LHC Higgs Cross Section Meeting (ttH WG), May 4, 2012

Search for ttH in ATLAS: Questions for Theorists

Aurelio Juste
ICREA/IFAE, Barcelona

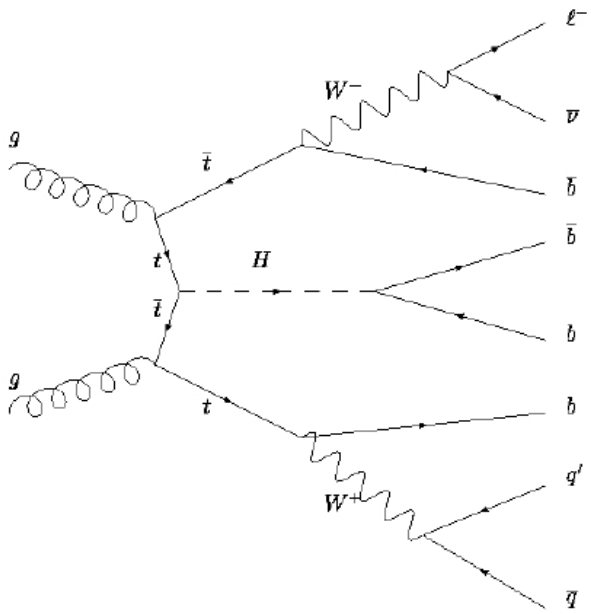
On behalf of the ATLAS HSG5 Working Group

The Challenge (I)

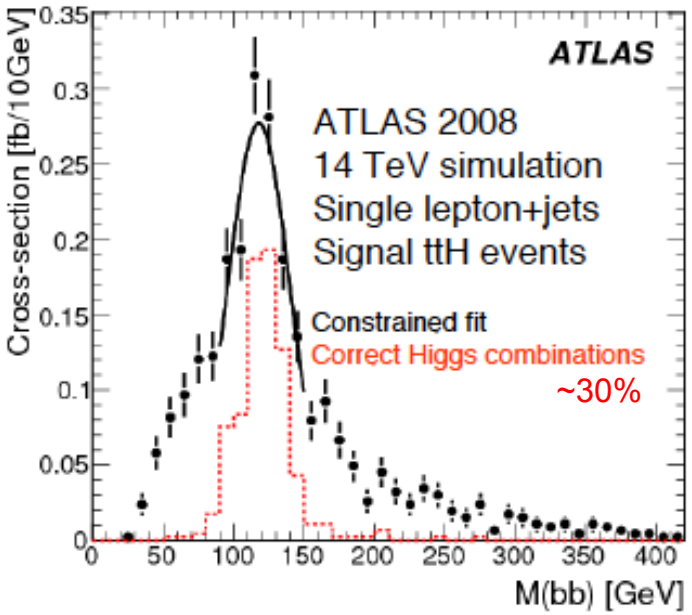
- Small signal cross section on top of huge tt+jets background. At $\sqrt{s}$=7 TeV:
  $\sigma(ttH) \times BR(H \rightarrow bb) \sim 65$ fb @ m_H=120 GeV
  $\sigma(tt) \sim 160$ pb

- Currently focusing on lepton+jets channel:
  - =1 e or μ, $p_T(e)$>25 GeV, $p_T(\mu)$>20 GeV
  - E_T^{miss}>30 GeV (e+jets), >20 GeV (μ+jets)
  - ≥6 jets p_T>25 GeV, $|\eta|$<2.5
  - =3 or ≥4 jets are tagged (~70% b-jet eff., ~0.7% mistag rate)
  - ➔ Very sensitive to b-tagging systematics

- Signal reconstruction via constrained kinematic fit:
  - Hadronic W resonance: $m_{jj} \sim m_W$
  - Leptonic W resonance: $m_{l\nu} \sim m_W$
  - Top quark resonances: $m_{jjb} \sim m_{l\nu b} \sim m_t$
  - ➔ Large combinatorial background (~20% correct H→bb combinations)



ATLAS CSC book (CERN-OPEN-2008-020)

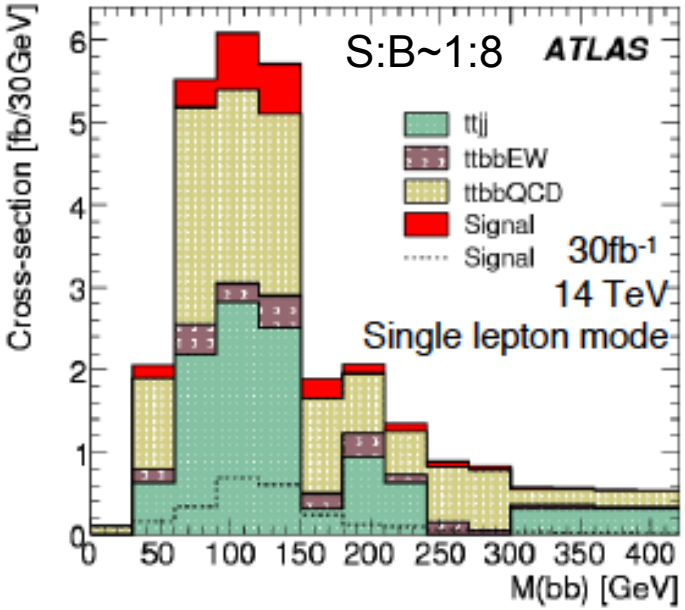


The Challenge (II)

- Large tt+jets (incl tt+HF) background that must be precisely estimated from a combination of
  - ME+PS MC
  - NLO calculations
  - Data-driven techniques

  We not only care about normalization but also shape! Signal is where background peaks…

- The question is: how to do this consistently?
  - ➔ This is less trivial than it seems.
  - ➔ W+HF is an example where after more than a decade of data analysis and increasingly more sophisticated MCs and high-order calculations, still unsatisfactory procedures are being followed to both normalize W+HF and assess modeling uncertainties.

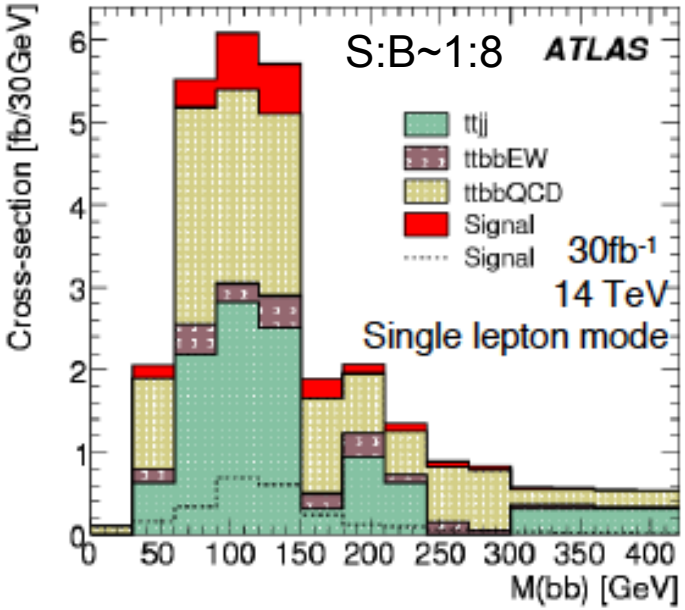


The Challenge (II)

- Large tt+jets (incl tt+HF) background that must be precisely estimated from a combination of
  - ME+PS MC
  - NLO calculations
  - Data-driven techniques

  We not only care about normalization but also shape! Signal is where background peaks…

- The question is: how to do this consistently?
  - ➔ This is less trivial than it seems.
  - ➔ W+HF is an example where after more than a decade of data analysis and increasingly more sophisticated MCs and high-order calculations, still unsatisfactory procedures are being followed to both normalize W+HF and assess modeling uncertainties.



This goal of this talk is not to provide details on how advanced the ATLAS analyses are (still internal information), but rather to start a conversation on how to address the background modeling issue.

Available Tools

- To my knowledge, what's currently available (may not be completely accurate):

ttH:

fixed-order calculation up to NLO

MCs: POWHEG, aMC@NLO for NLO ttH production

ttbb:

fixed-order calculation up to NLO for ttbb

MCs:

- AcerMC: LO, QCD+EW contributions
- ALPGEN: LO, QCD ME+PS (PYTHIA or HERWIG)
- SHERPA: LO, QCD ME+PS

tt+jets:

fixed-order calculation up to NLO for ttjj

MCs:

- ALPGEN: LO, QCD ME+PS (up to high parton muliplicity)
  ➔ needs overlap between PS and ME for ttbb, ttcc
- SHERPA: LO, QCD ME+PS (up to high parton muliplicity)
- MC@NLO: tt incl @ NLO, tt+1j @ LO, extra-jets LL via PS (HERWIG)
- POWHEG: tt+1j @ NLO (PS from PYTHIA or HERWIG)

Questions on tt+jets Modeling

Basic requirements:

- Need to be able to describe tt+jets over a wide range in jet multiplicity spectrum.
- Need as a minimum a LO calculation for ttbb.

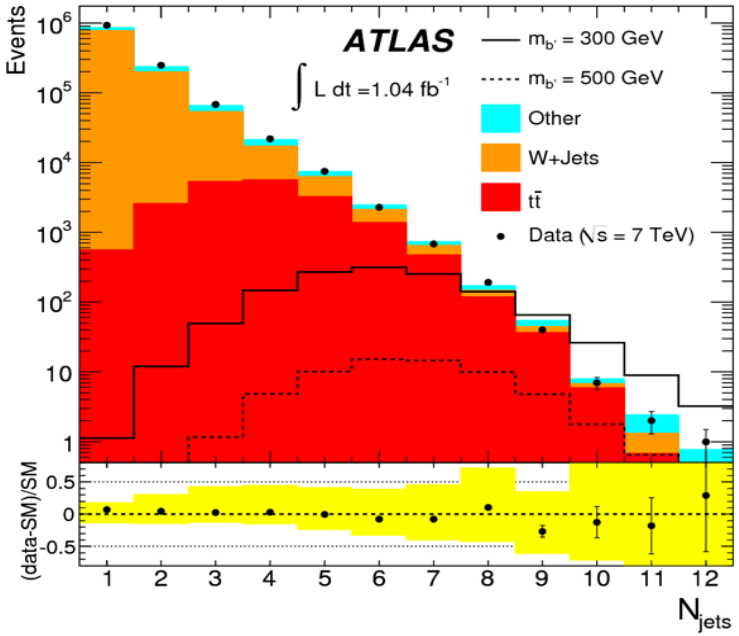
➔ Currently using ALPGEN+HERWIG.

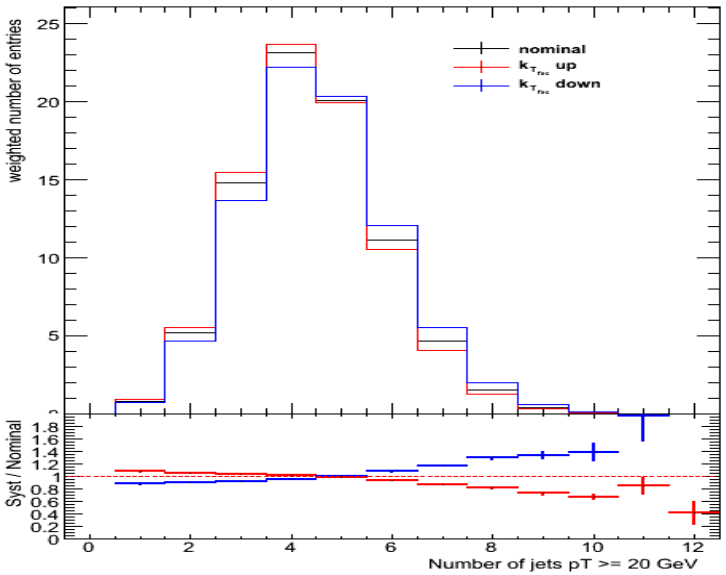
Q1: What variations in generator parameters should be considered to cover for possible modeling uncertainties?

Can such variations (e.g. functional form of factorization scale) be considered correlated between ttjj and ttbb?

What would be ttbb-specific systematics?

arXiv:1202.6540





Questions on tt+jets Modeling

Basic requirements:

- Need to be able to describe tt+jets over a wide range in jet multiplicity spectrum.
- Need as a minimum a LO calculation for ttbb.

➔ Currently using ALPGEN+HERWIG.

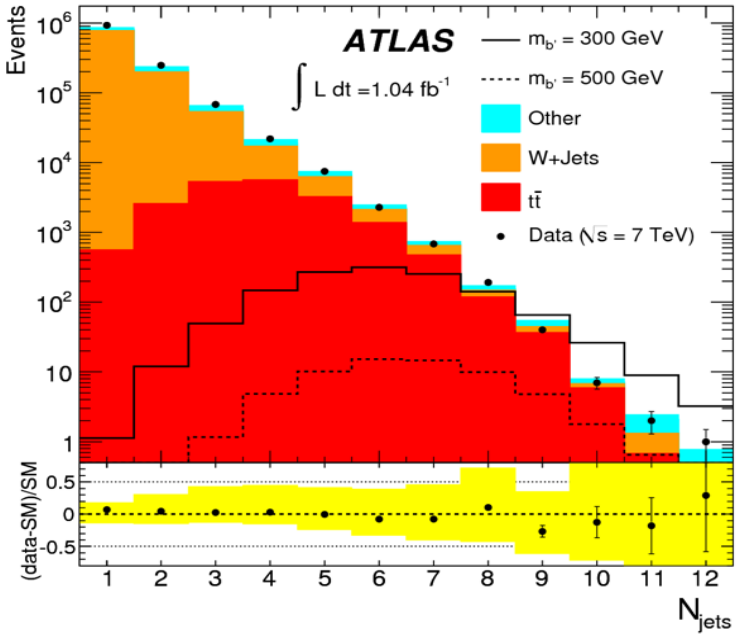
Q2: Can one use existing NLO calculations to "tune" ALPGEN generation parameters to better describe shapes and/or constrain range of variation in parameters to explore?

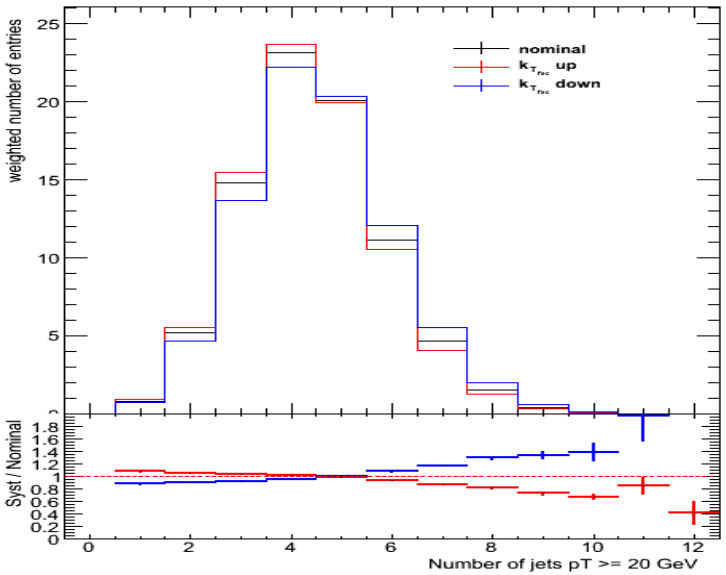
Q3: How to use existing NLO calculations to "normalize" ALPGEN at particular jet multiplicity bins? What are the related uncertainties?

Q4: Are NLO calculations available at 7 and 8 TeV?

Are there "user-friendly" tools that can be run by experimentalists for ME+PS vs NLO comparisons of differential distributions, etc?

arXiv:1202.6540



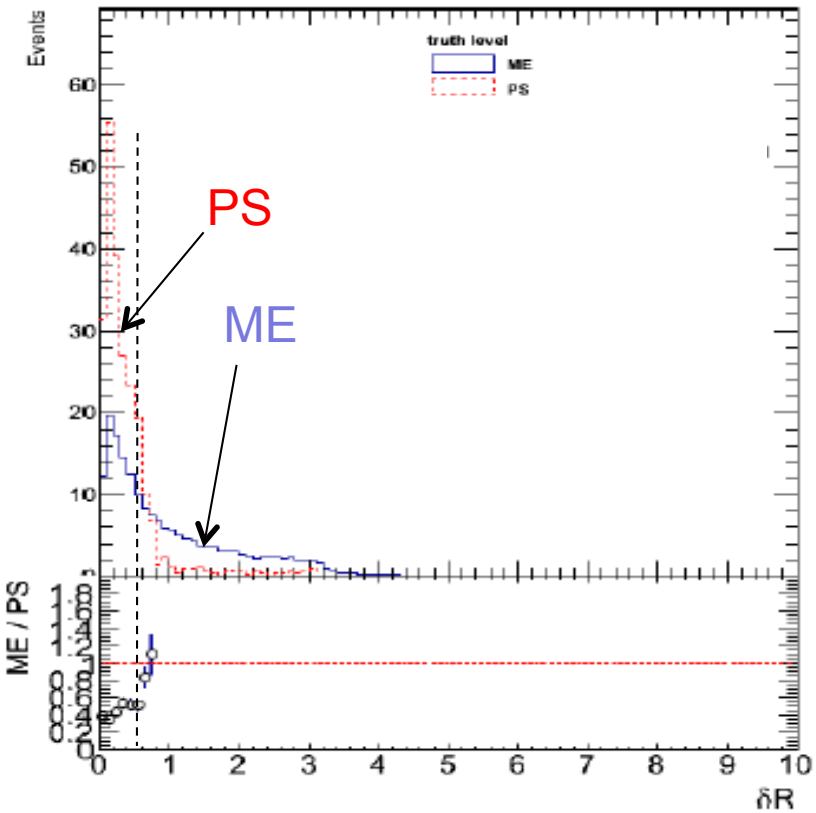


Questions on tt+jets Modeling

ALPGEN does not implement a procedure for overlap removal between ME and PS for ttQQ. This is currently done “by hand” with a relatively ad-hoc prescription:

- Generate tt+light partons ME sample.
  (QQ pairs will be generated in PS)
- Generate ttQQ (Q=b,c) ME sample
- Use ttQQ ME sample if $\Delta R(Q,Q)>0.4$
  Otherwise use QQ from PS (from tt+light partons sample).

Rationale is that low angle/soft QQ pairs will be more accurately described by PS (this is ~50% of ttbb events with ≥6 jets/≥4 tags!)



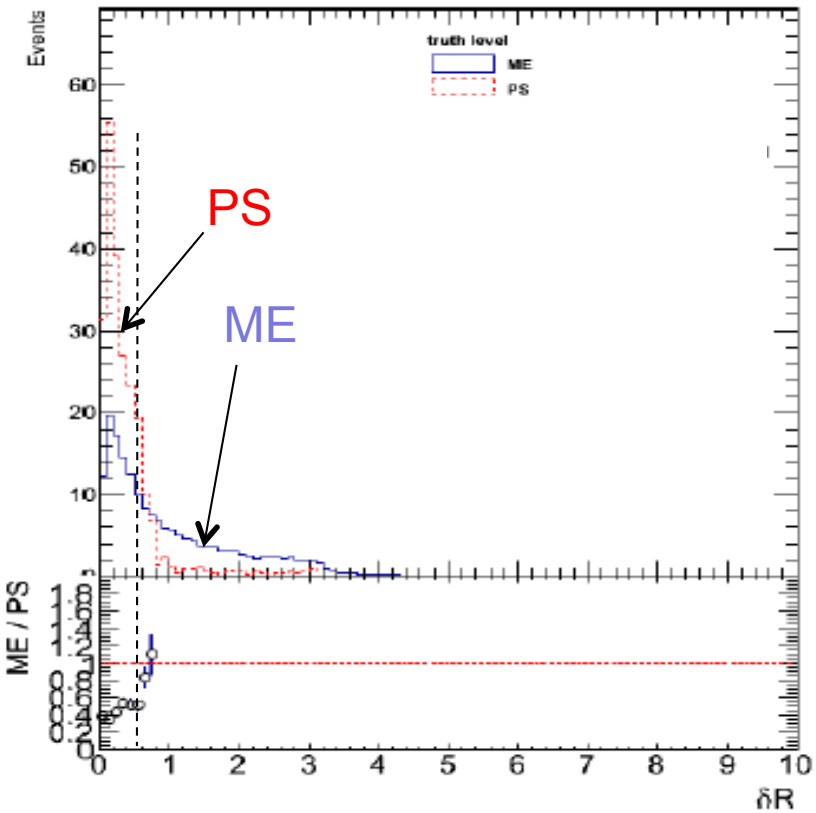
Q5: Does this make sense? How do we assess systematic uncertainties on the relative fraction of tt(QQ) and ttQQ events? [This is needed because these events have different topology and contribute differently to =3-tag and ≥4-tag bins.]

Questions on tt+jets Modeling

ALPGEN does not implement a procedure for overlap removal between ME and PS for ttQQ. This is currently done "by hand" with a relatively ad-hoc prescription:

- Generate tt+light partons ME sample.
  ($Q\bar{Q}$ pairs will be generated in PS)
- Generate ttQQ (Q=b,c) ME sample
- Use ttQQ ME sample if $\Delta R(Q,Q)>0.4$
  Otherwise use QQ from PS (from tt+light partons sample).

Rationale is that low angle/soft QQ pairs will be more accurately described by PS (this is ~50% of ttbb events with ≥6 jets/≥4 tags!)



Q6: In order to have a more accurate background prediction it would be beneficial to normalize the ratio ttbb/ttjj to the NLO calculation. Does such ratio and related uncertainty exist at 7 and 8 TeV?

How does one use it given the above prescription?

Is the NLO calculation trustworthy for $\Delta R(Q,Q)<0.4$ or do resummation effects become important?