Top modelling and tuning in ATLAS

status, latest developments and wishlist

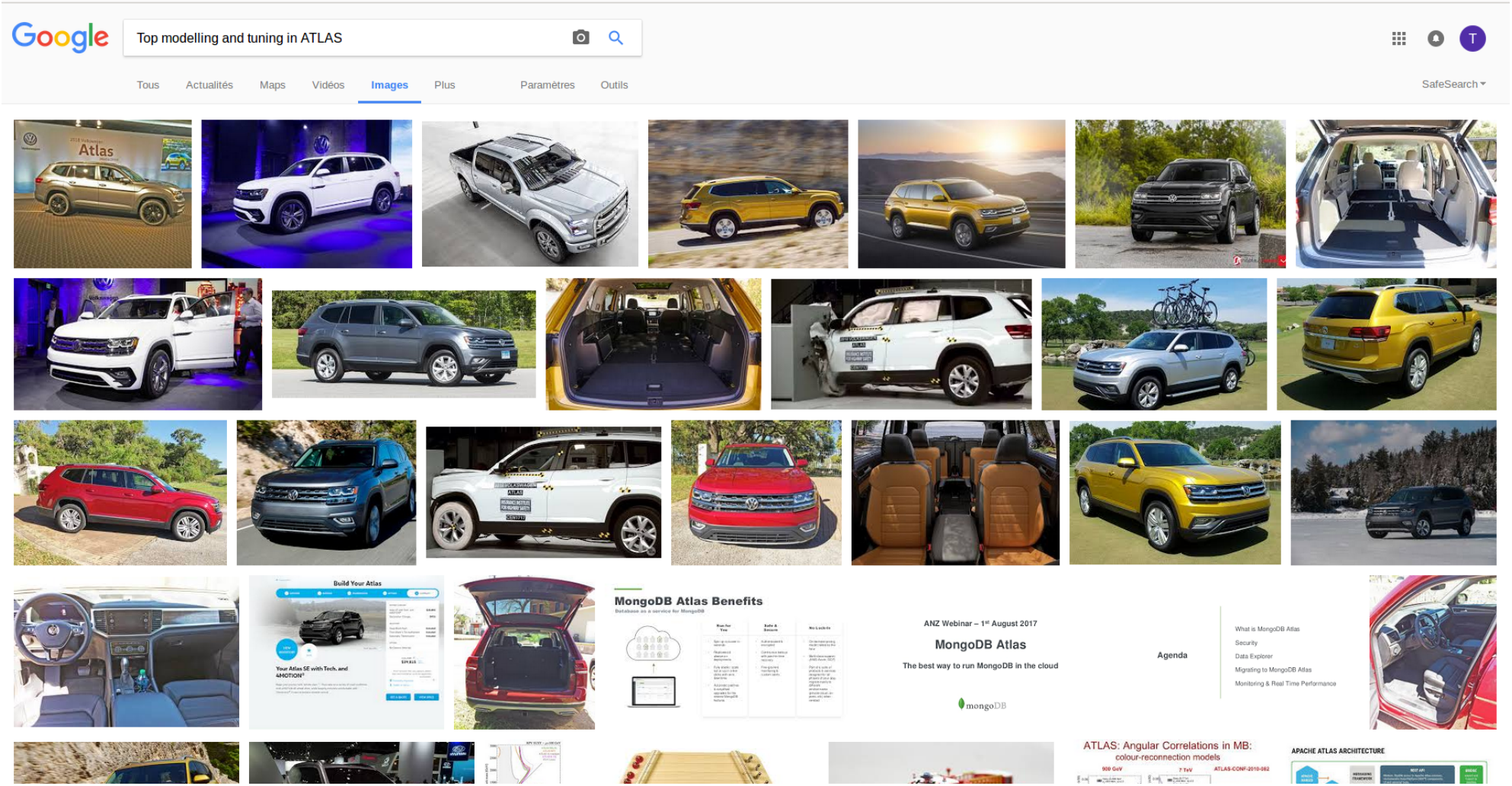
Timothée Theveneaux-Pelzer

on behalf of the ATLAS collaboration

DESY - Zeuthen
(previously at CPPM-Marseille)

Tuesday, September 19th 2017

“Top modelling and tuning in ATLAS” : let’s google it



Introduction

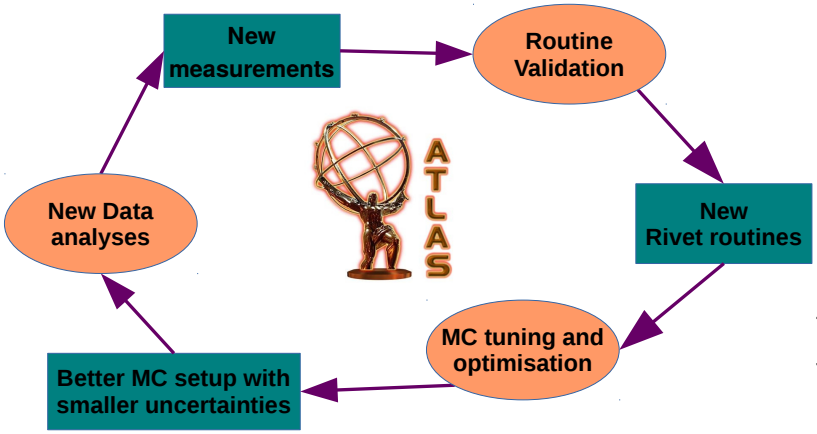
- We need Monte-Carlo generators to model top-quark processes
  - → top-quark as signal - modelling of detector acceptance and response
  - → top-quark as background - reliable prediction of any variable in the full phase-space
- MC generators setups are made of several components
  - → need to choose parameters which cannot be obtained from first principles
  - → each configuration requires to be wisely optimised (or tuned)
- Many studies peformed by ATLAS since run-1
  - → 13 PUB-notes in TopPublicResults page + a few more in MCPublicResults page
  - → comparisons between MC generators setups
  - → with more and more measurements available : comparisons of these setups to data
- In this talk : status and recent developments in ATLAS
  - → mostly speaking about $t\bar{t}$

Overview

1. With several ingredients...
   - Available ATLAS unfolded data
   - MC generators setups and tunes
2. ...we define a good nominal setup...
   - Baseline ttbar model : Powheg+Pythia8
3. ...with meaningful and reliable modelling uncertainties...
   - Additional radiation
   - Hadronisation and Parton Shower model
   - NLO matching scheme
4. ...while working on further developments
   - NLO multileg generators
   - Other processes
   - Future improvements and wishlist
5. Conclusions

Available ATLAS unfolded data

- Feeding community with results useable for tuning
  - → well-defined observables
  - → unfolded at particle-level
  - → Rivet routines (and HepData)



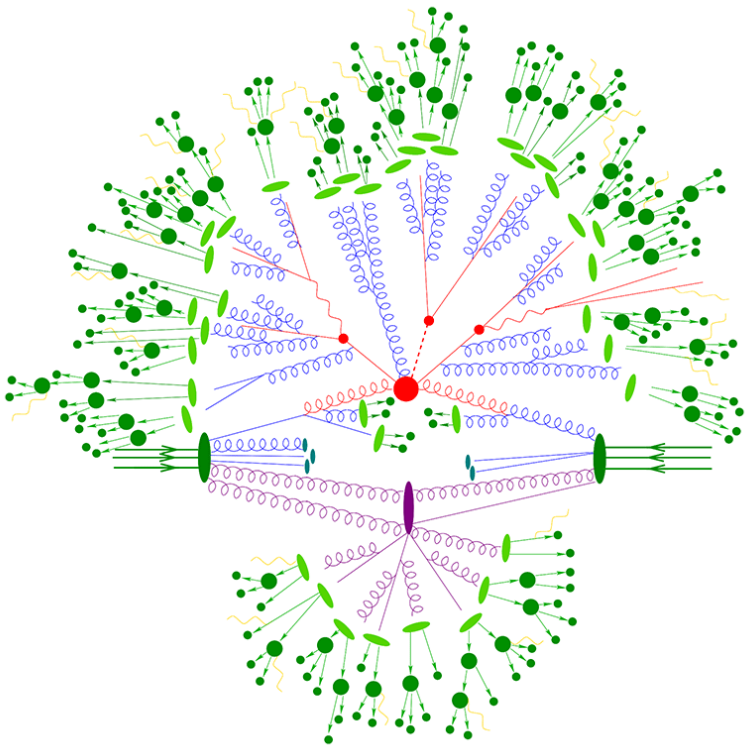
- Many available measurements (mostly $t\bar{t}$)
- Several new results this week !
- New data to improve our understanding
  - → new channels
  - → new processes
  - → new variables
- Not shown on this list :
  - → results unfolded at parton-level
  - → ongoing analyses

$\sqrt{s}$	channel	description	reference
		Public Rivet routines	
7 TeV	2ℓ	$t\bar{t}$ gap fractions	TOPQ-2011-21
7 TeV	1ℓ, 2ℓ	$t\bar{t}$ jet shapes	STDM-2011-48
7 TeV	1ℓ	$t\bar{t}$ Njets, jets p_T	TOPQ-2012-03
7 TeV	1ℓ	$t\bar{t}$ diff. (resolved)	TOPQ-2013-07
8 TeV	1ℓ	$t\bar{t}$ jet pull observables	TOPQ-2014-09
8 TeV	1ℓ	$t\bar{t}$ diff. (boosted)	TOPQ-2014-15
8 TeV	1ℓ, 2ℓ	$t\bar{t}$+HF inclusive	TOPQ-2014-10
8 TeV	2ℓ	Wt+$t\bar{t}$ inclusive	TOPQ-2012-20
8 TeV	1ℓ	$t\bar{t}$ diff. (resolved)	TOPQ-2015-06
		Other publications	
8 TeV	1ℓ	t-channel diff.	TOPQ-2015-05
8 TeV	1ℓ	$t\bar{t}$ spin density matrix	TOPQ-2015-13
8 TeV	1ℓ	$t\bar{t}+\gamma$ diff.	TOPQ-2015-21
8 TeV	2ℓ	$t\bar{t}$ Njets, jets p_T, gap fractions	TOPQ-2015-04
13 TeV	2ℓ	$t\bar{t}$ Njets, jets p_T, gap fractions	TOPQ-2015-17
13 TeV	2ℓ	$t\bar{t}$ diff. (resolved)	TOPQ-2016-04
13 TeV	1ℓ	$t\bar{t}$ diff. (resolved, boosted)	TOPQ-2016-01
		New results shown this week !	
13 TeV	0ℓ	$t\bar{t}$ diff. (boosted)	TOPQ-2016-09
13 TeV	2ℓ	Wt diff.	TOPQ-2016-12
...	...	...	TOPQ-2017-13



Modelling $pp \to t\bar{t} + X$ events

- Several ingredients to build a good model, for each step of the $pp \to t\bar{t} + X$ event
  - $\to$ Parton Distribution Functions
  - $\to$ Matrix Element generator
  - $\to$ Parton Shower and hadronisation generator
  - $\to$ possibly : additional legs in the ME generator
  - $\to$ possibly : afterburners for decay of some particles

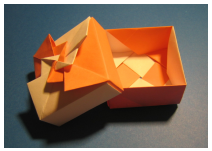


- For each ingredient :
  choice determined by different kind of arguments
  - $\to$ trustable theoretical predictions
  - $\to$ existing measurement of a parameter
  - $\to$ comparisons of predictions to measurements
    - $\Rightarrow$ tuning (or optimisation)

- At each step, uncertainties are assessed
  - $\to$ propagating the uncertainty of the external input (e.g. PDFs)
  - $\to$ comparing options which give reasonable modelling of relevant variables

Available MC generators setups and tunes

- Latest ATLAS setup for $t\bar{t}$: Powheg+Pythia8
  - → switched from Powheg+Pythia6 used at the beginning of run-2
  - → Pythia8 tuned on 7 TeV ATLAS data : A14 tune
  - → PDF : NNPDF3.0NLO in the ME
  - → all PS generators use EvtGen (except Sherpa)
- Other generators used :
  - → MG5_aMC@NLO
  - → Herwig7, H7-UE-MMHT tune
  - → Pythia6, Perugia 2012 tune (for single-top)
  - → Herwig++, UE-EE-5 tune (for single-top)
  - → Sherpa, author tune
- Other PDF sets used :
  - → NNPDF2.3LO (for Pythia8 A14 tune)
  - → CT10, CT10f4 (for single-top)
  - → PDF4LHC15 (for PDF uncertainty estimation)

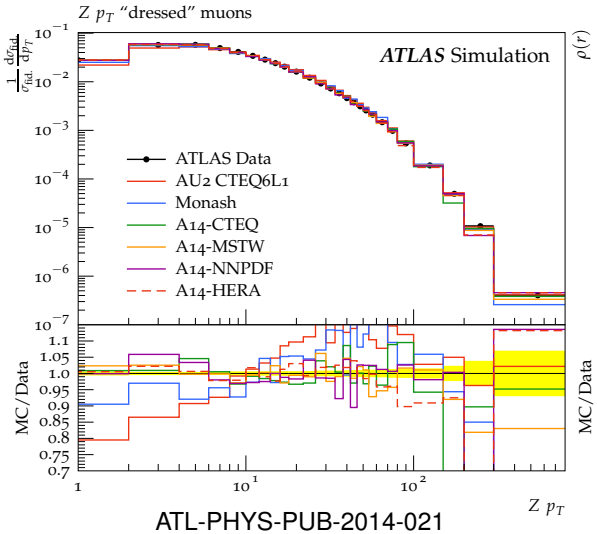




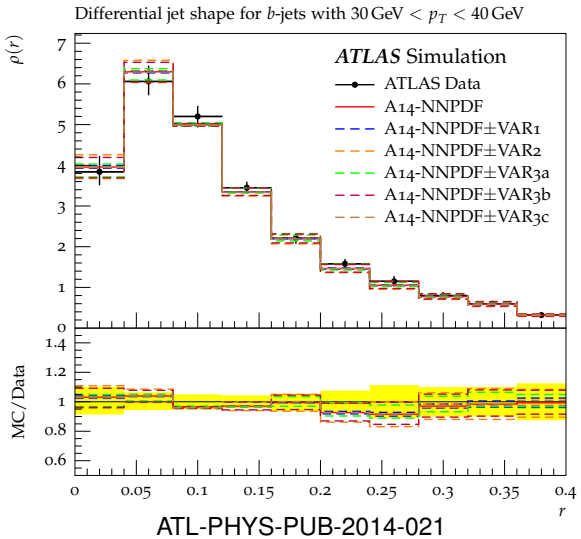

Pythia8 tunes to ATLAS 7 TeV data

ATL-PHYS-PUB-2014-021 - ATL-PHYS-PUB-2015-007 - ATL-PHYS-PUB-2015-048

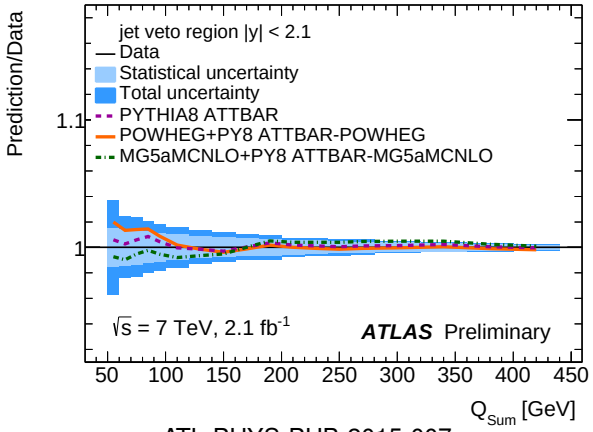
- ATLAS baseline Pythia8 tune : A14
  - → tuning performed with Professor
  - → provided in Pythia 8.2
- Used measurements :
  - → inclusive jet prod., multijets, $p_T(Z)$
  - → $t\bar{t}$ gap fractions and jet shapes
- Tuned Pythia8 parameters :
  - → hard process α_S, MPI (α_S, p_T cutoff)
  - → ISR (α_S, p_T cutoff, etc.), FSR (α_S)
  - → CR strength, beam remnants
  - → different PDF sets - best is NNPDF2.3LO
- Several variations for uncertainties
  - → uncertainties on MPI+CR, ISR/FSR
  - → 5 up/down variations ("eigentunes")
- Two other tunes with $t\bar{t}$ data only
  - → same $t\bar{t}$ data as A14 + N_{jets}, jet p_T
  - → ATTBAR
  - → A15-MG5aMCNLO-TTBAR
  - → some $t\bar{t}$ distributions better modelled
  - ⇒ gain not significant enough to replace A14 yet



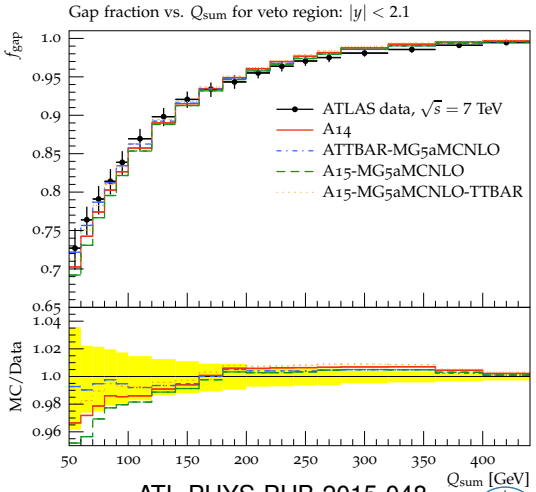
ATL-PHYS-PUB-2014-021



ATL-PHYS-PUB-2014-021



ATL-PHYS-PUB-2015-007

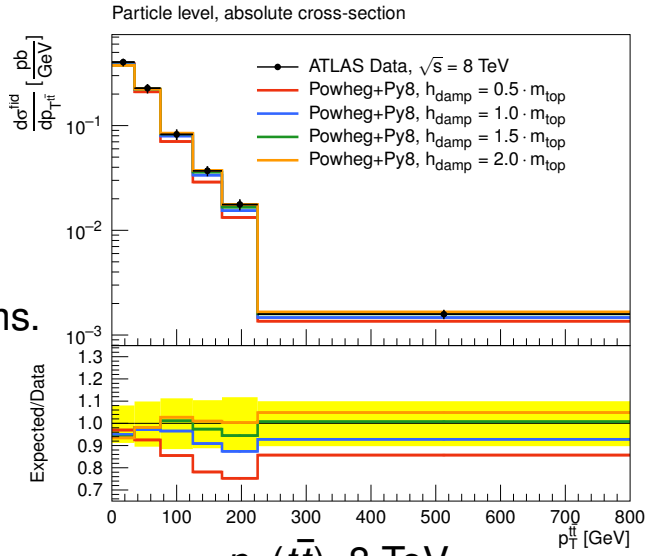


ATL-PHYS-PUB-2015-048

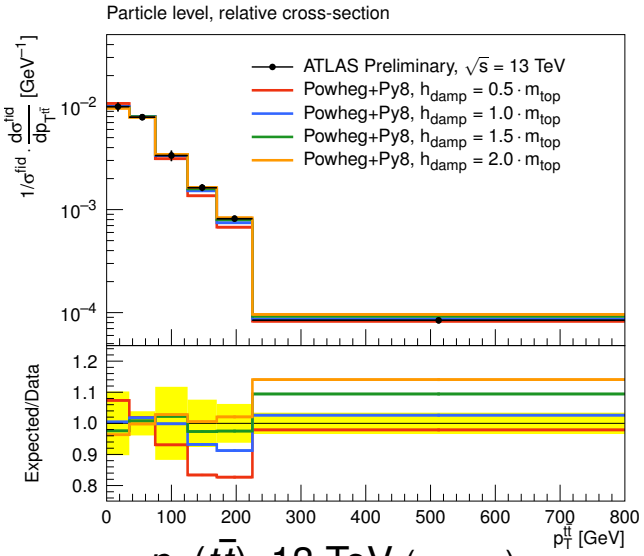
Powheg+Pythia8 : new baseline model

ATL-PHYS-PUB-2016-020

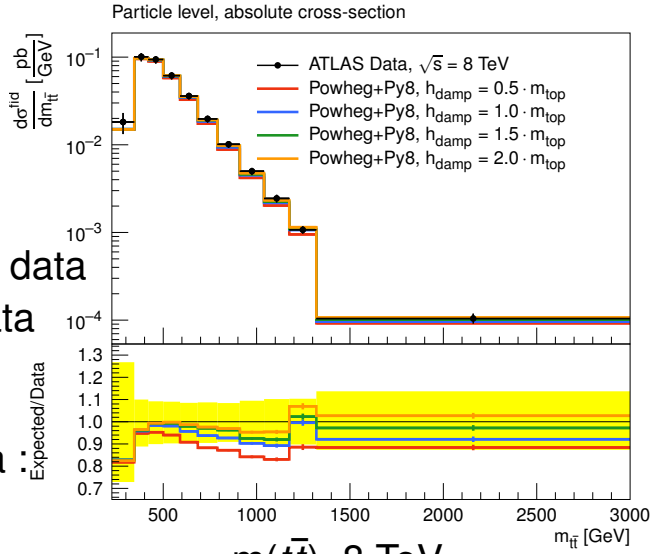
- Used Powheg+Pythia6 in run1
  - → $h_{\mathrm{damp}}=\infty$, then m_t
  - → kept for early run-2 analyses
- Switch to Powheg+Pythia8 :
  - → using Pythia8 A14 tune
  - → updated NLO matching params.
  - → re-optimisation of h_{damp}
- Effect of h_{damp} value
  - → ...as shown last year
  - → 1.5 m_t gives better $p_{\mathrm{T}}(t\bar{t})$
  - → very small effect on m($t\bar{t}$)
  - → works for 8 and 13 TeV data
- P6->P8 switch is crucial
  - → better model of soft regime
  - → Perugia tune : almost no LHC data
  - → A14 tune : only 7 TeV LHC data
- No Professor tuning of hdamp
  - → exercise done with 7 TeV data : ATTBAR-POWHEG tune



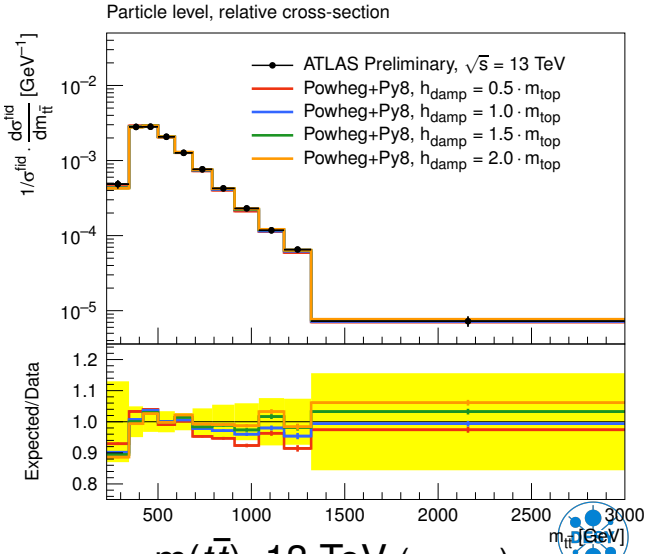
$p_{\mathrm{T}}(t\bar{t})$, 8 TeV



$p_{\mathrm{T}}(t\bar{t})$, 13 TeV (norm.)



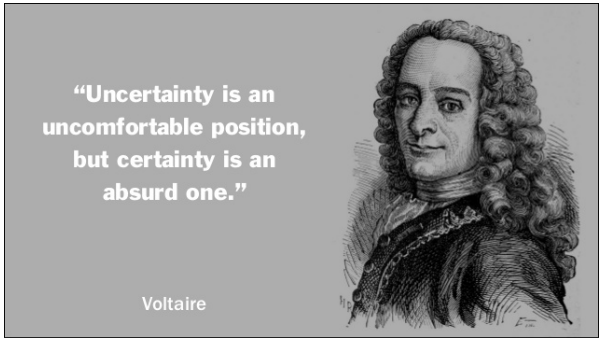
m($t\bar{t}$), 8 TeV



m($t\bar{t}$), 13 TeV (norm.)

Definitions of the modelling uncertainties

- Using factorisation approach to define $t\bar{t}$ modelling systematic uncertainties
  - → each ingredient of the model gets its own uncertainty
  - → split driven by physical effects rather than by the number of tunable parameters
  - → for each term, looking at relevant unfolded distributions (when available)
  - ⇒ long and continuing debate within the LHCTopWg

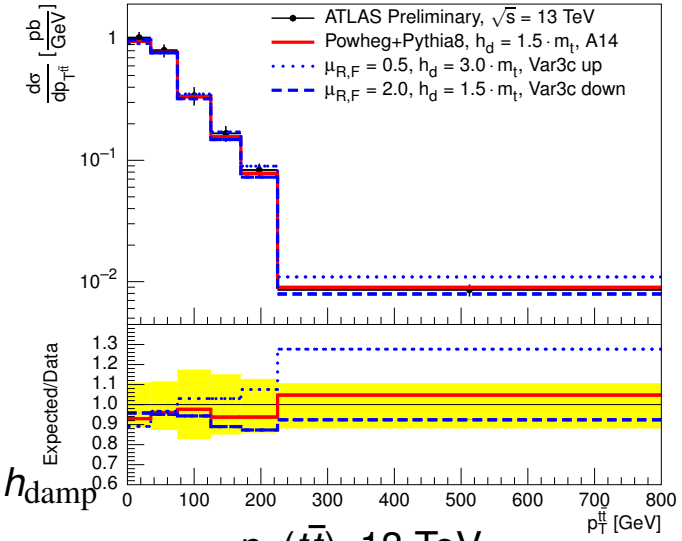


- Additional radiation (Initial- and Final-State Radiation)
  - → radiations come from various sources
  - → simultaneous variation of μ_R and μ_F, h_{damp}, α_S^{ISR}
- Hadronisation & Parton Shower model (PS)
  - → compare Powheg-Box+Pythia8 with Powheg-Box+Herwig7
  - → in the past : Pythia6 vs. Herwig++ (or fHerwig)
- NLO matching scheme
  - → Powheg-Box+P8 vs. MG5_aMC@NLO+P8
  - → in the past : Powheg-Box+Herwig++ vs. MG5_aMC@NLO+Herwig++ (or MC@NLO+fHerwig)
- Colour Reconnection / Underlying Event / Multi Parton Interactions
  - → for now : vary some Pythia8 parameters (A14 tune var1)
  - → ongoing effort to improve this term - see studies with UE observables : ATL-PHYS-PUB-2017-008
- PDFs : PDF4LHC15 prescriptions

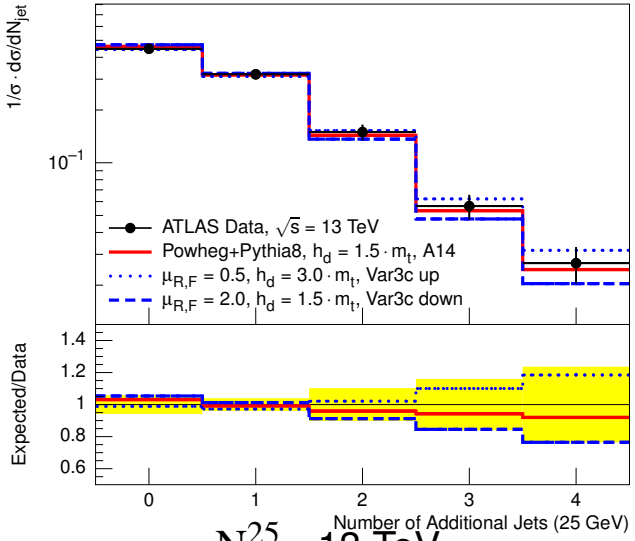
Additional radiation (ISR/FSR)

ATL-PHYS-PUB-2017-007

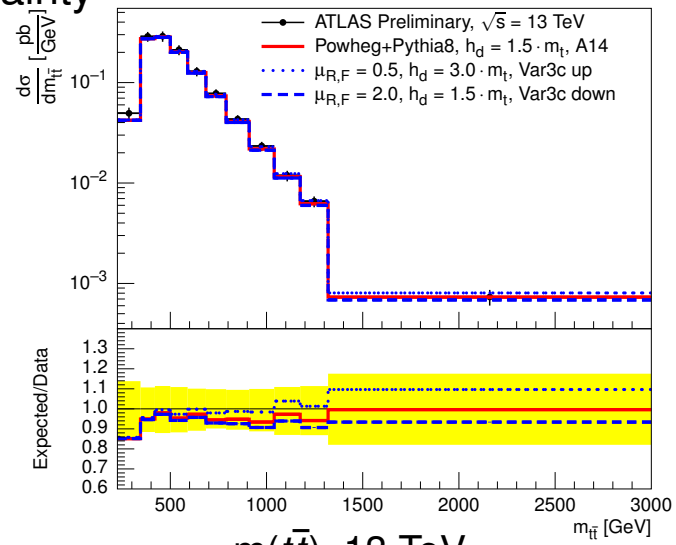
- Many variations were tested
  - → μ_R and μ_F, independently
  - → h_{damp}
  - → A14 tune var3c (i.e. α_S^{ISR})
- Final choice :
  - → up : $0.5 \cdot \mu_{R,F}$, var3c up, $2 \cdot h_{\text{damp}}$
  - → down : $2 \cdot \mu_{R,F}$, var3c down
- Good coverage of N_{jets} uncertainty
- Effect on $t\bar{t}$ kinematics
  - → $p_T(t\bar{t})$, $m(t\bar{t})$
  - → no effect on $p_T(t^{\text{had}})$



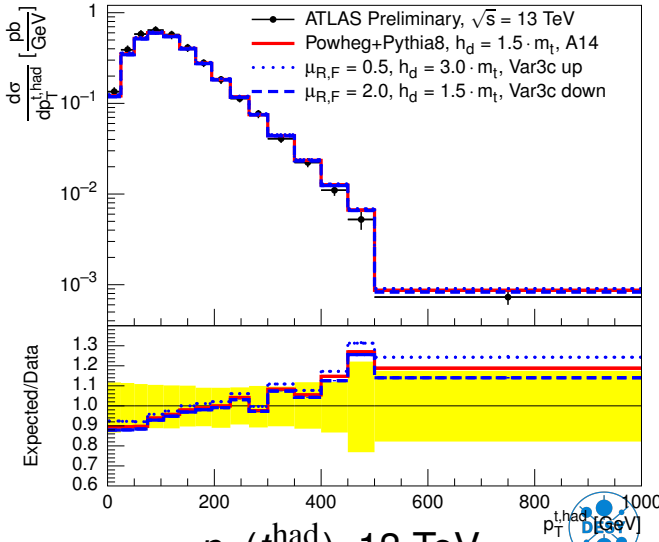
$p_T(t\bar{t})$, 13 TeV



N_{jets}^{25}, 13 TeV



$m(t\bar{t})$, 13 TeV

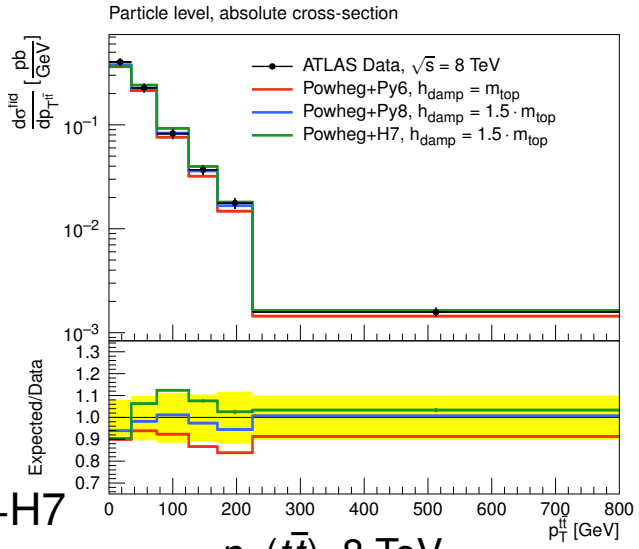


$p_T(t^{\text{had}})$, 13 TeV

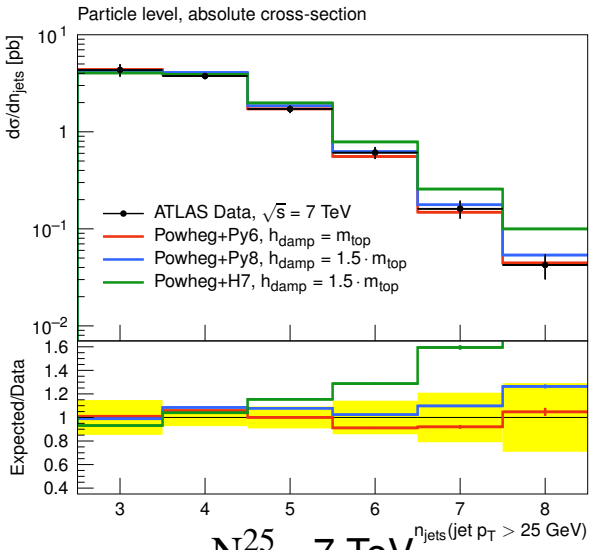
Hadronisation and Parton Shower model

ATL-PHYS-PUB-2016-020

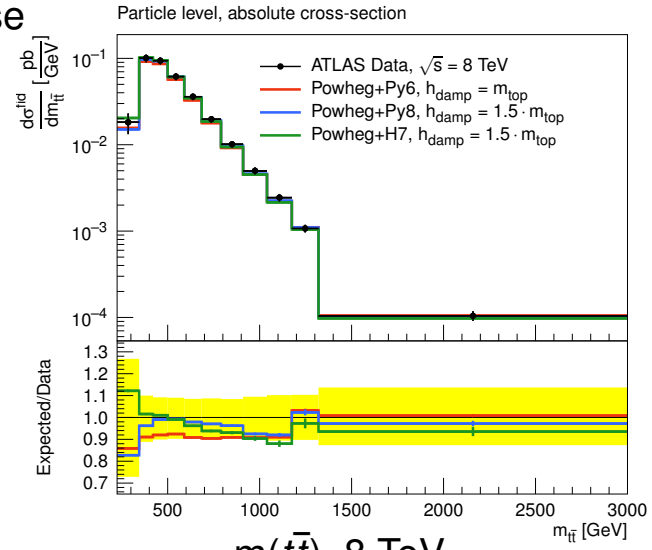
- Hadronisation and PS uncertainty
  - → replace Pythia8 by Herwig7
  - → keeping Powheg-Box for both
- Different top and $t\bar{t}$ kinematics
  - → more pronounced $p_T(t\bar{t})$ shape
  - → large difference at low m($t\bar{t}$)
  - → better p_T(top) modelling !
- Bad N_{jets} modelling with Powheg+H7
  - → mis-configuration in first setup
  - → issue now fixed, setup ok to use
- Additional fixes being worked on
- Hadronisation and PS combined
  - → split them in the longer term ?



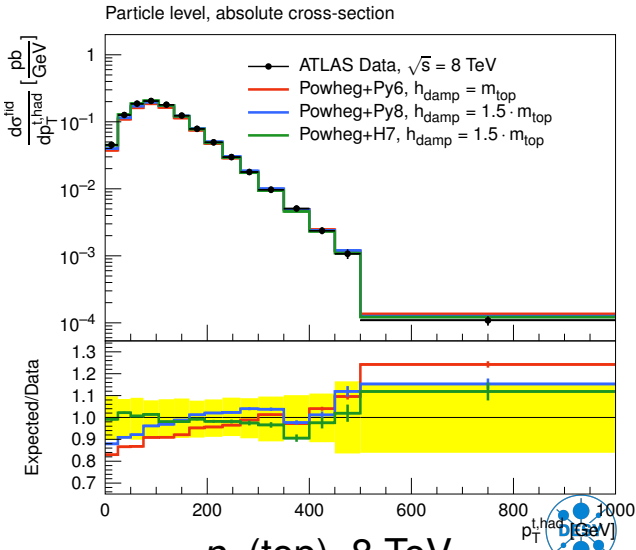
$p_T(t\bar{t})$, 8 TeV



N_{jets}^{25}, 7 TeV



m($t\bar{t}$), 8 TeV

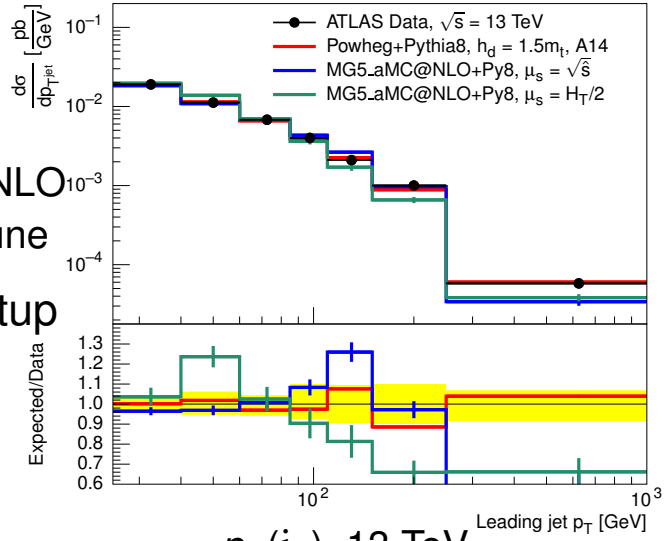


p_T(top), 8 TeV

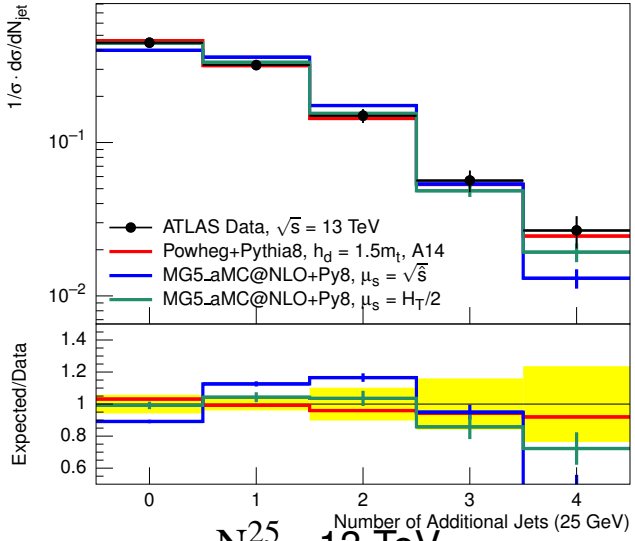
NLO matching scheme : the MG5_aMC@NLO issue

ATL-PHYS-PUB-2017-007

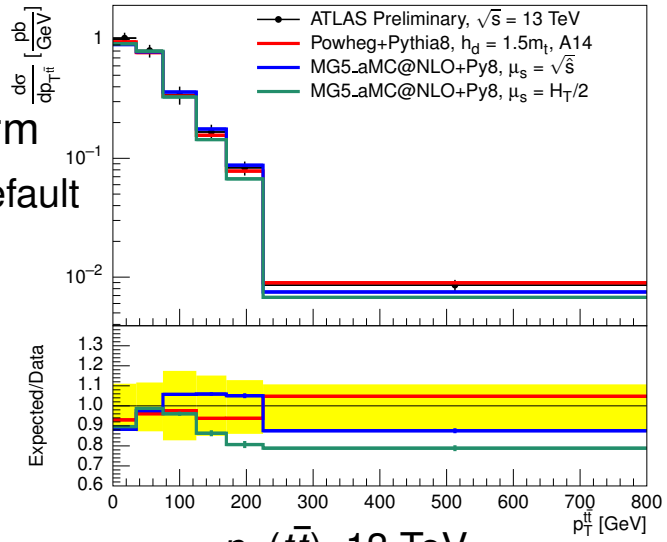
- NLO matching uncertainty
  - → replace Powheg by aMC@NLO
  - → keeping Pythia8 with A14 tune
- Mismodelling with baseline setup
  - → studies perform last spring
  - → bump in N_{jets}, $p_T(j_1)$
- Effect of shower starting scale
  - → test $H_T/2$ instead of $\sqrt{\hat{s}}$
  - → improves N_{jets}, not $p_T(j_1)$
  - → $p_T(t\bar{t})$ is worse
- Since then : new functional form
  - → $H_T(=\sum m_T)$ - new MG5 default
  - → no public plot so far
  - → looks promising



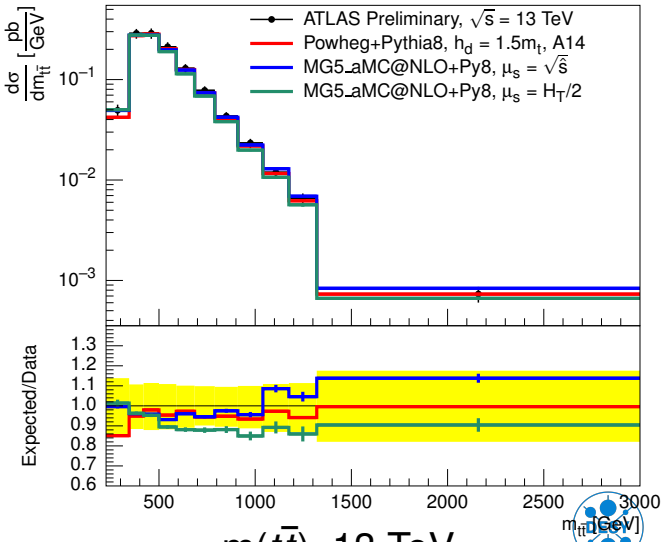
$p_T(j_1)$, 13 TeV



N_{jets}^{25}, 13 TeV



$p_T(t\bar{t})$, 13 TeV

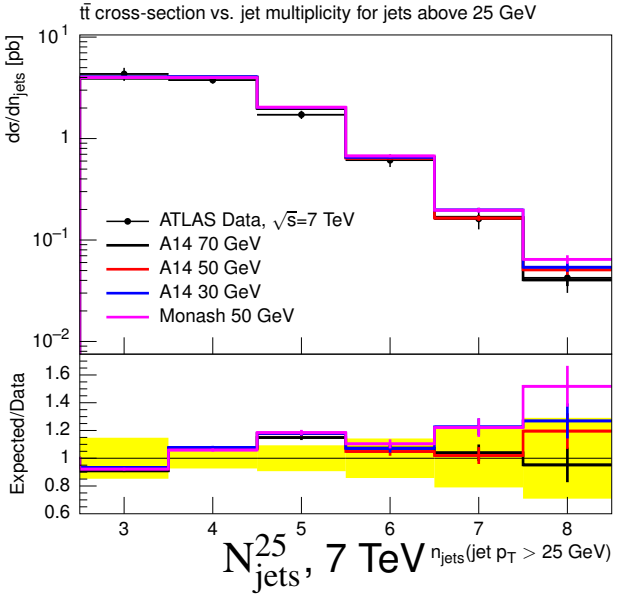


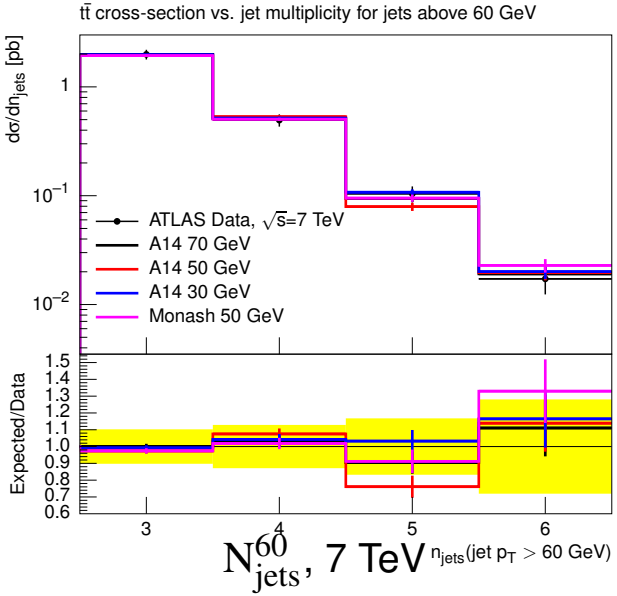
$m(t\bar{t})$, 13 TeV

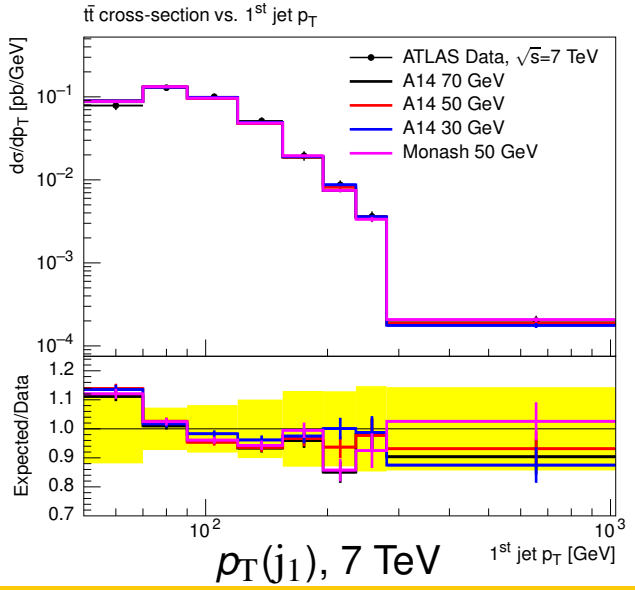
NLO multileg generators : MG5_aMC@NLO+Pythia8 [FxFx]

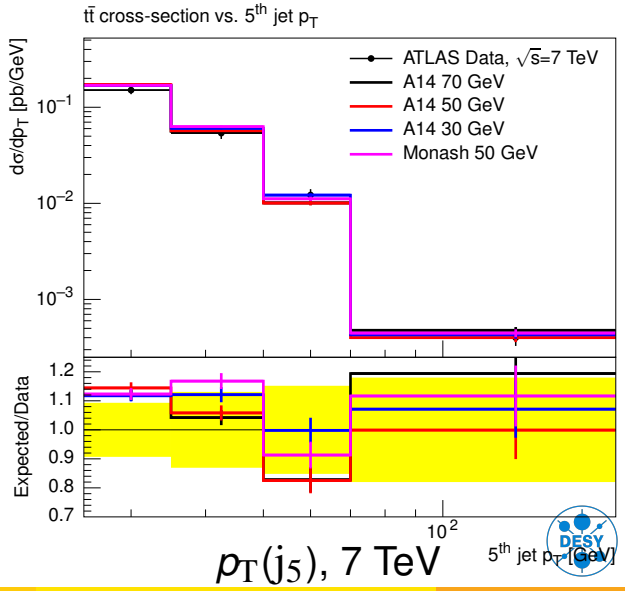
ATL-PHYS-PUB-2016-016

- $t\bar{t} + 0, 1, 2$j@NLO
  - → better model of additional jets ?
- First optimisation last year
  - → MG5_aMC@NLO v2.3.3
  - → Pythia8 A14 tune
  - → 7 TeV data
- Effect of the merging scale
  - → no obvious optimal value
- New shower starting scale
  - → FxFx studies to be re-done
- New unfolded data will be crucial
  - → Njets, jet pT @13TeV
  - → $t\bar{t}$ kinematics vs. Njets





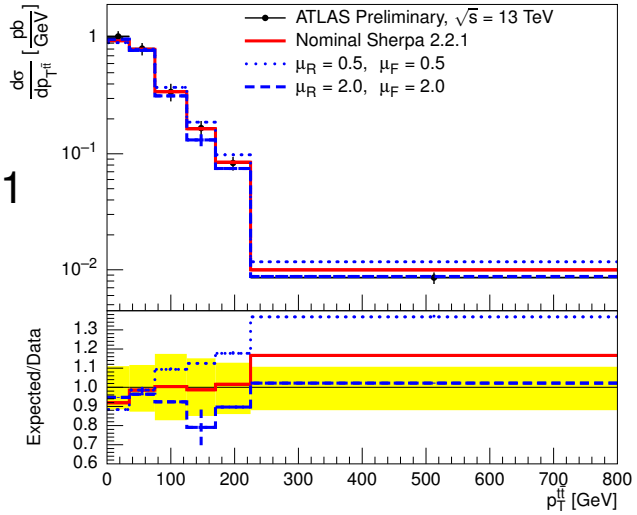




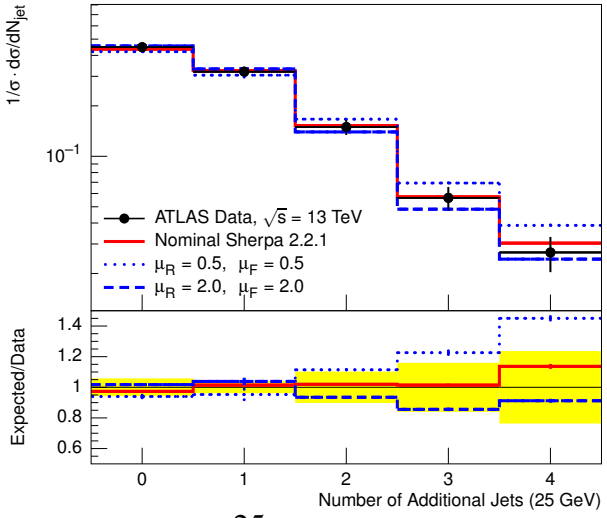
NLO multileg generators : Sherpa

ATL-PHYS-PUB-2017-007

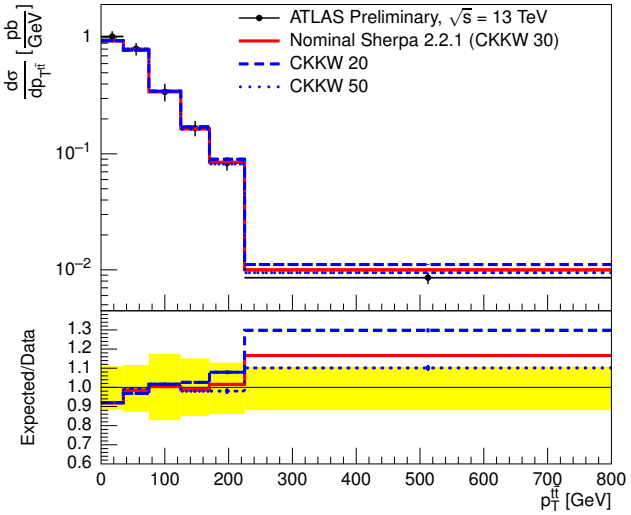
- $t\bar{t} + 0, 1\text{j@NLO} + 2, 3, 4j\text{@LO}$
  - → another NLO multileg option
- Various studies with Sherpa 2.2.1
  - → rather good nominal model
  - → several tunable parameters
- E.g. : $\mu_{R,F}$
  - → effect on additional radiation
  - → larger than for Powheg
- E.g. : ME matching scale (CKKW)
  - → nominal : 30 GeV
  - → alternatives : 20 and 50GeV
  - → smaller effect on radiation
- Other possible parameters :
  - → resummation scales
  - → PS recoil scheme
  - → DIRE shower vs. CSShower
  - → ...



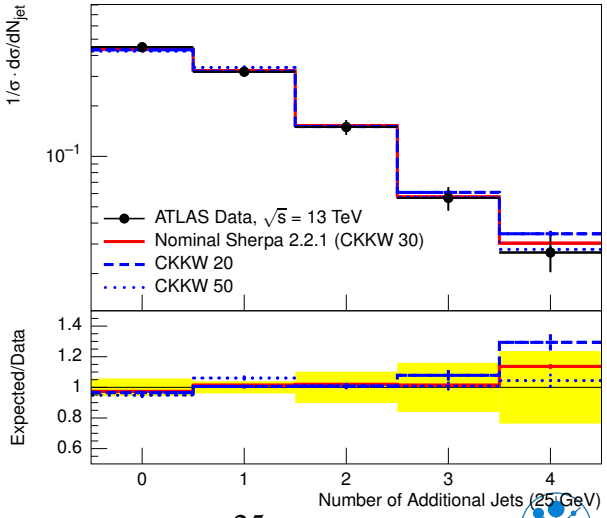
$p_T(t\bar{t})$, 13 TeV



N^{25}_{jets}, 13 TeV



$p_T(t\bar{t})$, 13 TeV

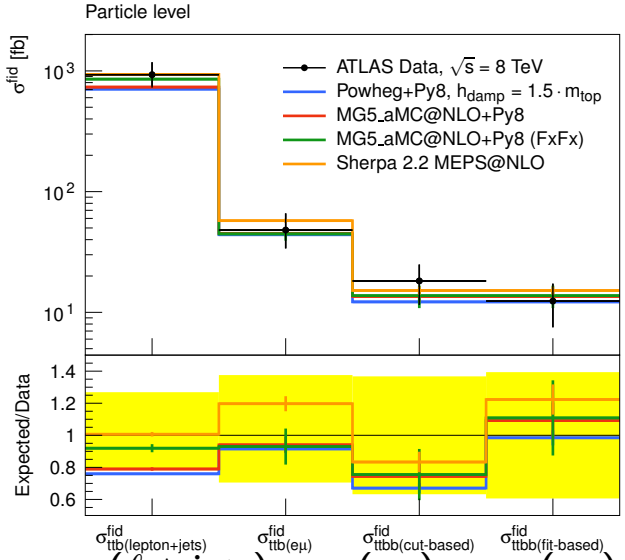


N^{25}_{jets}, 13 TeV

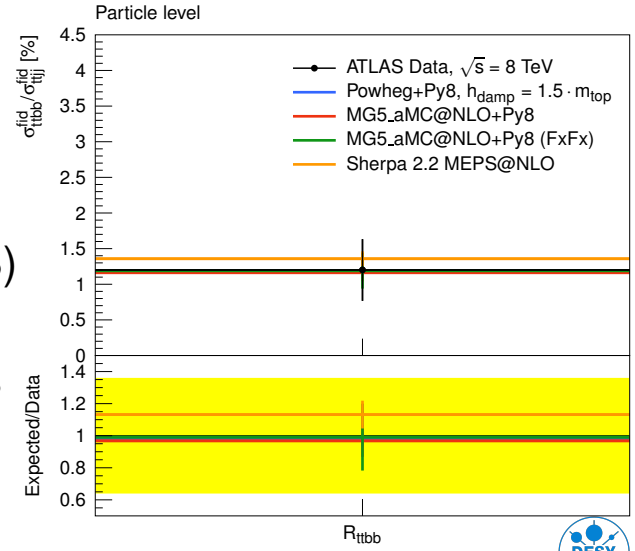
Other processes

ATL-PHYS-PUB-2016-020

- Single-top t-, s-channel and tW
  - → still using Powheg+P6 - migration to Powheg+P8 will follow
  - → now in Powheg-Box v2 - can now use ME reweighting !
  - → first unfolded data are coming - toward better optimisation ?
  - → new $t\bar{t}/tW$ overlap treatment in MG5_aMC@NLO (DR2)
  - → a lot of activity for $WWbb$ production (only dilepton for now)
- New process : single-top tZ
  - → MG5_aMC@NLO+P6 - use similar settings as t-channel
  - → migrating to P8 as well
- $t\bar{t} + V$: using MG5_aMC@NLO+P8
  - → systematics : $\mu_{R,F}$ variations, A14 variations, Herwig++, Sherpa
  - → no unfolded data yet
- $t\bar{t}$+HF : mostly with inclusive $t\bar{t}$ generators (HF-jets from the PS)
  - → Sherpa 4FS $t\bar{t} + b\bar{b}$ prediction available - very CPU consuming
  - → 8 TeV measurement : inclusive cross-sections in fiducial volumes
  - → no actual tune yet possible
  - → large differences between generators for the $t\bar{t}$+HF fractions...



$\sigma_{ttb}(\ell + \text{jets})$, $\sigma_{ttb}(e\mu)$, $\sigma_{ttbb}(e\mu)$



$R_{ttbb} = \sigma_{ttbb}/\sigma_{ttjj}$

Future improvements and wishlist

- Powheg+Pythia8 with A14 tune describes data well - no full retuning planned
- Developments mostly in two directions :
  - → all-in-one setups - all systematics with the same generator
  - → NLO multileg - to improve modelling in exclusive jet bins
- Herwig7 could replace several if not all variations
  - → allows Powheg and MC@NLO matching
  - → angular- and dipole-ordered showers (only string model, though)
  - → FxFx merging available - plans for UNLOPS maybe ?
  - → issue with photon radiation being worked on
- MG5_aMC@NLO+P8 [FxFx]
  - → need systematics - e.g. merging scale, UNLOPS, Herwig7
  - → new shower starting scale in MG5_aMC@NLO should help
  - → also : ongoing studies for MG5_aMC@NLO+H7
- Sherpa
  - → need meaningful systematic uncertainties - which parameters should we vary ?
  - → can/should we plug-in EvtGen ?
  - → $t\bar{t} + b\bar{b}$ 4FS setup available - recipe to mix with $t\bar{t}$ 5FS ?
  - → lack of control in the shower - tune to LHC data ?
- Large fractions of negative weights increases the statistical uncertainty !
  - → negative weights are inherent to MC@NLO matching method
  - → some setups have 1/3 negative-weighted events in some phase-space
  - → in such case, we need to generate 9 times more events...

Conclusions

- Improvement of top-quark MC models is a continuous process
  - → requires more and more unfolded data to better model more variables
  - → ...and a wiser choice of tunable parameters with more and more complex generators
- Controversial points not addressed in this talk :
  - → overtuning and predictivity
  - → two-point systematic uncertainties
- ATLAS has reached a satisfactory NLO+PS setup
  - → Powheg+Pythia8 (A14), h_{damp}=1.5 $\cdot\, m_t$ as nominal
  - → comparison to other NLO+PS setups for systematics
- Ongoing effort to go beyond that
- Our input : unfolded ATLAS data - see the many results shown this week !
  - → $t\bar{t}$ cross-sections : Michele's talk (yesterday), Petr's poster (yesterday)
  - → single-top cross-sections : Lidia's talk (earlier today), Rui's and Irina's posters (yesterday)
  - → top-quark properties : Fabian's (later today) and Frederic's (tomorrow) talks
  - → $t\bar{t}+\gamma$ cross-sections : Lisa's talk (tomorrow), Mazuza's poster (yesterday)
- ~~Stay tuned !~~ Keep tuning !