF AVERAGES
- SOME SUB-AVERAGES (E.G. DIS) INCLUDE MUTUALLY INCONSISTENT/INCOMPATIBLE DATA/EXTRACTIONS



- SOME SUB-AVERAGES (E.G. τ OR JETS) INCLUDE DETERMINATIONS WHICH DIFFER FROM EACH OTHER BY EVEN FOUR-FIVE σ
- AVERAGING THE TWO MOST RELIABLE VALUES (GLOBAL EW FIT & τ, BOTH N^{3}LO, NO DEP. ON HADRON STRUCTURE) GIVES

$$\alpha_s = 0.1196 \pm 0.0010$$

NEW PDF4LHC AGREEMENT

- PDG UNCERTAINTY CONSERVATIVELY MULTIPLIED BY 2
- CENTRAL VALUE & UNCERTAINTY ROUNDED: PDF SETS USUALLY GIVEN IN STEPS OF $\Delta\alpha_s(M_z) = 0.001$

$$\alpha_s(M_Z) = 0.118 \pm 0.001$$

-PDFs all evaluated at same value of α_s (0.118).
-α_s uncertainty added in quadrature with PDF uncertainty
-α_s uncertainty is one of the dominant errors now

S. Forte Higgs XSWG meeting June 8, 2015

Updating the PDF4LHC prescription

- We are working on an updated prescription, at NNLO and NLO, using information from CT14, MMHT14, NNPDF3.0, that have similar theoretical treatments/data sets
- We are currently examining two techniques for reducing the number of error PDFs needed
  - Hessian
    - META PDFs
    - MC2Hessian
  - Compression
    - CMC PDFs

Note that measurements should be compared to individual PDFs. Error PDFs derived in this way are useful when a more general definition of the PDF uncertainty is required.

Specialized PDFs can also be made available, i.e. to look at directions sensitive to Higgs physics, W mass, etc.

- See for example the presentation and discussion from PDF4LHC meeting in April
  - https://indico.cern.ch/event/355287/
- …and the one here last Thursday
  - https://indico.cern.ch/event/399439/
- Followup meeting later this month at CERN; paper in preparation

5-flavour vs 4-flavour in Z+b(b)

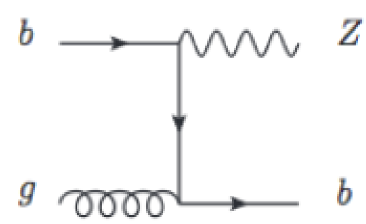
Emerging pattern:

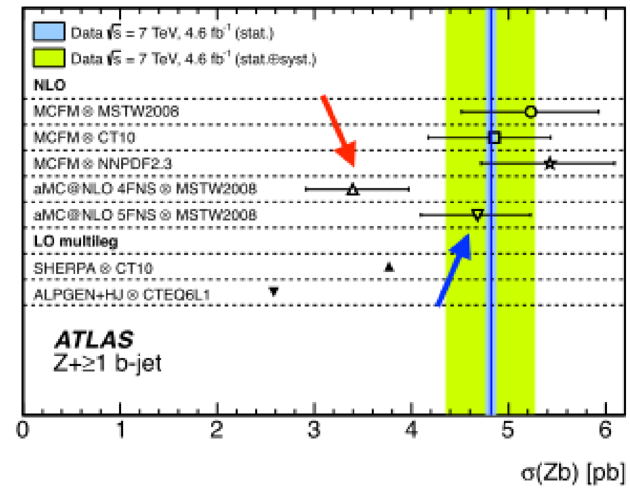
- 5-flavour is better for Z+1b
- 4-flavour is better for Z+2b

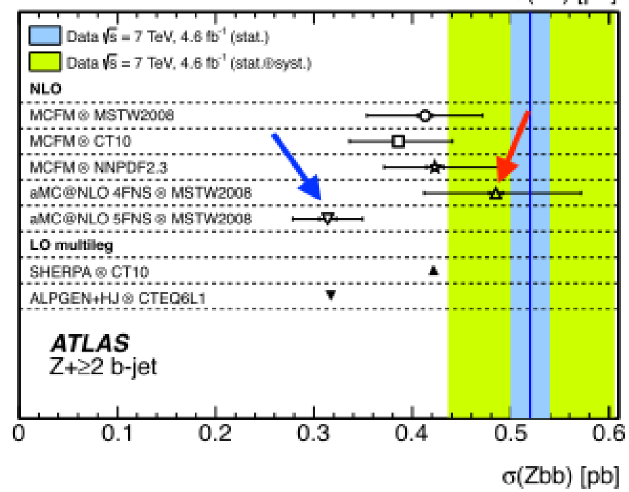
at least for aMC@NLO…

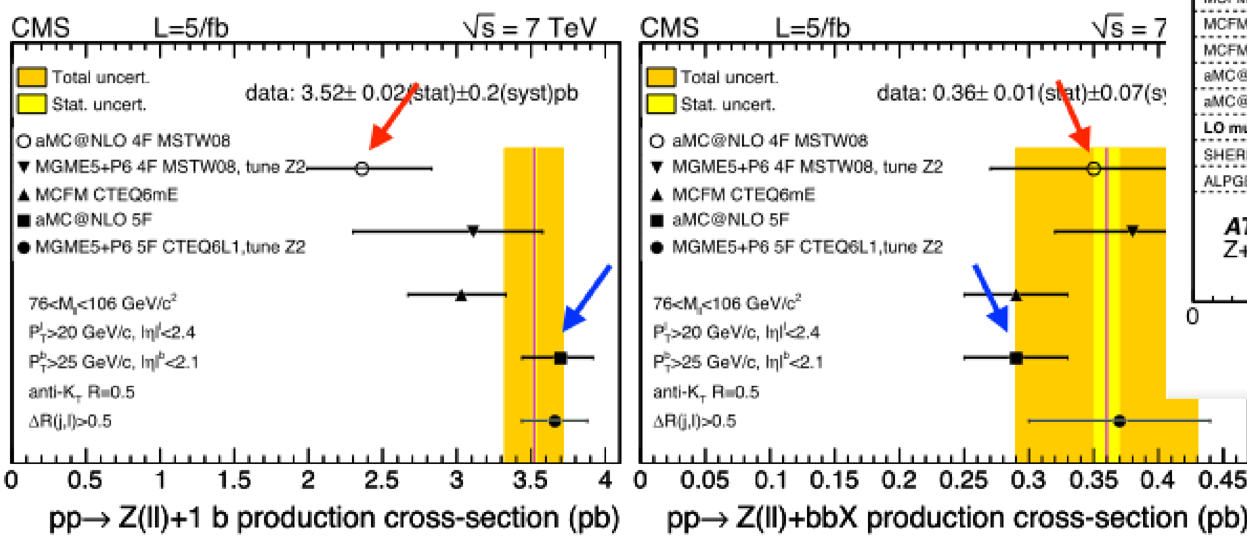
Somewhat reasonable but is it fully understood?

What must be used to evaluate background e.g. for ZH?





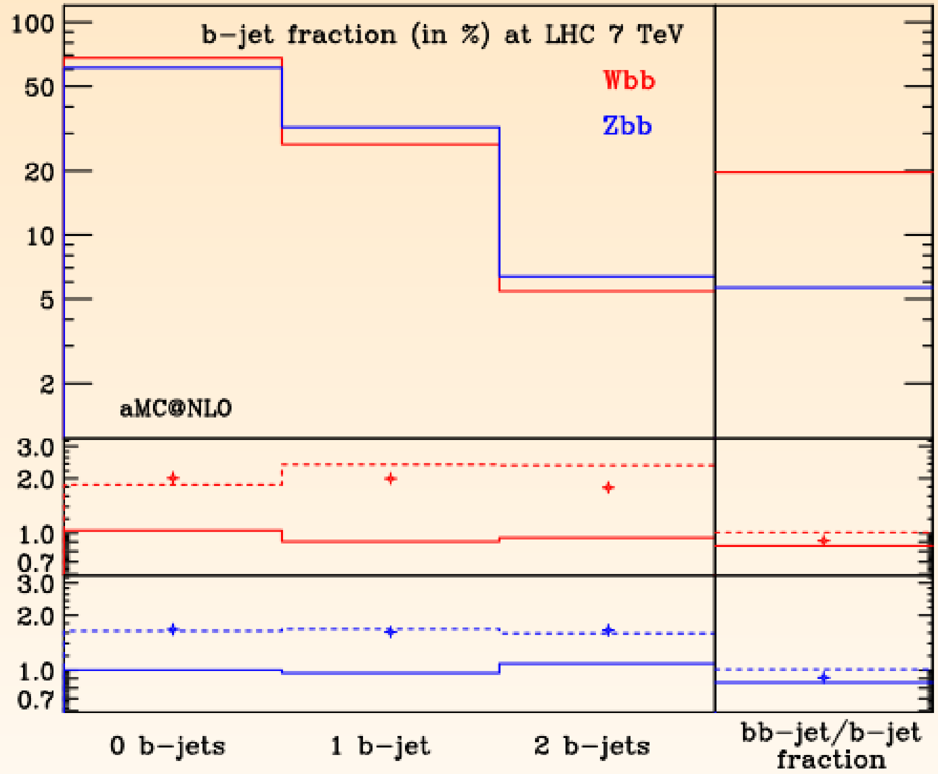


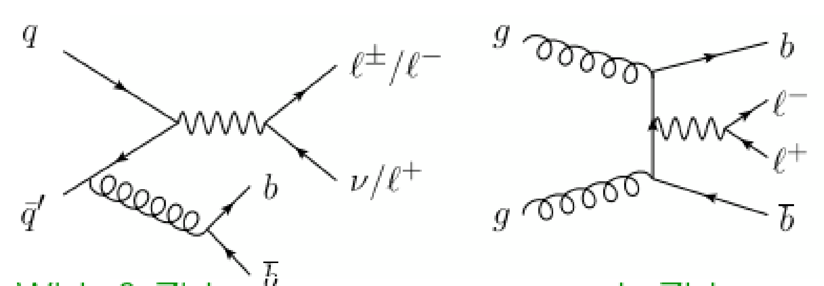


BB-JETS IN 4 FLAVOUR

- Fraction of events containing 0, 1 and 2 b-jets for Wbb and Zbb processes in the 4-flavour scheme, at NLO+PS accuracy
- Important differences due to different kinematic structures
- Not obvious if shower approach (which resums large logs), or fixed order (which take mass effects correctly into account) is the best description

RF et al., 2011



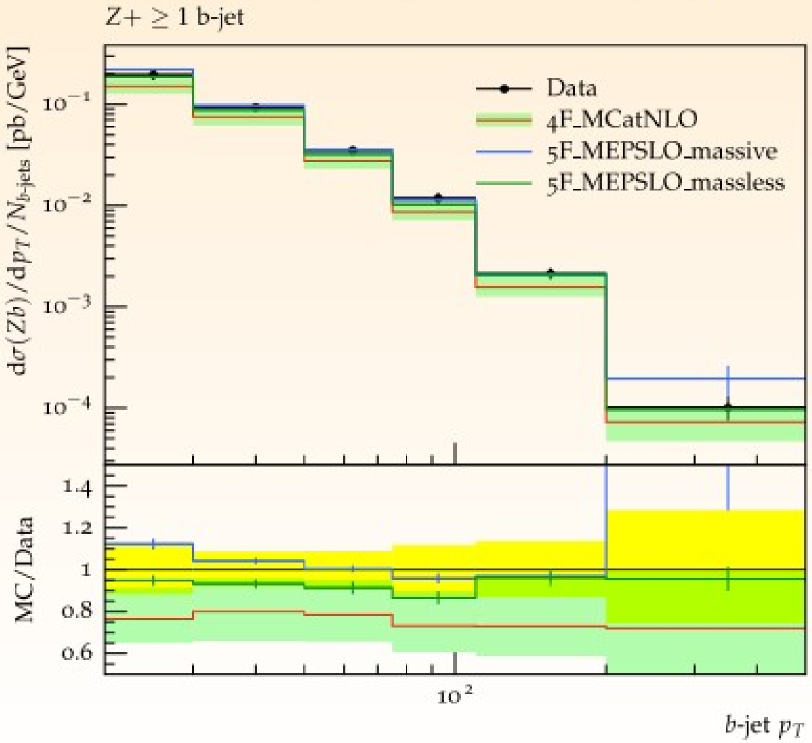


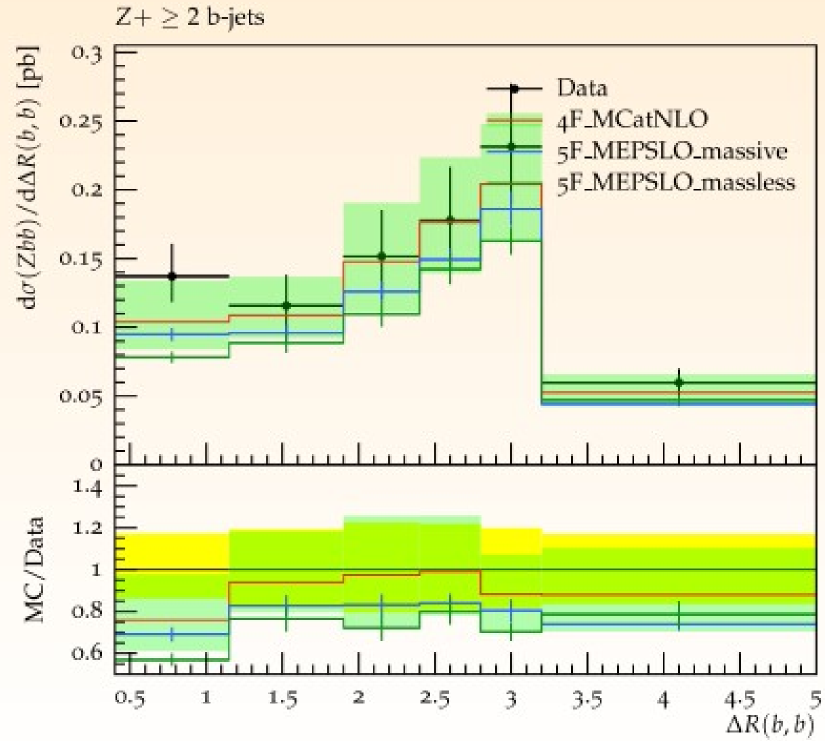
Wbb & Zbb

only Zbb

First results

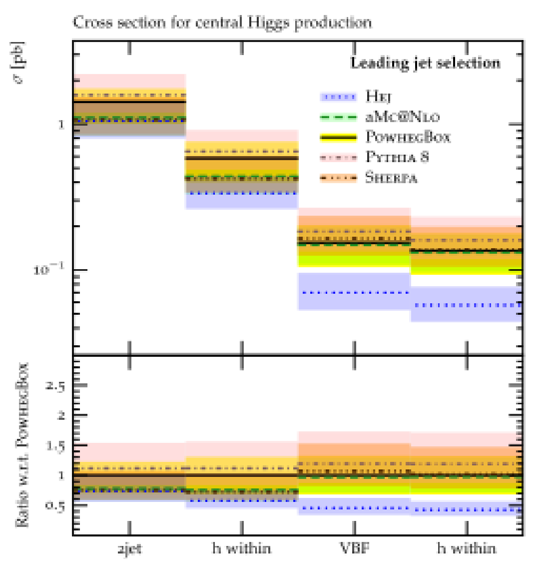
- First results look promising: pT(b-jet) is well-modelled; no significant differences in shape for 5F and 4F
- All predictions agree with each other also at small DeltaR(b,b): no sign for the need of resummation or inclusion of mass effects
  - although maybe undershooting the data a bit there
- Thanks to Davide Napoletano for the plots

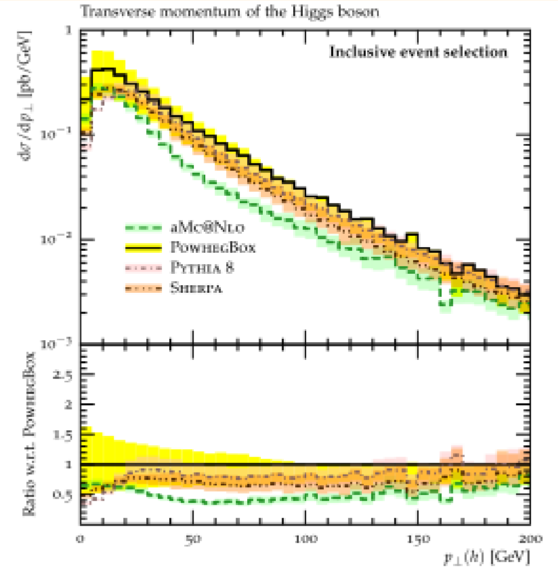




Les Houches, 2013

Higgs+jets

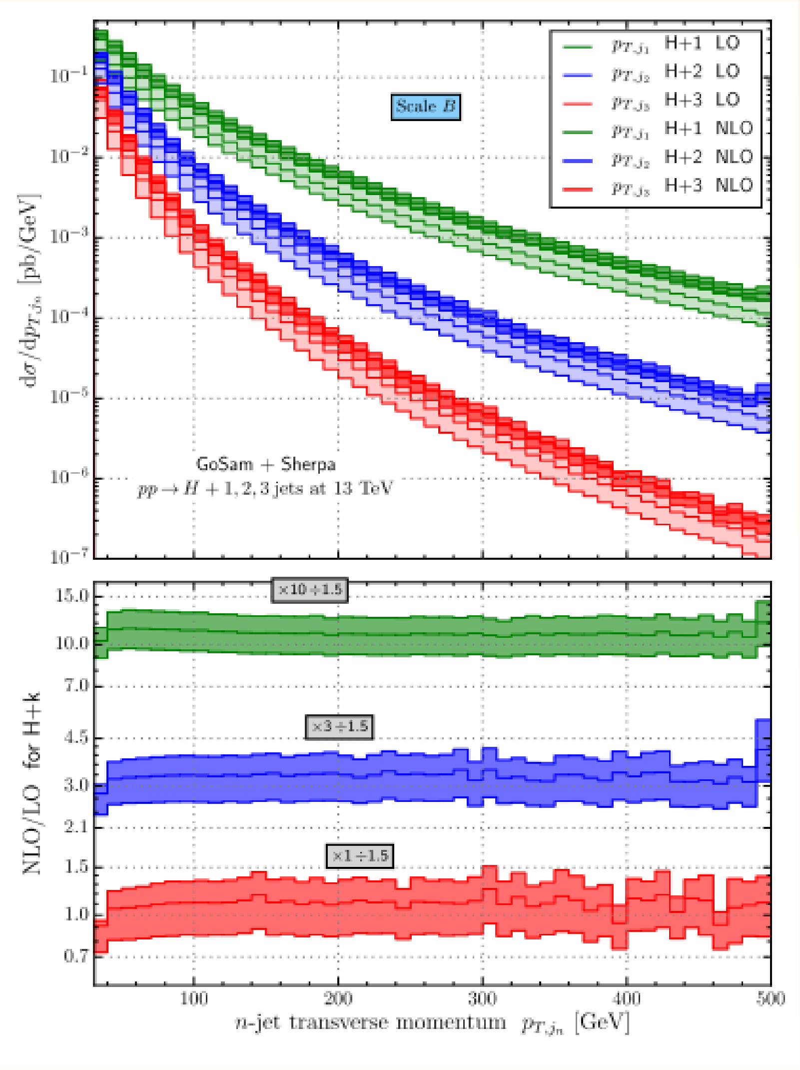




- At last Les Houches, a comparative study for Higgs (+jets) has been made at the NLO+PS level, including merging for various multiplicities
- In general, good agreement has been found between the various codes
- Missing in previous comparison is to compare to higher order calculation with/without (analytic) resummation
- Common project with SM group

Higgs+jets

Greiner et al., 2015



- In the recent years, great improvements have been made in analytic computations
- For example, NNLO for H+j known, NLO H+3j, jet-veto including NNLL resummation, …
- For non-trivial observables it is not obvious which ones can be decently described by fixed order calculations

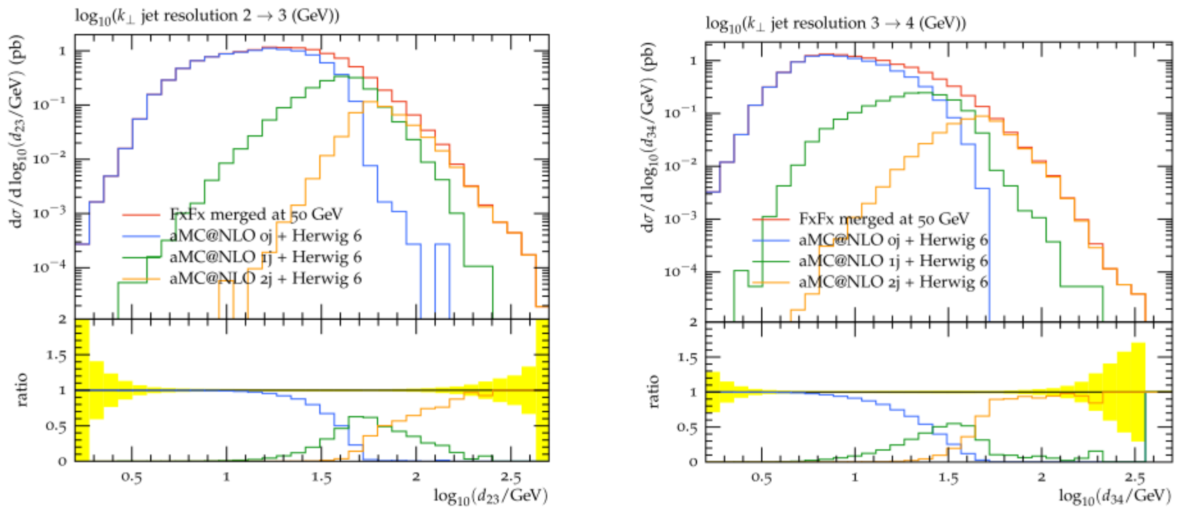
Backup

VH signal and control regions

Table 7.4: Kinematic cuts for the different selection regions (in GeV where applicable). The exclamation marks denote the inversion of a cut.

	Control Regions			MVA signal regions		m_{jj} signal regions			SF fit regions	
variable	Z+light jets	Z+b jets	$t\bar{t}$	low-p_T	high-p_T	low-p_T	medium-p_T	high-p_T	Z+jets	$t\bar{t}$
$p_T(Z)$	>100	-	-	50-100	>100	50-100	100-150	>150	>50	50-100
$p_T(H)$	>100	-	>100	-	-	-	-	-	-	-
$m(Z)$	-	-	!(75-120)	75-105	75-105	75-105	75-105	75-105	75-105	!(75-105)
$m(H)$	-	!(90-145)	-	40-250	<250	40-250	40-250	40-250	!(80-150) & <250	!(80-150) & <250
highest b-tag	0-0.898	>0.898	>0.898	>0.244	>0.5	>0.679	>0.679	>0.679	>0.244	>0.244
second b-tag	>0	>0.5	>0.5	>0.244	>0.244	>0.5	>0.5	>0.5	>0.244	>0.244
#add. jets	<1	<2	-	-	-	-	-	-	-	-
$\not{E}_T$	-	-	-	-	-	<60	<60	<60	-	-
$\Delta R(jj)$	-	-	-	-	-	-	-	<1.6	-	-
$\Delta\varphi(V,H)$	>2.9	>2.9	-	-	-	-	-	-	-	-

Validation of the aMC@NLO merging scale for ZH

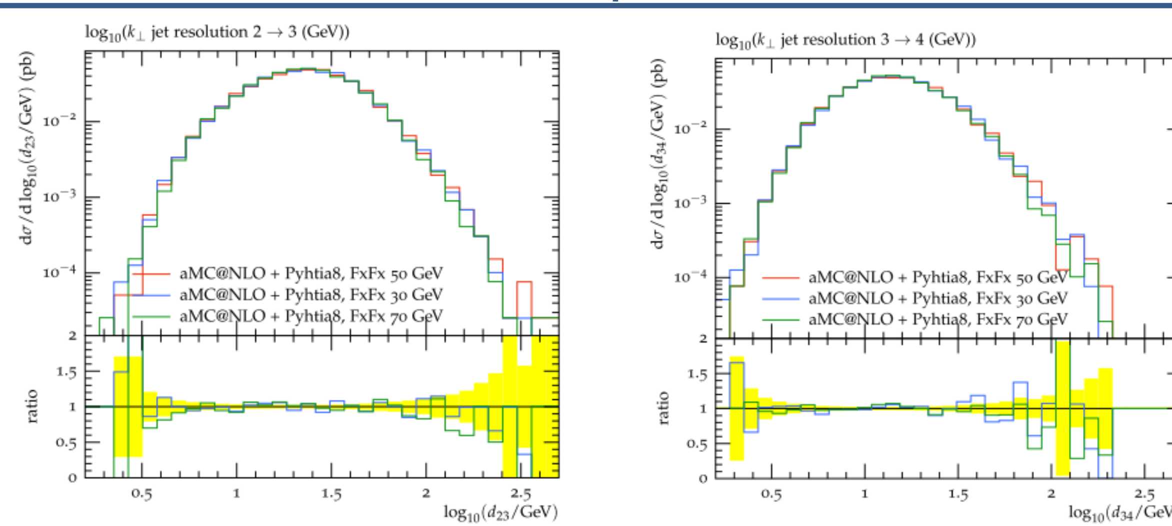


(a) d_{23} (b) d_{34}

Figure 10.1: FxFx merging validation plots for the k_T distances d_{23} and d_{34} of the aMC@NLO + Herwig 6 ZH_{125} sample for a merging scale of 50 GeV. The ratio plot is over the merged sample (sum).

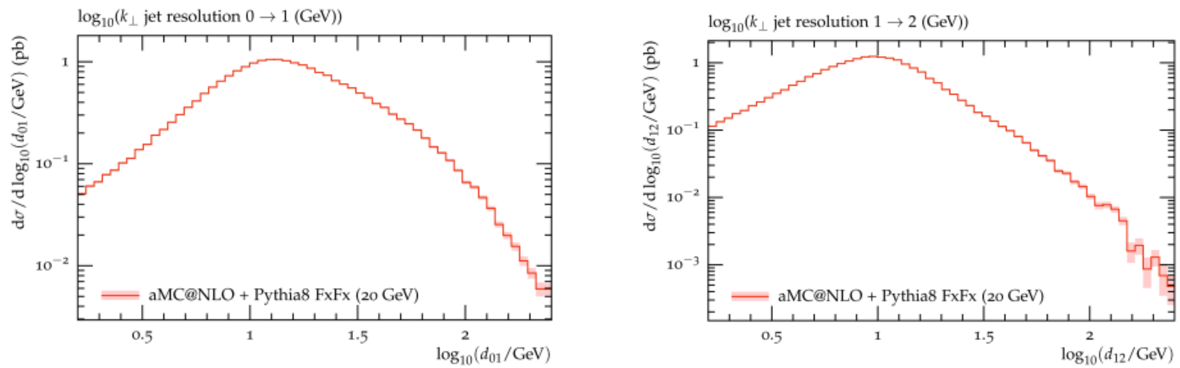
Contributions from the different matched jet samples to the FxF merged sample

Check of different merging scales



(a) d_{23} (b) d_{34}

Figure 10.2: FxFx merging validation plots for the k_T distances d_{23} and d_{34} of the aMC@NLO + Pythia 8 ZH_{125} sample for different merging scales of 30, 50 and 70 GeV. The ratio plot is over the sample with scale 50 GeV.



(a) d_{01}

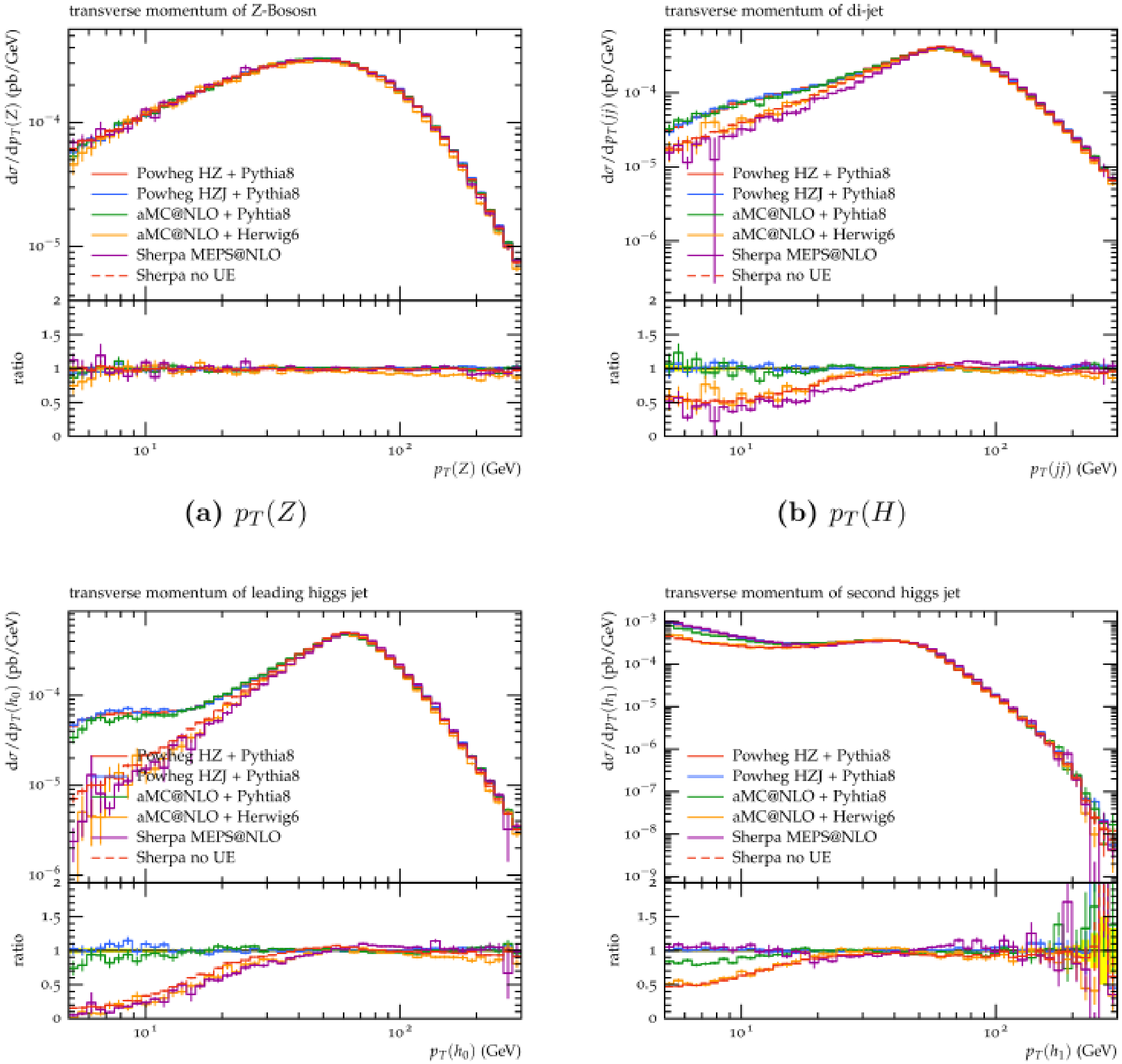
(b) d_{12}

Figure 10.3: FxFx merging validation plots for the k_T distances d_{01} and d_{12} of the aMC@NLO + Pythia 8 Z+jets sample for a merging scale of 20 GeV.

Comparisons using Rivet

- ZllH analysis designed to cover both fully inclusive and boosted (i.e. signal) regions
  - Require at least 2 jets originating from a b-quark and 2 oppositely charged leptons compatible with Z mass
  - Inclusive case: pseudorapidity of the objects up to |5|
  - Signal region: $|\eta|<2.5$ and pT > 20 GeV for leptons and jets, pT(Z) > 100 GeV

- Plots are normalized in the inclusive region to the theory cross section
  - They are NOT renormalized after the signal region cuts, allowing for changes due to acceptance differences



(a) $p_T(Z)$ (b) $p_T(H)$

(c) $p_T(h_0)$ (d) $p_T(h_1)$

Figure 11.1: p_T distributions of the Z and the H systems, as well as the two Higgs daughter jets h_0 & h_1 in the inclusive region and their ratio over Powheg HZ, errors include statistical uncertainty only.

NB: inclusive region

Differences arise in very low pT phase space (not surprising)

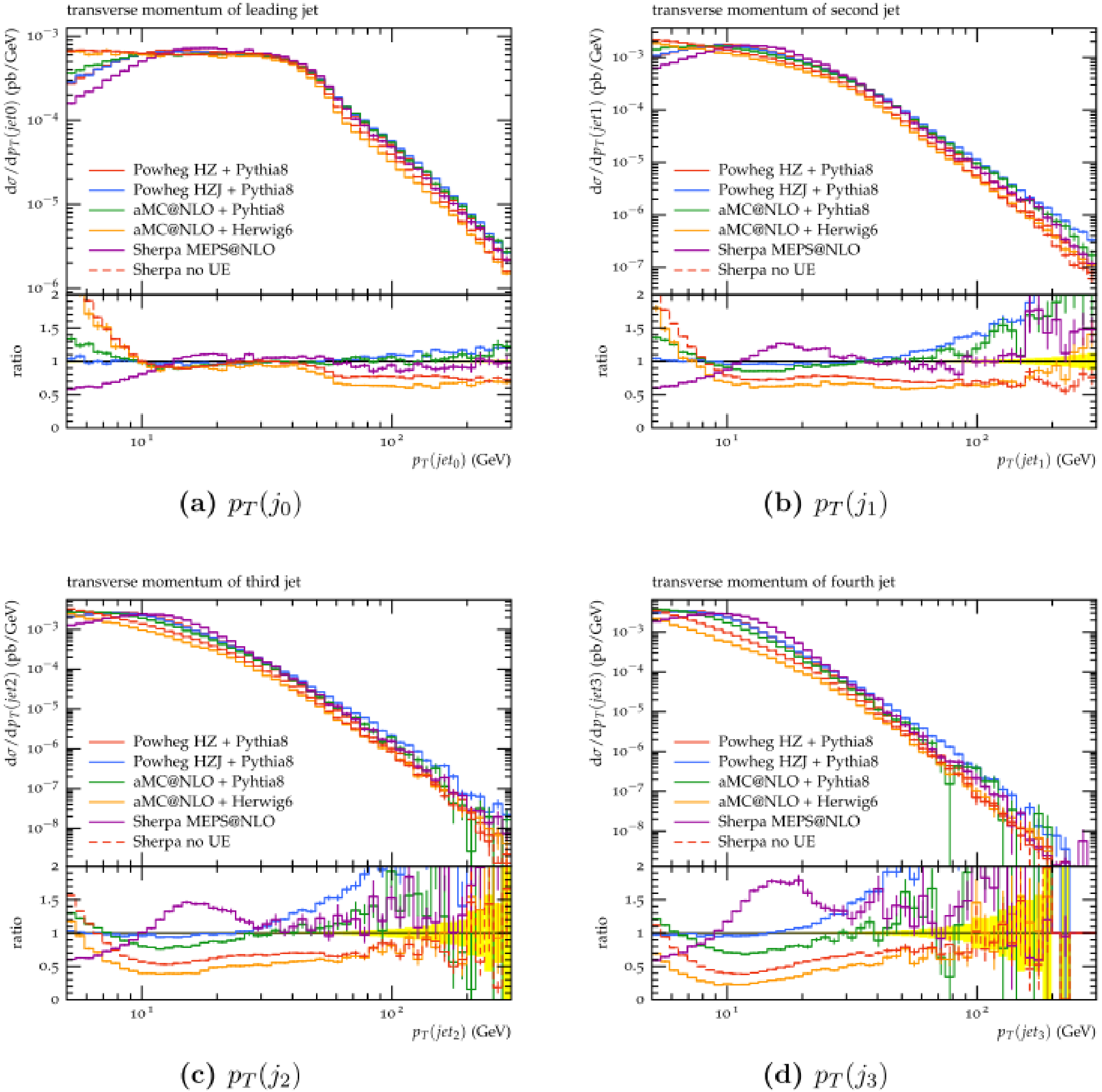
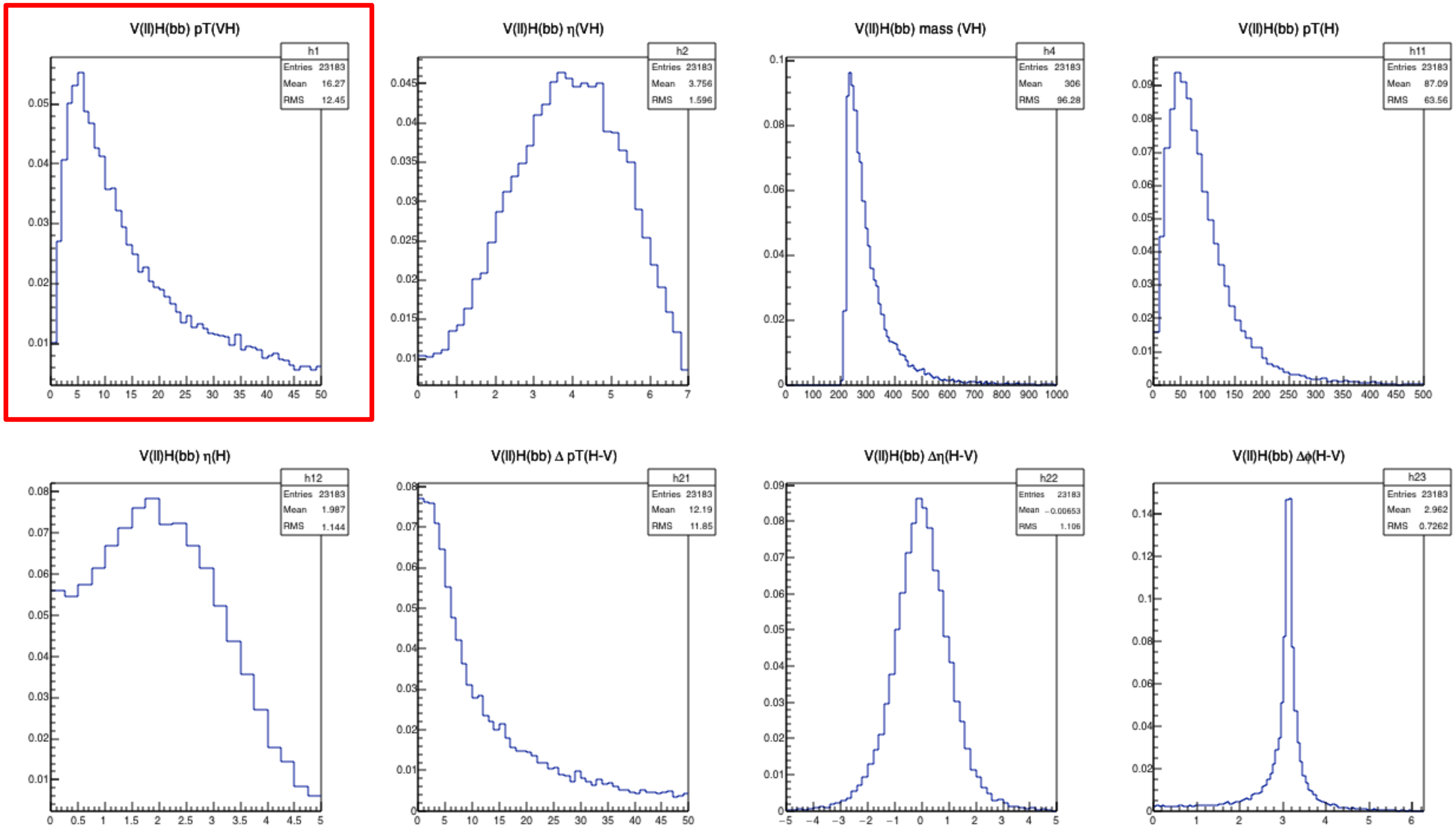


Figure 11.2: p_T distributions of the first 4 additional jets $j_0 - j_3$ and their ratio over Powheg HZ in the inclusive region. Errors include statistical uncertainty only.

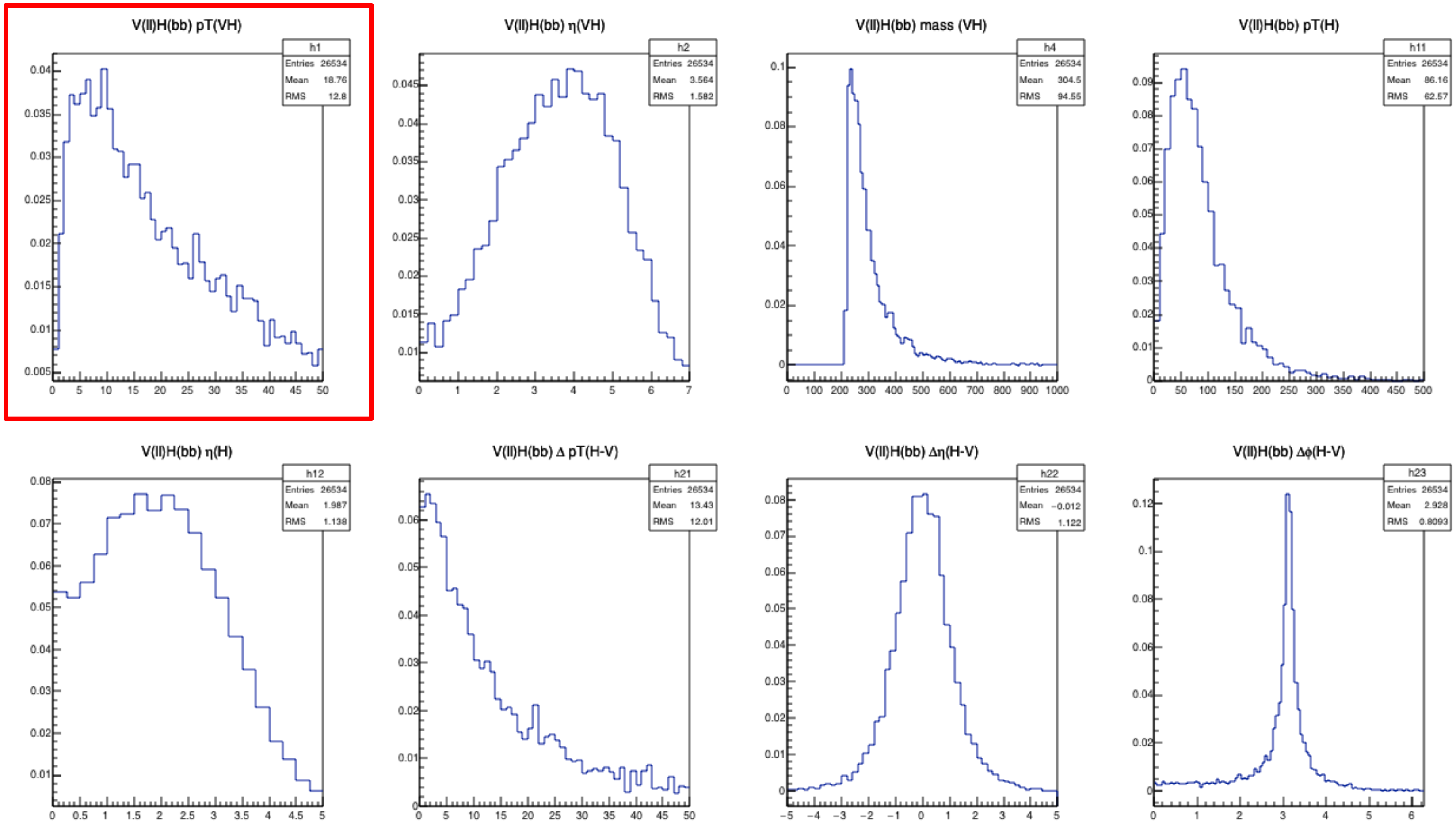
NB: inclusive region

Differences are larger for high numbers of additional jets (mainly due to different number of matched jets)

POWHEG + Herwig^{++}



aMC@NLO + Pythia8



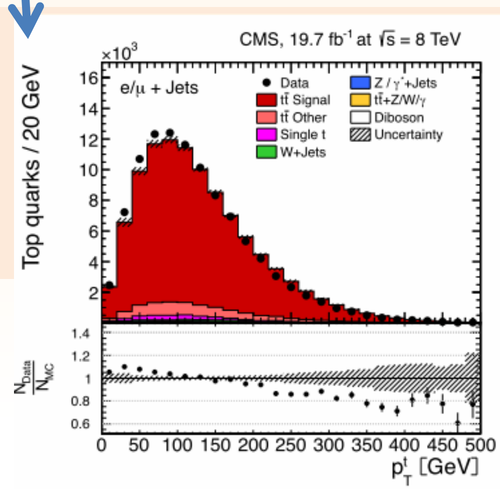
TTBAR+JETS COMPARISON

- Agreement has been made on the generator setup and observables to study; analysis routine is being written
  - On June 15 there will be a ttH/tH subgroup of LHCHXSWG meeting dedicated to MC validation/simulations within ATLAS & CMS
  - Probably after that, we can finalize the analysis and start making the predictions
- Extensive list of simulation programs and methods, MadGraph5_aMC@NLO, Sherpa, OpenLoops, MEPS@NLO, Herwig++, Pythia8, UNLOPS, FxFx merging, POWHEG, …
- CMS has agreed that we can use some of their event samples to perform this analysis. ATLAS hasn't agreed just yet…

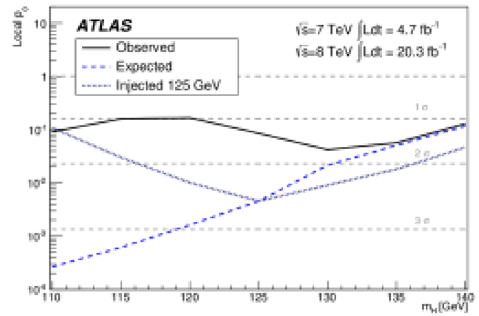
Observables:

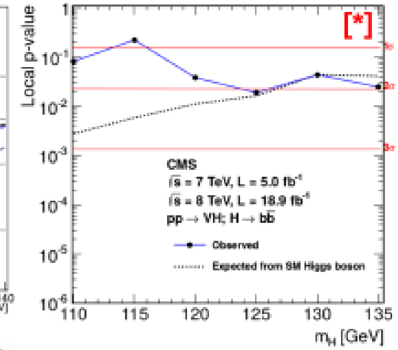
- pT(top)
- pT(ttbar)
- pT(ttbar+jet)
- m(t,tbar)
- m(top)
- DeltaPhi(l1,l2)
- DeltaPhi(lj1,lj2)
- m(b1,b2)
- # of jets
- # of b-jets
- # of light jets
- pT(j1), pT(j2), pT(j3), pT(j4)
- pT(b1), pT(b2), pT(b3), pT(b4)
- pT(lj1), pT(lj2), pT(lj3), pT(lj4)
- asymmetries:
  - lepton
  - ttbar
- Gap fraction: Q0, inclusive in Delta y
- tt+jets as background to ttH. Input from the

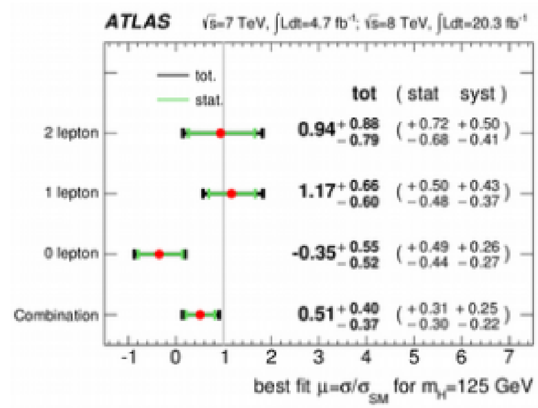
LesHouches2015, wiki

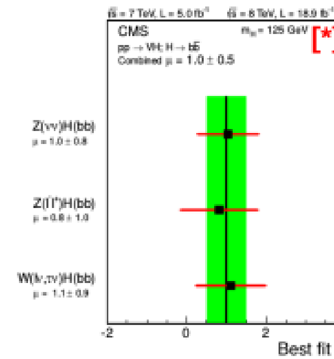


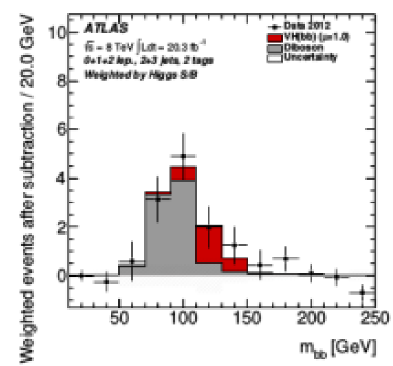
- Observed (expected) p-value:
  - ATLAS: **1.4σ (2.6σ)**;
  - CMS: **2.1σ (2.5σ)**.
- Signal strength:
  - ATLAS: $\boldsymbol{\mu = 0.52 \pm 0.40}$;
  - CMS: $\boldsymbol{\mu = 0.89 \pm 0.43}$.

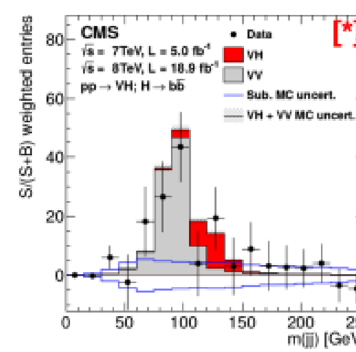












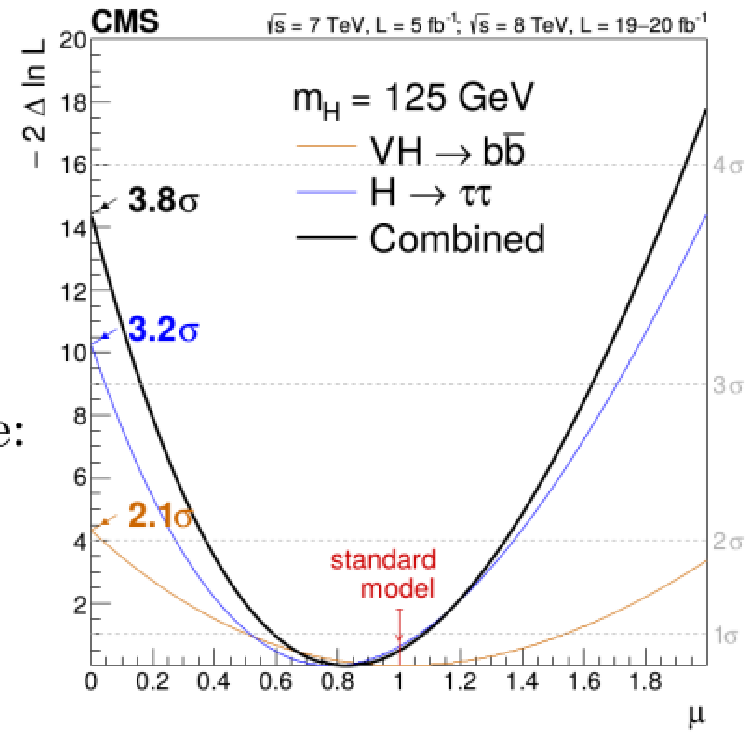
[*] : plots obtained excluding gg → ZH contribution.

- Combination of all H → bb analysis, signal strength:

[new!]

  - CMS (VH, VBF, ttH$^{(*)}$): $\boldsymbol{\mu = 1.03^{+0.44}_{-0.42}}$.
  - ATLAS (VH, ttH): $\boldsymbol{\mu = 0.63^{+0.39}_{-0.37}}$.

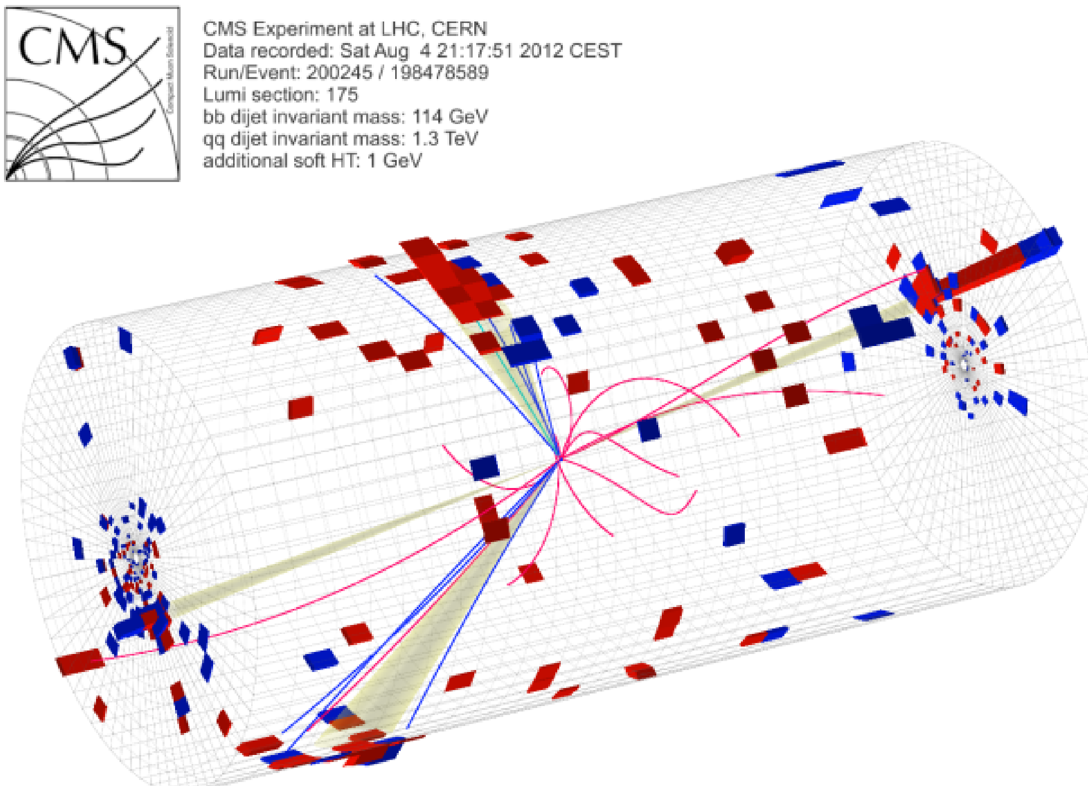
- Higgs to fermions (H→ττ, VH→bb) p-value:
  - Observed (exp.), CMS: **3.8σ (4.4σ)**.
  - Observed, ATLAS: **~4.5σ**.



[new!] VBF H → bb

☺ Highest branching ratio ~ 60 % and large cross section;

☹ Fully hadronic final state → large QCD background.

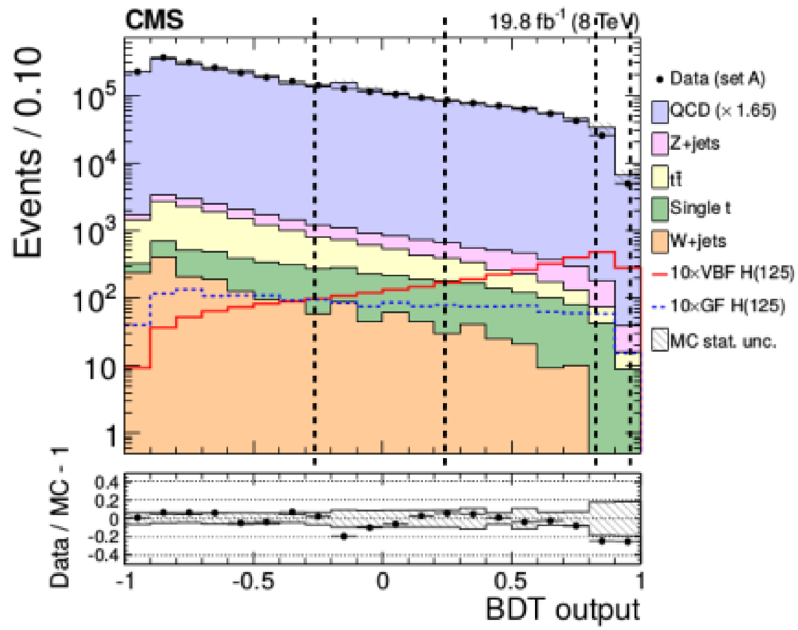


☺ Peculiar final states:

- Two b-jets;
- Two quark-jets with large $\Delta\eta$;
- No additional hadronic activity between them.

[new!] VBF $H \to bb$

- Events are divided in 7 categories, with different S/B, using a multivariate discriminator (uncorrelated with m_{bb}).
- Signal is extracted with a simultaneous fit on m_{bb} in all categories.
- QCD is fitted in all categories with a common fifth order polynomial.
- QCD shape corrected with a category-dependent quadratic transfer function.



Signal strength

Normalization and shapes of signal, Z and top

Free parameter

Fixed parameter (simulation)

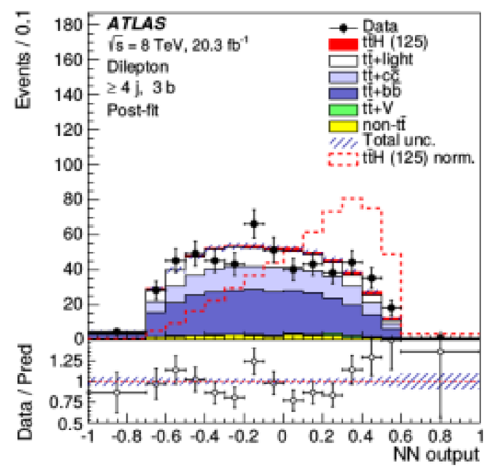
$$f_i(m_{bb}) = \mu_H \cdot N_{i,H} \cdot H_i(m_{bb}; k_{JES}, k_{JER}) + N_{i,Z} \cdot Z_i(m_{bb}; k_{JES}, k_{JER}) + N_{i,t} \cdot T_i(m_{bb}; k_{JES}, k_{JER}) + N_{i,QCD} \cdot K_i(m_{bb}) \cdot B(m_{bb}; \vec{p}_{set}),$$

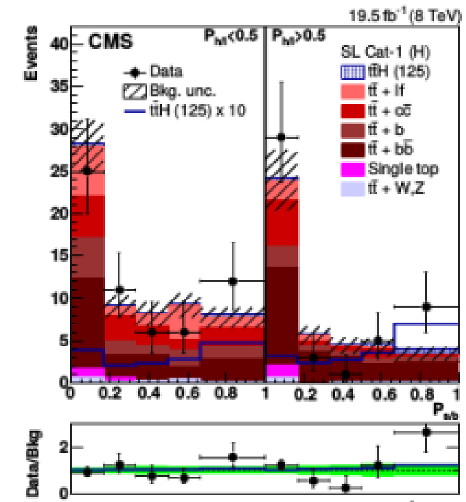
QCD normalization

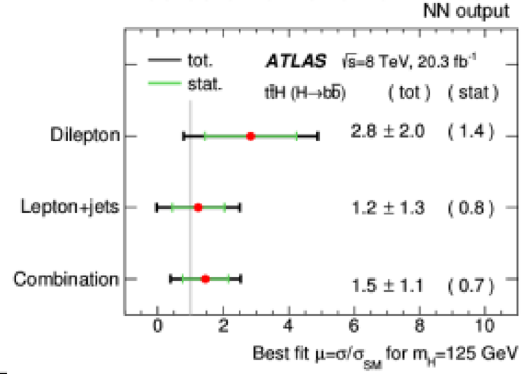
Transfer function (linear or quadratic)

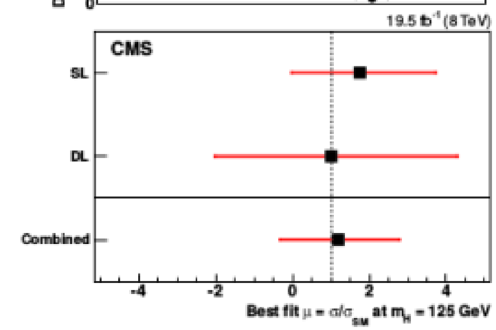
Polynomial QCD shape with free parameters.

- Signal extraction:
  - Fit on multivariate discriminant that includes MEM [ATLAS];
  - 2D fit using MEM and a heavy/light jets discriminant [CMS].
- Observed (expected) 95% CL upper limit:
  - ATLAS: **3.4 (2.2)**;
  - CMS: **4.2 (3.3)**.
- Signal strength:
  - ATLAS: $\boldsymbol{\mu = 1.5\,^{+1.1}_{-1.1}}$;
  - CMS: $\boldsymbol{\mu = 1.2\,^{+1.6}_{-1.5}}$.



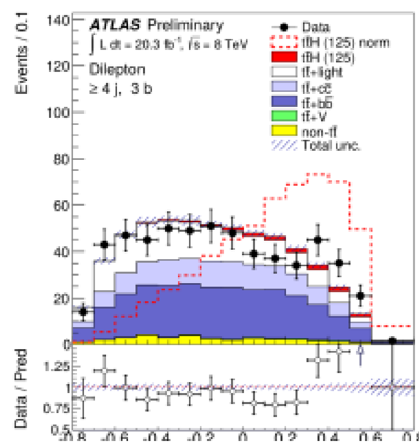


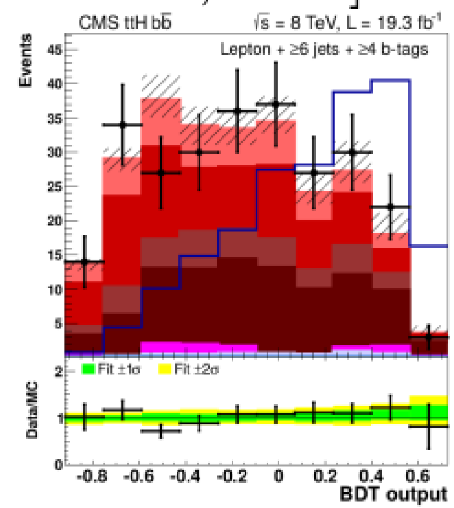


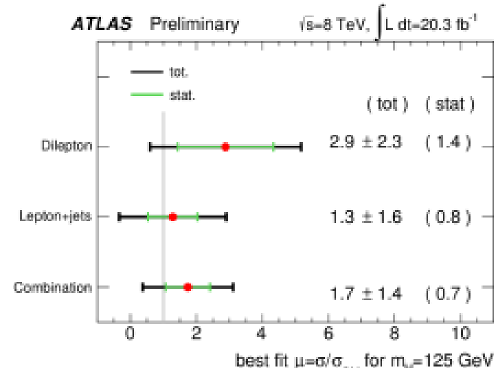


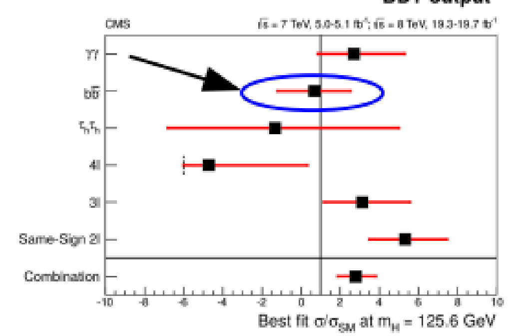
ttH → bb (legacy)

- Previous analysis without using Matrix Element method.
- Signal extraction:
  - Fit on multivariate discriminant distribution [ATLAS,CMS].
- Observed (expected) 95% CL upper limit:
  - ATLAS: **4.1 (2.6)**;
  - CMS: **4.1 (3.5)**.
- Signal strength:
  - ATLAS: $\boldsymbol{\mu = 1.7^{+1.4}_{-1.4}}$;
  - CMS: $\boldsymbol{\mu = 0.7^{+1.9}_{-1.9}}$.



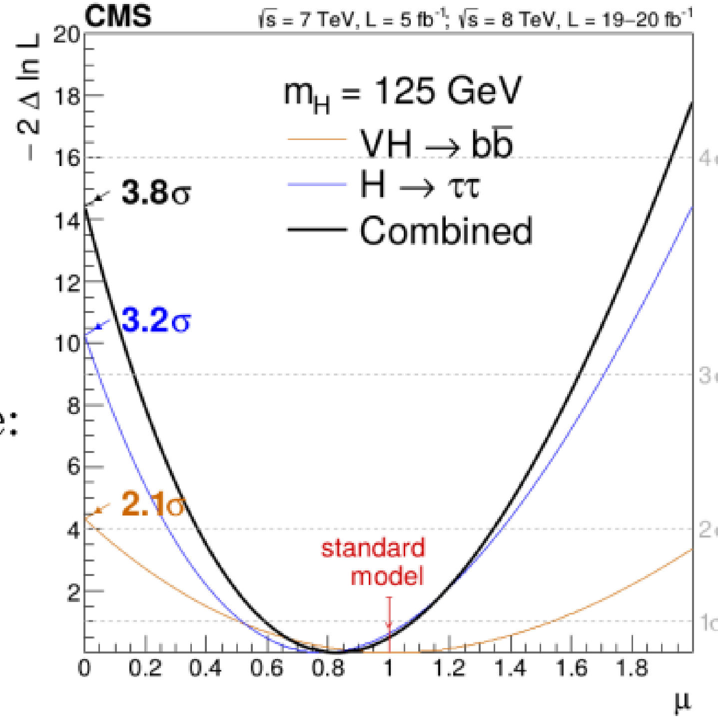






$H \rightarrow$ fermions combination

- Combination of all $H \rightarrow bb$ analysis, signal strength:
  - [new!] CMS (VH, VBF, ttH$^{(*)}$): $\mu = 1.03^{+0.44}_{-0.42}$.
  - ATLAS (VH, ttH): $\mu = 0.63^{+0.39}_{-0.37}$.
- Higgs to fermions ($H\rightarrow\tau\tau$, $VH\rightarrow bb$) p-value:
  - Observed (exp.), CMS: $\mathbf{3.8\sigma}$ $\mathbf{(4.4\sigma)}$.
  - Observed, ATLAS: $\sim\mathbf{4.5\sigma}$.



(*) Legacy analysis (no Matrix Element)