Challenging Measurements with *b*-jets @ATLAS GE

F. Sforza - Riunione di Gruppo 1 INFN GE- 24/02/2021

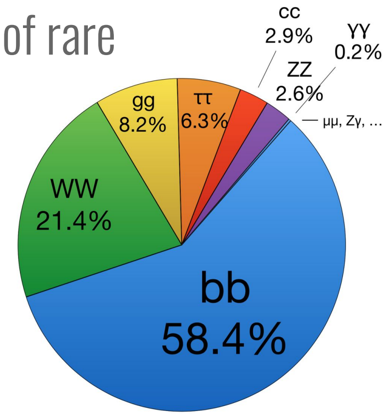
The importance of Beauty (quark) at LHC

Beauty (or *b*) quarks - of m_b ~ 4.2 GeV - **are key items when exploring fundamental physics**

⇒ *Two examples where ATLAS Genoa has leading role & recent results*

A window on the Higgs properties:

- Dominating $H \to b\bar{b}$ branching ratio
- High statistics (but large background) test of rare Higgs production modes or in challenging phase space



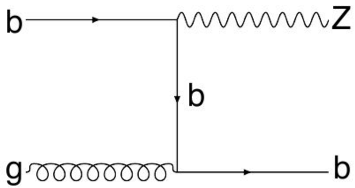
⇒ *BSM sensitivity!*

A probe of complex QCD processes (new@GE !)

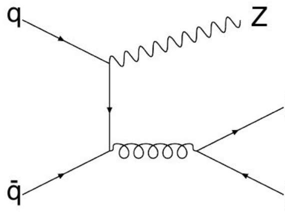
- Heavy-flavour (HF) quark mass in QCD predictions?
- What about the HF content of the proton PDFs?
- Reliability of state-of-the-art MC generators at LHC?

E.g. LO diagram for Z+*b* in different QCD “Flavour Schemes” (***FS***):

5FS: Massless *b* in proton PDF



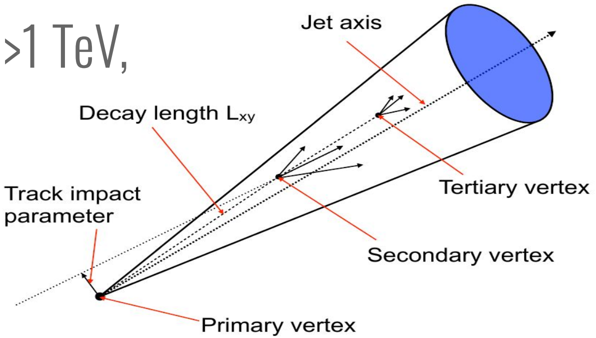
4FS: gluon splitting to massive *b*’s



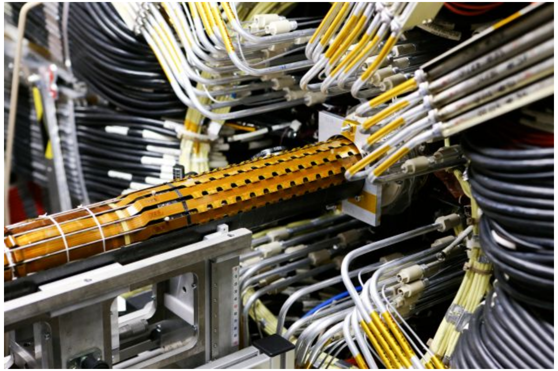
High-p_T *b*-jet identification in ATLAS: Detector and Algorithms

Identification of *b*-quarks inside high-p_T jets, from ~20 GeV to >1 TeV, relies on typical characteristics of B-hadrons in the *b*-jet:

- Long lifetime, secondary vertices, decay pattern, etc.
- Greatly improved in LHC Run 2 thanks to installation of new pixel detector (IBL) very close to beamline (~3cm)

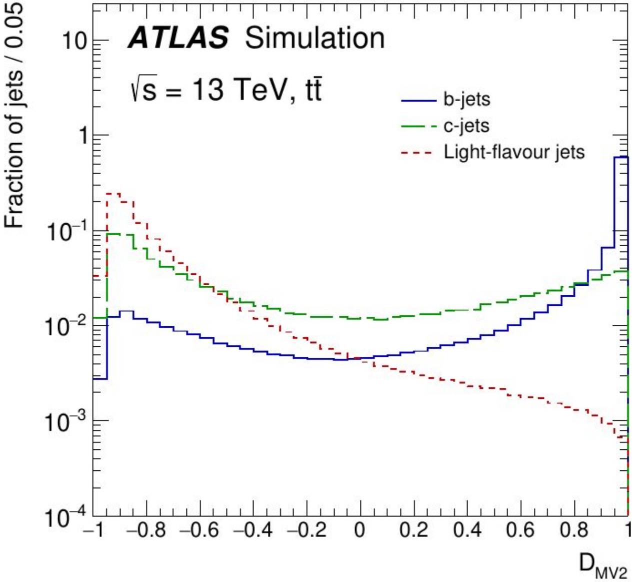


JINST 13 T05008 (2018)



INFN Genoa led IBL construction and now upgrade towards HL-LHC

- Multiple tracking and jet information condensed using multivariate (MV) algorithms, and *Deep NN,* for optimal identification of *b*-jets vs different flavour jets (c or light-flavour jets)



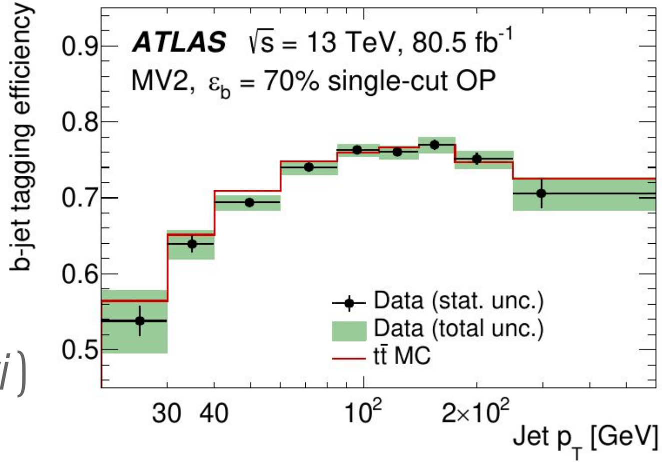
High-p_T *b*-jet identification: Selection and Efficiency Calibration

- In practical terms the *b*-jet identification proceeds with a *“cut”* on the MV algorithm output

WP	Cut value X	b-jet efficiency (ε_b)	c-jet mistag rate (ε_c)	LF-jet mistag rate (ε_{LF})
85%	0.1758	85%	32%	2.9%
77%	0.6459	77%	16%	0.77%
70%	0.8244	70%	8.3%	0.26%
60%	0.9349	60%	2.9 %	0.065%
50%	0.9769	50%	0.94 %	0.017%

- *b*-jet identification performance evaluated on MC but crucial to calibrate it with reference candles in *Data* ⇒ inaccuracy in detector description or QCD simulation of *b/c/light*-jets

- Example of $t\bar{t}$ events used for the extraction of the efficiency correction (EPJC 79(2019)970, with the leading contribution of *A. Coccaro*)



NB: MV algorithm optimization, MC efficiency calibration, etc. handled by *Flavour Tagging ATLAS Group* (convened by *C. Schiavi*)

ATLAS Run 2: measurements limited by systematics or statistics

All the (quickly) described, complex infrastructure needed to obtain challenging measurements!

Presenting in the next slides 3 recent measurements done using the ATLAS Run 2 dataset:

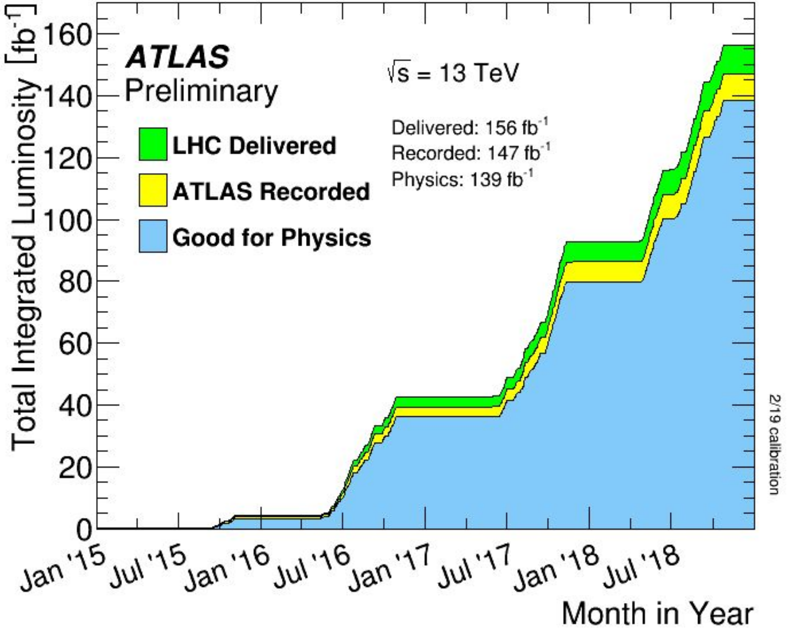
Discussed in more detail (leading author *F.S.*):

- *Z+b*- jets cross section [JHEP 07 (2020) 044]
  ⇒ 36 fb^{-1} of data, systematic limited

Presenting only the results:

- VHbb in boosted regime [arXiv:2008.02508]
  ⇒ 139 fb^{-1}, statistically limited measurement
  (Chair of internal review board was *A.Coccaro*)

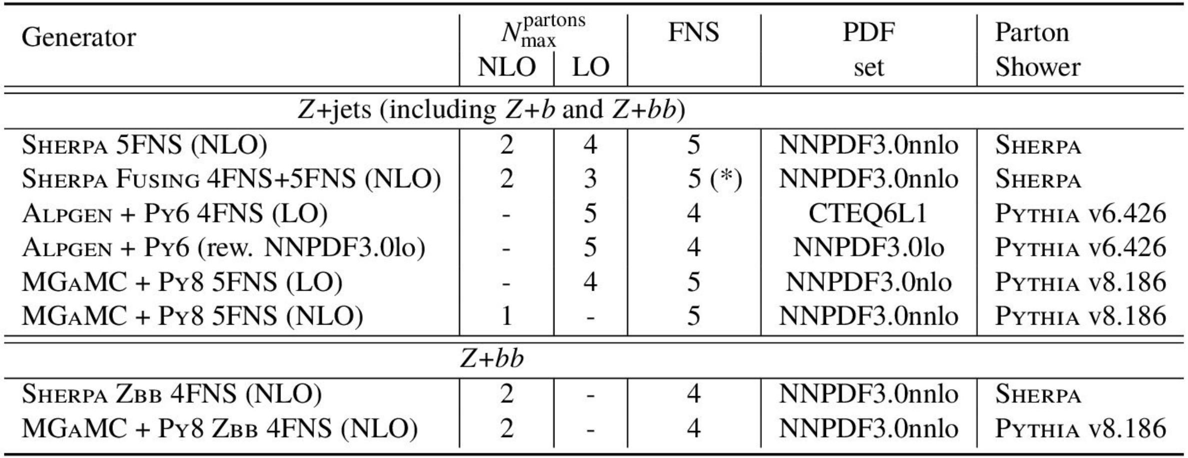
- Search for VBF(+γ) Hbb [arXiv:2010.13651]
  ⇒ 139 fb^{-1}, stat. limited (ATLAS GE research line in VBF, chair of internal review board was *F.Parodi*)



State-of-the-art of Z+*b*-jet theoretical predictions

Measurement challenges theoretical predictions for vector boson (Z here) production+b quarks:

Examples of the tested QCD MC predictions ⇒



Generator	$N_{\max}^{\text{partons}}$ NLO	$N_{\max}^{\text{partons}}$ LO	FNS	PDF set	Parton Shower
Z+jets (including *Z*+*b* and *Z*+*bb*)					
Sherpa 5FNS (NLO)	2	4	5	NNPDF3.0nnlo	Sherpa
Sherpa Fusing 4FNS+5FNS (NLO)	2	3	5 (*)	NNPDF3.0nnlo	Sherpa
Alpgen + Py6 4FNS (LO)	-	5	4	CTEQ6L1	Pythia v6.426
Alpgen + Py6 (rew. NNPDF3.0lo)	-	5	4	NNPDF3.0lo	Pythia v6.426
MGaMC + Py8 5FNS (LO)	-	4	5	NNPDF3.0nlo	Pythia v8.186
MGaMC + Py8 5FNS (NLO)	1	-	5	NNPDF3.0nnlo	Pythia v8.186
Z+*bb*					
Sherpa Zbb 4FNS (NLO)	2	-	4	NNPDF3.0nnlo	Sherpa
MGaMC + Py8 Zbb 4FNS (NLO)	2	-	4	NNPDF3.0nnlo	Pythia v8.186

NB: V+HF is also the main background for V+Higgs measurements ⇒ crucial to have reliable MC simulation

- Different levels of ME evaluation accuracy ⇒ different number of partons, LO, NLO

E.g. FxFx prescriptions by S. Frixione et al. [JHEP 12 (2012) 061]

- Different Flavour Scheme used for PDF set ⇒ 4 FS in PDFs allows use of b-quark mass in ME, while 5FS allows resummation of large logarithm inside the PDF itself
- Different approaches for approximate use of HF quark mass in parton-shower (PS)

Key analysis components: selection & data-driven background

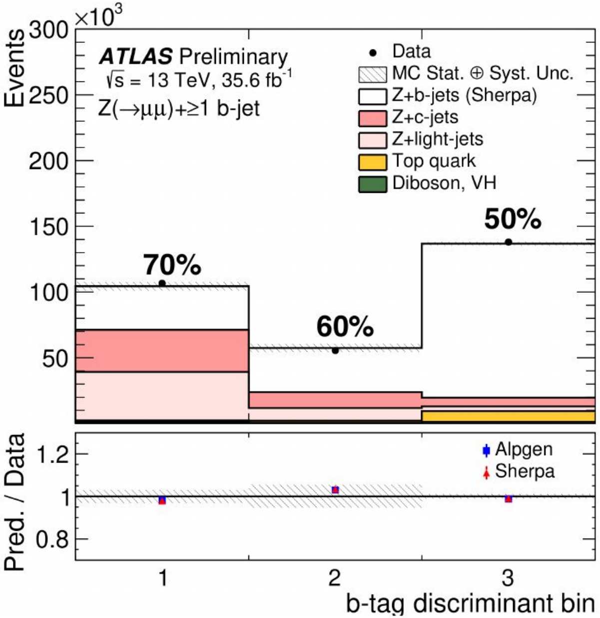
Main elements of the Z+b-jets signal candidate event selection:

- 2 opposite charge light leptons (ee/$\mu\mu$), with p_T>27 GeV, $|\eta|$<2.5, and 76<m_{ll}<106 GeV
- Jets passing "cut" on MV *b*-tagging algorithm corresponding of 70% efficiency for b-jets
- ≥1 or ≥ 2 b-jets with p_T>20 GeV and $|\eta|$<2.5 ⇒ measurement of *Z+≥1 b*-jet and *Z+≥2 b*-jets

Z+jets background affected by large uncertainties would provide poor background subtraction ...

⇒ Data-driven multi-component template fit using MV discriminant binned according to *b*-tag purity cuts

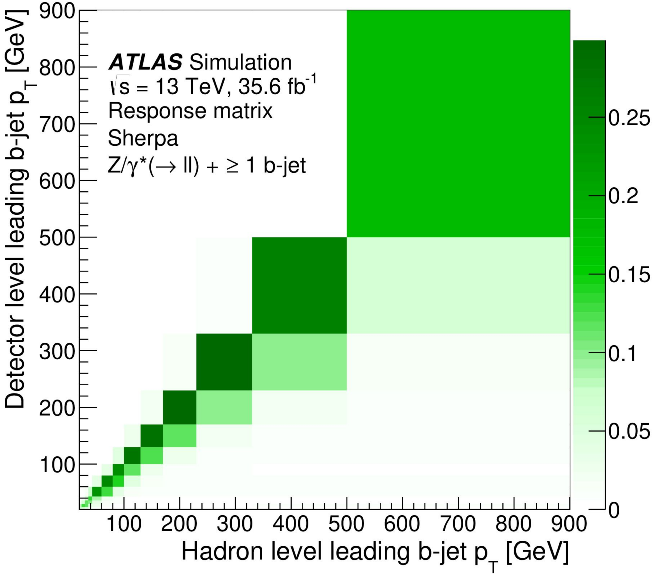
⇒ extracted *Z+c*-jets and *Z+light*-jets then subtracted from data (together with other minor backgrounds)



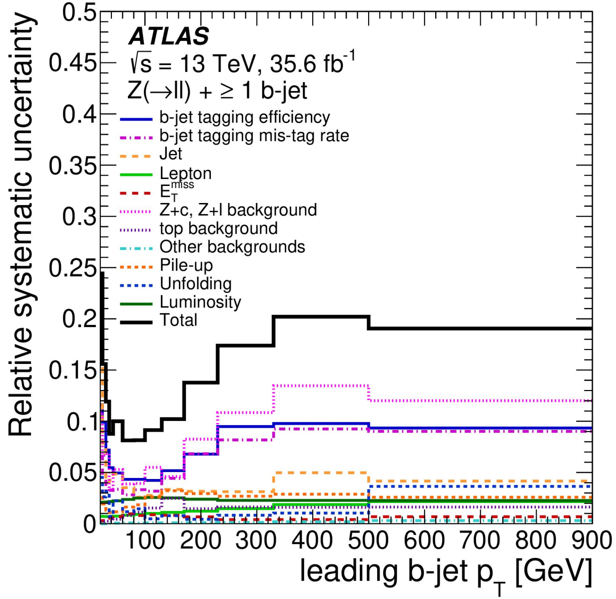
Correction to particle level (Unfolding) and systematic errors

Data background subtracted events brought to particle-level correcting for efficiency and detector resolution effects using *Z+b*-jets signal MC, after application of *b*-jet efficiency calibration

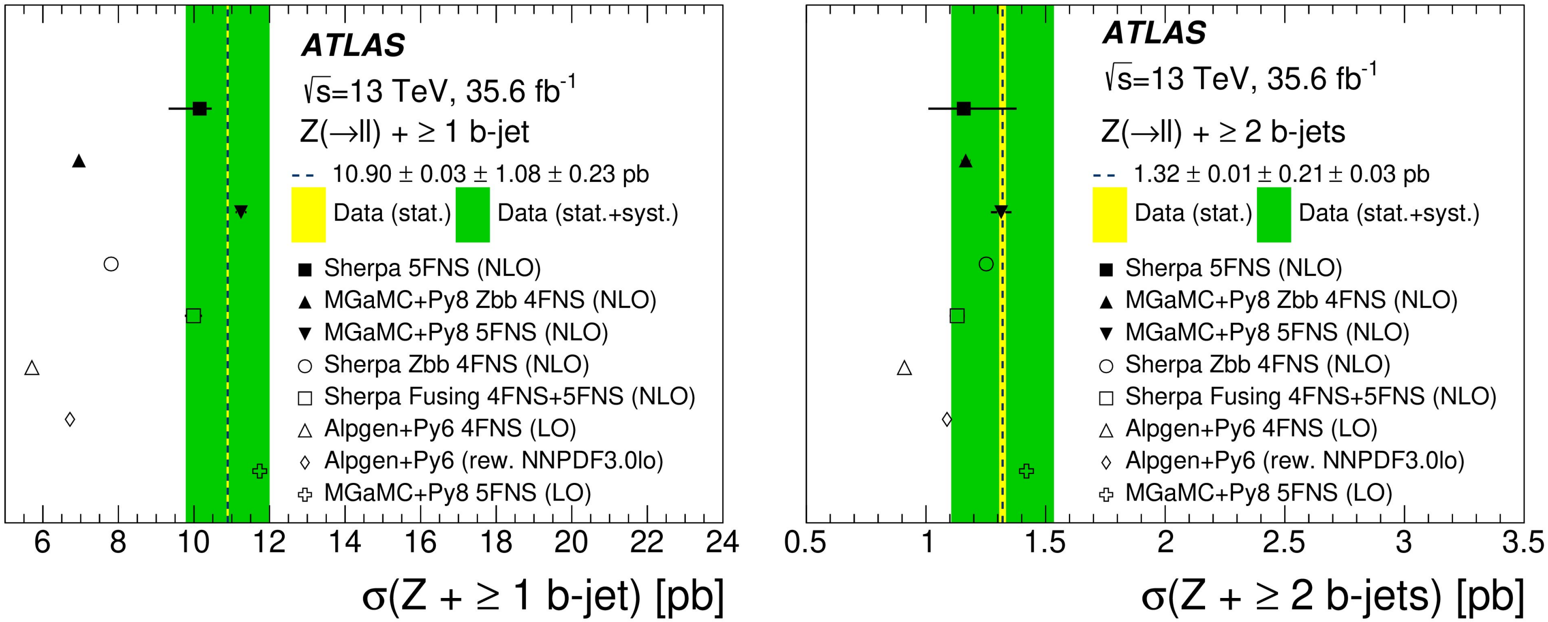
E.g. resolution effect on *b*-jet p_T (unfolding with iterative Bayesian method [arXiv:1010.0632])



Main syst. uncertainties due to *b*-jet eff. calibration and background subtraction:



Z+≥1 b-jet and Z+≥2 b-jets inclusive cross-section results



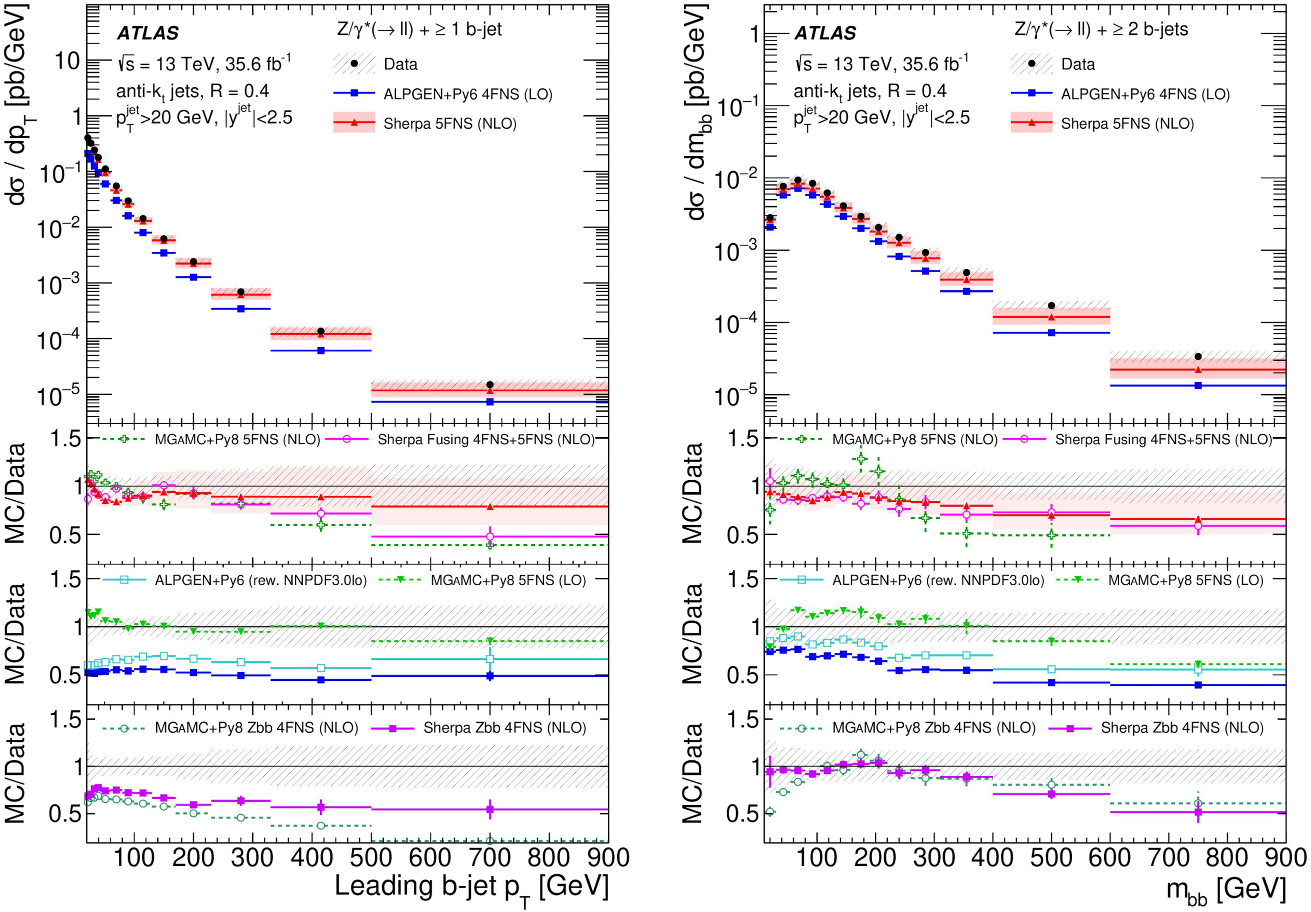
- 4FS undershoots *Z+≥1 b*-jets in all configuration ⇒ ***closing a long lasting discussion!***
- *Z+≥2 b*-jets uncertainties still too large to favour any of the more recent predictions

Differential cross-section measurements: high-p_T challenges

5FS predictions show reasonable agreement with data except in the very high-pT regime:

⇒ large uncertainties and tensions with data

Challenge for searches or other process test in such phase space…



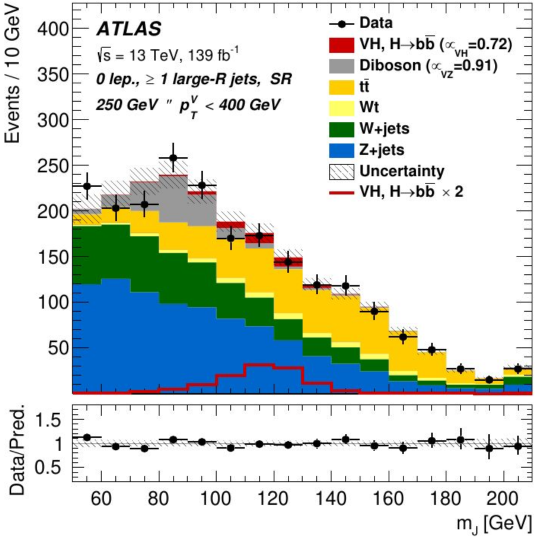
First glance to extreme high-p_T *b*-jets for Higgs measurements

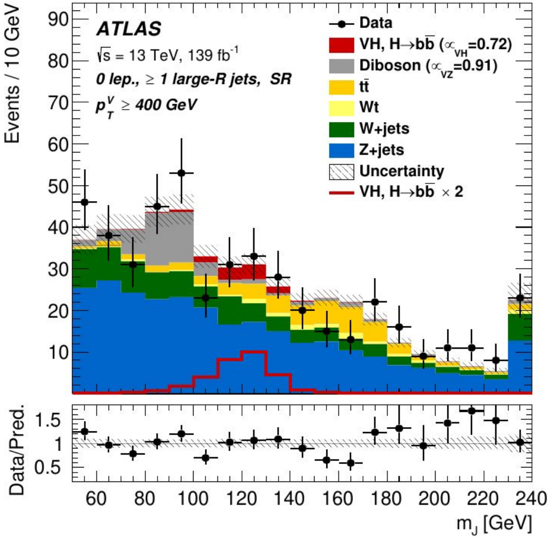
Boosted regime, e.g. V p_T>250 GeV, very challenging (as seen) but may be more affected by BSM!

Vector boson (V=W or Z) plus Higgs associated production is the main analysis channel for the study of H→bb properties because V leptonic decays allows trigger and background suppression. *Observation* of H→bb in VH [arXiv:1808.08238]⇒ now investigating differential cross section

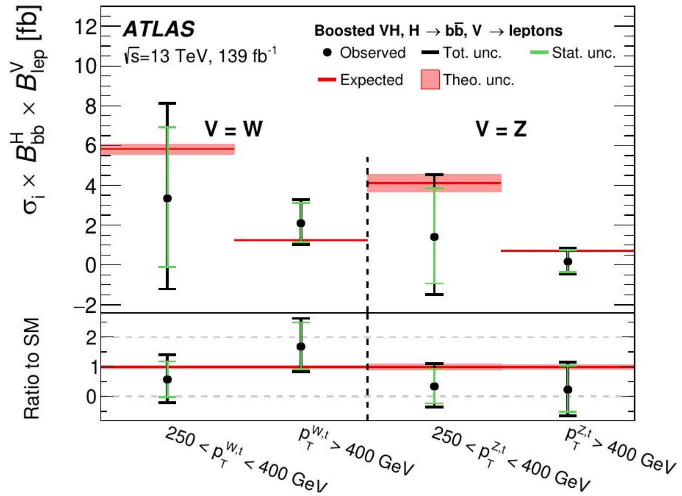
A new analysis [arXiv:2008.02508]

- Reconstruction of large radius *R=1.0* jets with two *b*-tagged sub-jets
- Two "differential" Vp_T bins





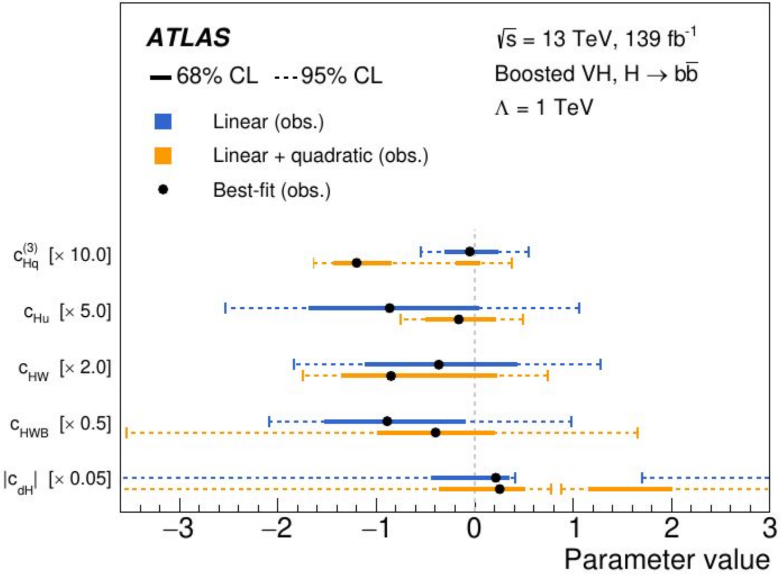
Boosted VH⟶ bb Interpretation in an EFT Framework



Differential result ⇒ statistically limited & large V+b-jets systematic uncertainty, but prepares us for future much larger LHC Run 3 and HL-LHC datasets

⇒ Also immediat probe of any strong BSM deviation in Effective Field Theory (EFT) coefficients:

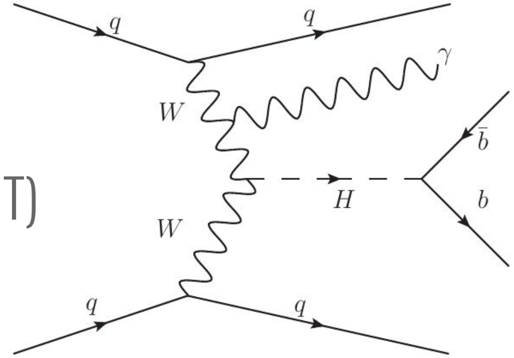
$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_d \frac{1}{\Lambda^{d-4}} \left(\sum_i c_i^{(d)} O_i^{(d)} \right)$$

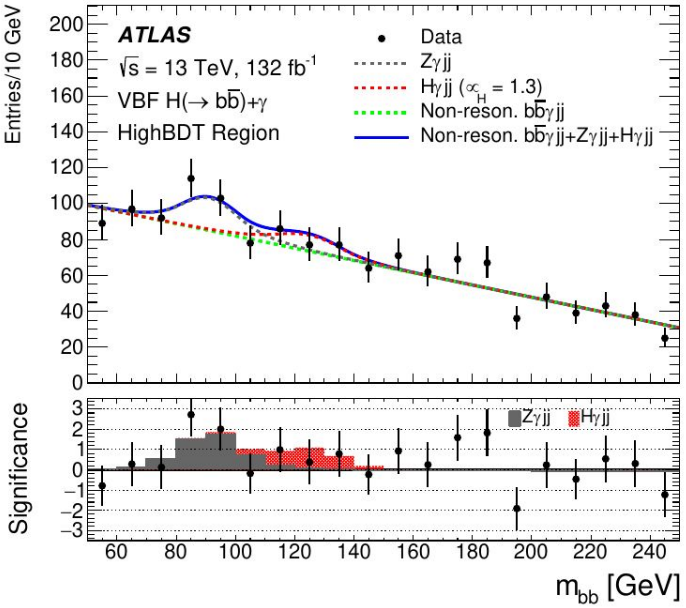


Higgs VBF: measuring another rare production mode with *b*-jets

Higgs VBF production mode is even harder to reach because of very large multi-jet background

- Full Run 2 dataset allows a first glance using prompt photon emission for trigger and background suppression





- Boosted Decision Tree (BDT) used to further reduce multi-jet background
- Data-driven lineshape fit for extraction of non-resonant background
- First hints of H⟶bb presence in this channel:

⇒ Signal significance 1.3σ

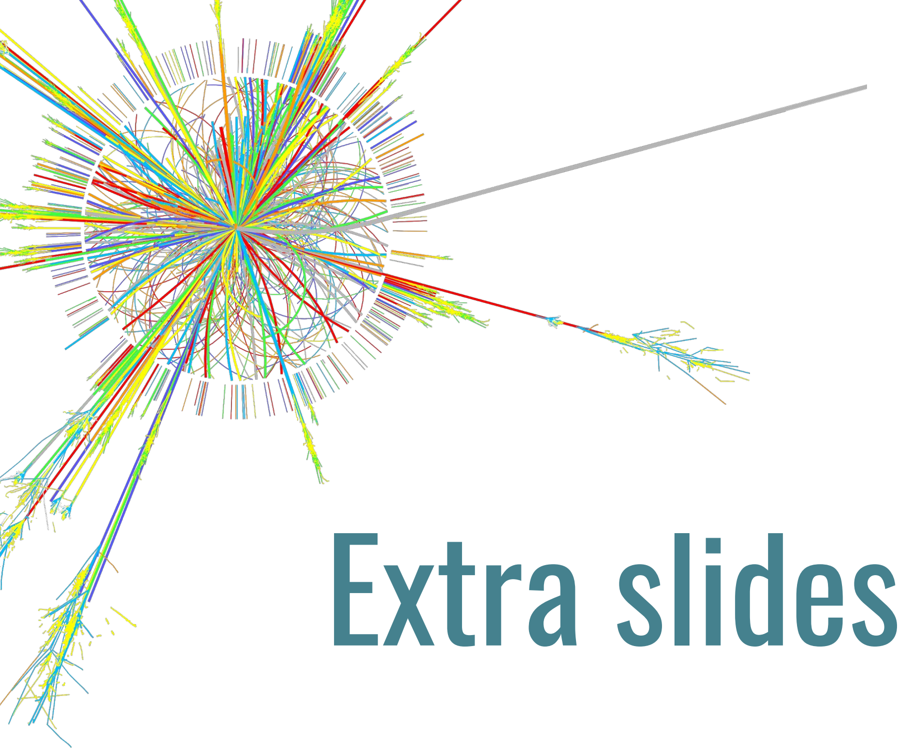
⇒ Cross section strength w.r.t. SM μ = 1.3±1.0

What’s next?

Towards Run 3: higher precision for the Higgs boson and beyond

What's next? ***Run 3 is coming...***

- Further effort in *b*-tagging algorithms optimization also thanks to a new PhD student (*M. Tanasini*, welcome!)
- Current analyses are paving the way towards higher precision measurements in Run 3:
  → Further optimized *b*-jet identification algorithms
  → Extended dataset statistics
  → Largely Improved MC simulation
- QCD theory predictions have comparable level of precision to measurement
  ⇒ it seems now at hand to bring measurements to the next level of precision
- Need to push precision for H$\rightarrow b\bar{b}$ Yukawa measurement and VH couplings beyond "simple" observation in order to be sensitive to BSM



Extra slides & Backup Material

References to recent papers with main ATLAS GE contribution

Z+b-jets differential measurement:
https://arxiv.org/abs/2003.11960

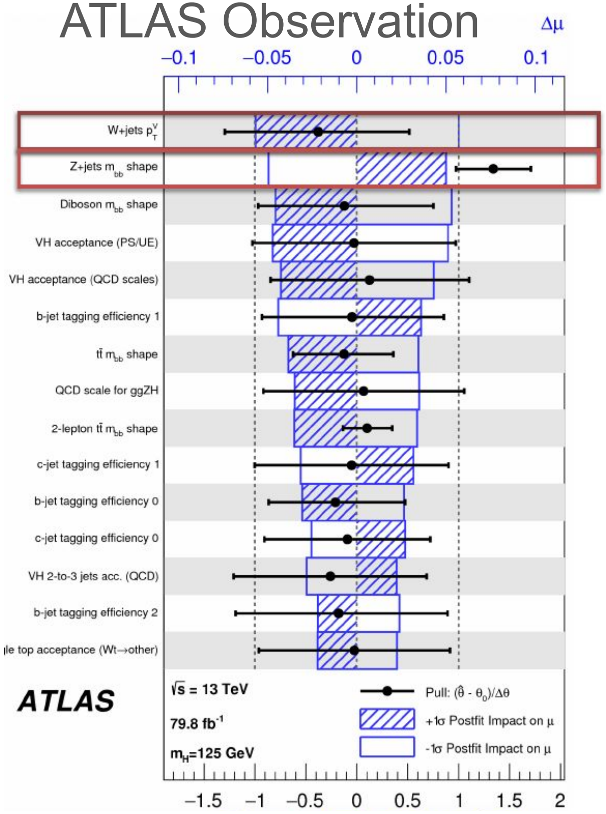
VH→ bb boosted:
https://atlas-glance.cern.ch/atlas/analysis/papers/details.php?id=12369
https://arxiv.org/abs/2008.02508

Hbb VBF(+photon): https://atlas-glance.cern.ch/atlas/analysis/papers/details.php?id=12805
https://arxiv.org/abs/2010.13651

b-tag performance:
https://atlas-glance.cern.ch/atlas/analysis/papers/details.php?id=11805
https://arxiv.org/abs/1907.05120

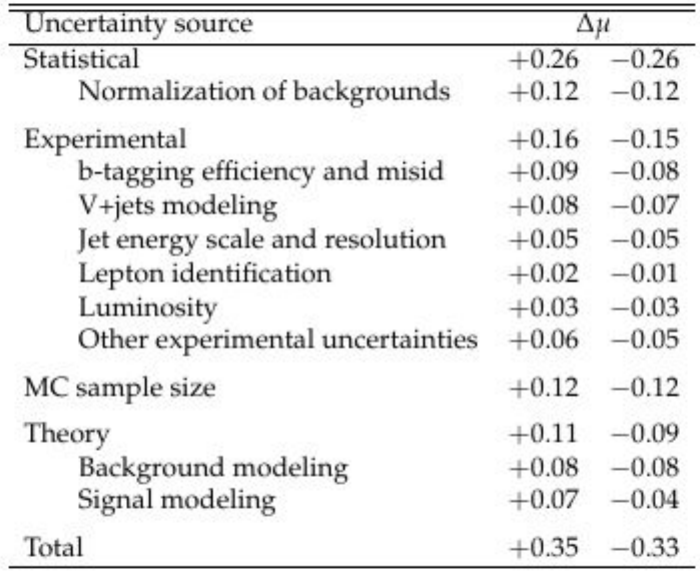
Impact of V+b-jets uncertainties on VHbb

ATLAS Observation



https://arxiv.org/abs/1808.08238

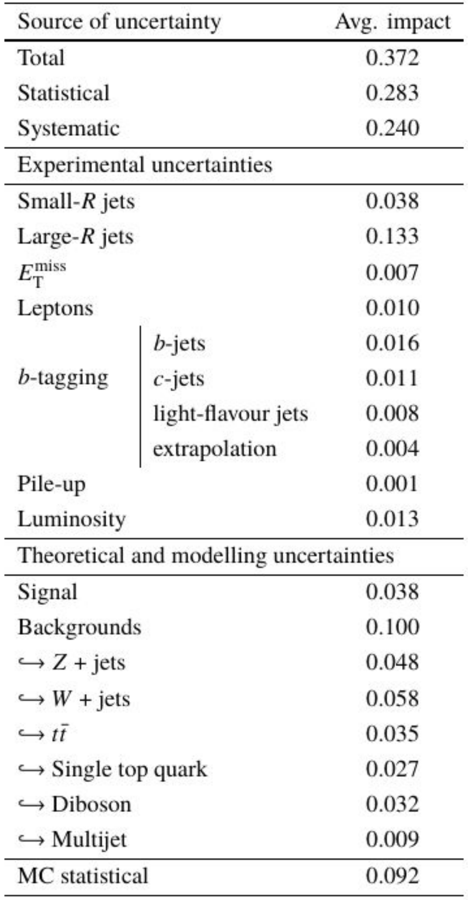
CMS Observation



Uncertainty source	$\Delta\mu$	
Statistical	+0.26	−0.26
Normalization of backgrounds	+0.12	−0.12
Experimental	+0.16	−0.15
b-tagging efficiency and misid	+0.09	−0.08
V+jets modeling	+0.08	−0.07
Jet energy scale and resolution	+0.05	−0.05
Lepton identification	+0.02	−0.01
Luminosity	+0.03	−0.03
Other experimental uncertainties	+0.06	−0.05
MC sample size	+0.12	−0.12
Theory	+0.11	−0.09
Background modeling	+0.08	−0.08
Signal modeling	+0.07	−0.04
Total	+0.35	−0.33

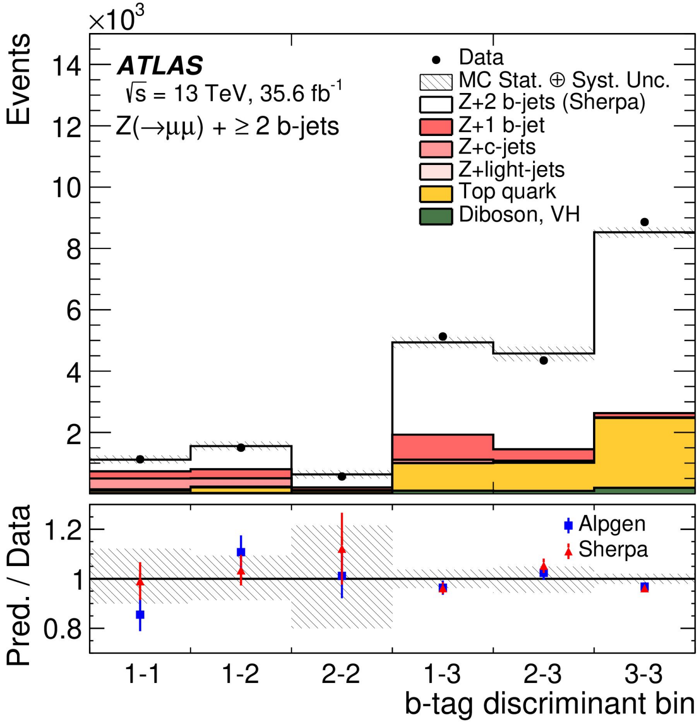
https://arxiv.org/abs/1808.08242

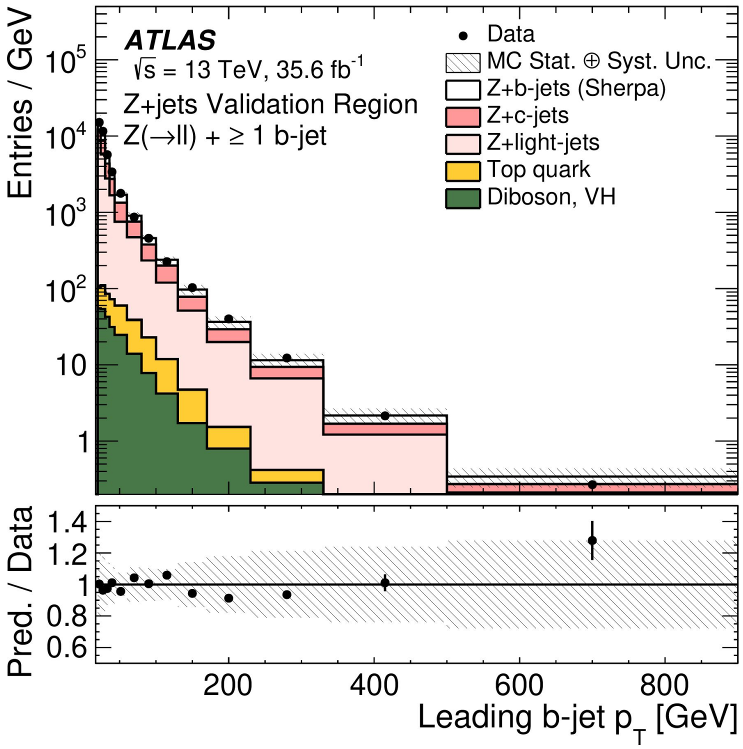
ATLAS VH boosted



Source of uncertainty		Avg. impact
Total		0.372
Statistical		0.283
Systematic		0.240
Experimental uncertainties		
Small-R jets		0.038
Large-R jets		0.133
E_T^{miss}		0.007
Leptons		0.010
b-tagging	b-jets	0.016
	c-jets	0.011
	light-flavour jets	0.008
	extrapolation	0.004
Pile-up		0.001
Luminosity		0.013
Theoretical and modelling uncertainties		
Signal		0.038
Backgrounds		0.100
↪ Z + jets		0.048
↪ W + jets		0.058
↪ $t\bar{t}$		0.035
↪ Single top quark		0.027
↪ Diboson		0.032
↪ Multijet		0.009
MC statistical		0.092

Data driven Z+jets background and validation



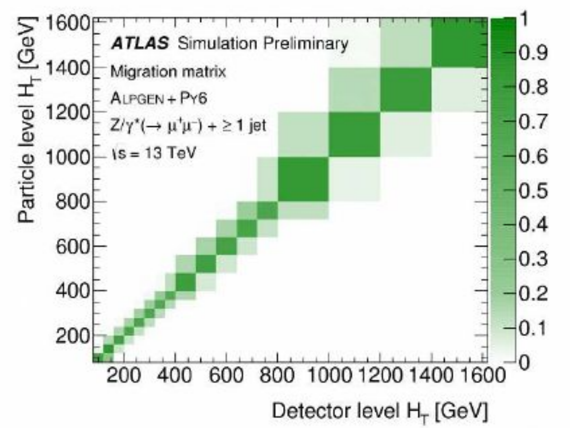


Iterative Unfolding with Bayesian method

Response matrix accounts for migrations using MC simulation:

$$M_{ij}=M(R_i|T_j)$$

Conditional probability that the effect R_i is produced by the cause T_j



How to extract *"prediction-unbiased"* probability using iterative Bayesian unfolding:

- Bayes theorem:

$$M(T_i|R_j) = M(R_i|T_j)\ P_0(T_j) \ /\ \mathrm{Sum}_l\ \ M(R_i|T_l)\ P_0(T_l)$$

- Particle level MC used as initial prior, $P_0(T_j)$, to determine a first estimate of the unfolded data distribution:

$$T_j = \mathrm{Sum}_i\ M(T_j\ |\ R_i\ )R_i$$

- In each further iteration the estimator of the unfolded distribution from previous iteration is used as a new prior

Other Z+b-jets differential distributions