PROCEEDINGS OF SCIENCE

$t\bar{t}b\bar{b}$ modeling for $t\bar{t}H$ analyses in ATLAS and CMS

Nihal Brahimi[a,*] **on behalf of the ATLAS and CMS collaborations**

[a] *TDLI-Tsung Dao Lee Institute, Shanghai Jiao Tong University*
Shanghai, China

E-mail: nihal.brahimi@cern.ch

An overview of the Monte Carlo modelling of top quark pair production in association with bottom quarks ($t\bar{t}b\bar{b}$), is presented. This process is a cornerstone for the $t\bar{t}H(b\bar{b})$ analyses performed by both ATLAS and CMS experiments at the Large Hadron Collider (LHC).

The Tenth Annual Conference on Large Hadron Collider Physics - LHCP2022
16-20 May 2022
online

*Speaker

https://pos.sissa.it/

PoS(LHCP2022)220

ATL-PHYS-PROC-2022-075
14/10/2022

The process of top quark pair production in association with bottom quarks ($t\bar{t}b\bar{b}$) is challenging to model theoretically due to its multi-scale nature (~ 350(10) GeV for $t\bar{t}(b\bar{b})$). Large theoretical uncertainties are thus associated with this process resulting from the significant differences among its Monte Carlo (MC) predictions. This is the limiting factor of sensitivity in measurements where $t\bar{t}b\bar{b}$ production is a dominant background such as the top quark pair production in association with a Higgs boson decaying to $b\bar{b}$, $t\bar{t}H(b\bar{b})$. These measurements were performed by both ATLAS [1] and CMS [2] Collaborations at the Large Hadron Collider (LHC) [3]. There are mainly two approaches to model the production of $t\bar{t}b\bar{b}$ at the LHC. One approach corresponds to predictions where the $t\bar{t}$ pair is described by the Matrix-Element (ME) part of the simulation, employing a five-flavour scheme (5FS) Parton Distribution Function (PDF). This includes simulations where the ME is calculated at the next-to-leading (NLO) in QCD and matched to a Parton Shower (PS) describing the additional b-quarks e.g through $g \to b\bar{b}$ splitting. This class of predictions considers the b-quarks as massless in the ME. The second approach makes use of $t\bar{t}b\bar{b}$ ME calculated at the NLO order with massive b-quarks, using a four-flavour scheme (4FS) PDF, matched to a PS. This type of predictions recently became the nominal approach for describing the $t\bar{t}b\bar{b}$ production at the LHC.

The latest $t\bar{t}H(b\bar{b})$ measurement performed by ATLAS [4] uses the full Run 2 p-p collision data amounting to an integrated luminosity of 139 fb^{-1}. The nominal prediction of the overwhelming $t\bar{t}b\bar{b}$ background in this analysis relies on a sample with 4FS $t\bar{t}b\bar{b}$ ME produced at NLO accuracy using the `Powheg Box Res` [5] generator matched to `Pythia8` [6] for the PS and hadronisation. The choice of the factorisation (μ_f) and renormalisation (μ_R) scales was set to $0.5 \times \sum_{i=t,\bar{t},b,\bar{b},j} m_T(i)$ [1] , where j stands for extra partons, and $\sqrt[4]{m_T(t).m_T(\bar{t}).m_T(b).m_T(\bar{b})}$, respectively, while the `Powheg` h_hdamp parameter was set to $0.5 \times \sum_{i=t,\bar{t},b,\bar{b}} m_T(i)$. Modeling uncertainties are assigned to the $t\bar{t}b\bar{b}$ prediction by comparing the nominal setup with different settings or samples. These uncertainties primarily account for variations in the amount of Initial State Radiation (ISR), including ME scale variations, and Final State Radiation (FSR), as well as alternative PS and hadronisation and NLO matching procedures. The latter two were derived based on comparisons among alternative $t\bar{t}$ 5FS predictions which are then propagated to the nominal prediction. The normalisation of $t\bar{t}b\bar{b}$ events was allowed to float freely in the signal extraction fit and was measured to be 1.28 ± 0.08, indicating that a larger $t\bar{t}b\bar{b}$ production cross section is favoured by data. The modeling of the $t\bar{t}b\bar{b}$ background was found to be the leading source of uncertainty impacting the $t\bar{t}H(b\bar{b})$ measurement sensitivity as can be seen in Table 1a.

A measurement of $t\bar{t}H(b\bar{b})$ was also carried out by CMS, where the most recent one uses collision data recorded in 2016 and 2017 amounting to about 78 fb^{-1} [7]. To model the $t\bar{t}b\bar{b}$ background, a 5FS $t\bar{t}+\text{jets}$ prediction was used. The latter was generated using `Powheg` [8] for the $t\bar{t}$ ME simulation at NLO accuracy, matched to `Pythia8` for the PS. The ME scales μ_F and μ_R were set to $m_T(t)$ while the `Powheg` h_damp parameter was set to $1.379 m_t$. The $t\bar{t}+\text{jets}$ events yield is normalised using the inclusive $t\bar{t}$ production cross-section of 831.76 pb, calculated at the next-to-next-to-leading order and next-to-next-to-leading-logarithm precision. Events where the $t\bar{t}$ pair is produced with additional b-jets form a sub-sample of the $t\bar{t}+\text{jets}$ sample, and is further divided

[1] m_T corresponds to the transverse mass of a particle i.e $m_T = \sqrt{m^2 + p_\text{T}^2}$ where m is the particle mass and p_T its transverse momentum

into 3 contributions based on the number of additional b-jets: tt+bb, tt+b and tt+2b corresponding respectively, to at least two additional b-jets, only one additional b-jet from a single B hadron and one additional b-jet from two B-hadrons close in direction. A normalisation uncertainty of 50% was assigned to each component. Additional systematic uncertainties affecting both the normalisation and shape of the $t\bar{t}b\bar{b}$ background were also considered e.g ME and PS scale variations, NLO matching (h_{damp} variation) and Underlying Event tune. Compared to the ATLAS background modeling, no dedicated comparison of different MC generators (2-point systematic) was used as a source of uncertainty. However, such comparison was performed in the context of testing the robustness of the signal extraction fit w.r.t the background modeling. The $t\bar{t}b\bar{b}$ uncertainties have a large impact on the $t\bar{t}H(b\bar{b})$ measurement sensitivity (cf. Table 1b), however, it is smaller compared to the ATLAS case where it is driven by the 2-point NLO matching systematic. The cross section of $t\bar{t}b\bar{b}$ is pulled by the fit towards higher values which indicates its underestimation in the MC prediction.

Table 1: Breakdown of the contributions to the uncertainties in the $t\bar{t}H(b\bar{b})$ signal strength $\Delta\mu$ for both (a) ATLAS [4] and (b) CMS [7]. The "tt+hf" row in the CMS table corresponds to the impact from the modeling uncertainties of the $t\bar{t}$ pair production with additional b- or c-jets, collectively referred to as heavy-flavour (hf).

(a)

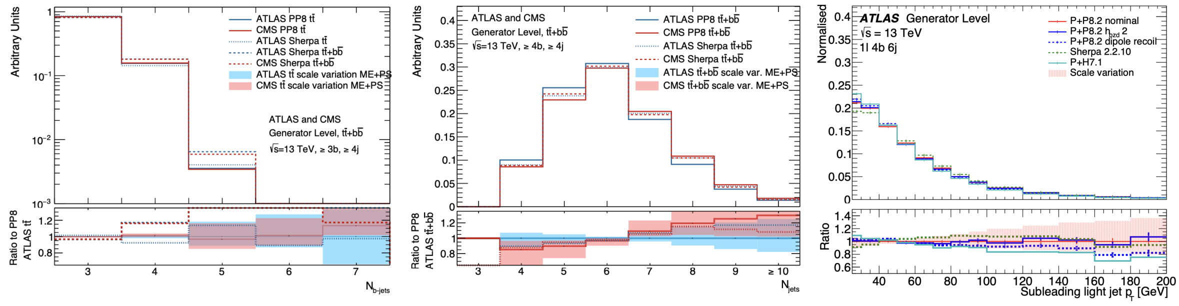
Uncertainty source	$\Delta\mu$	
Process modelling		
$t\bar{t}H$ modelling	+0.13	−0.05
$t\bar{t}+\geq 1b$ modelling		
$t\bar{t}+\geq 1b$ NLO matching	+0.21	−0.20
$t\bar{t}+\geq 1b$ fractions	+0.12	−0.12
$t\bar{t}+\geq 1b$ FSR	+0.10	−0.11
$t\bar{t}+\geq 1b$ PS & hadronisation	+0.09	−0.08
$t\bar{t}+\geq 1b$ p_T^{bb} shape	+0.04	−0.04
$t\bar{t}+\geq 1b$ ISR	+0.04	−0.04
$t\bar{t}+\geq 1c$ modelling	+0.03	−0.04
$t\bar{t}+$light modelling	+0.03	−0.03
tW modelling	+0.08	−0.07
Background-model statistical uncertainty	+0.04	−0.05
b-tagging efficiency and mis-tag rates		
b-tagging efficiency	+0.03	−0.02
c-mis-tag rates	+0.03	−0.03
l-mis-tag rates	+0.02	−0.02
Jet energy scale and resolution		
b-jet energy scale	+0.00	−0.01
Jet energy scale (flavour)	+0.01	−0.01
Jet energy scale (pile-up)	+0.00	−0.01
Jet energy scale (remaining)	+0.01	−0.01
Jet energy resolution	+0.02	−0.02
Luminosity	+0.01	−0.00
Other sources	+0.03	−0.03
Total systematic uncertainty	+0.30	−0.28
$t\bar{t}+\geq 1b$ normalisation	+0.04	−0.07
Total statistical uncertainty	+0.20	−0.20
Total uncertainty	+0.36	−0.34

(b)

Uncertainty source	$\Delta\hat{\mu}$
Total experimental	+0.15/−0.13
b tagging	+0.08/−0.07
jet energy scale and resolution	+0.05/−0.04
Total theory	+0.23/−0.19
signal	+0.15/−0.06
$t\bar{t}$+hf modelling	+0.14/−0.15
QCD background prediction	+0.10/−0.08
Size of simulated samples	+0.10/−0.10
Total systematic	+0.28/−0.25
Statistical	+0.15/−0.15
Total	+0.32/−0.29

Dedicated cross-section measurements of the $t\bar{t}b\bar{b}$ production, carried out by both ATLAS [9] and CMS [10, 11], yield 30%- 40% larger cross-section than the predictions. This is well in agreement with the $t\bar{t}H(b\bar{b})$ measurements presented above.

In view of a full Run 2 combination of $t\bar{t}H(b\bar{b})$ measurements between ATLAS and CMS, efforts were launched within the LHC Higgs Working Group since 2019 [12] to compare the modeling of the $t\bar{t}b\bar{b}$ background between the two experiments. These efforts aim to understand the possible differences in the setups used by each experiment and decide on a common strategy for the treatment of the background. The modeling comparisons for $t\bar{t}$ and $t\bar{t}b\bar{b}$ predictions, mainly used in the last rounds of $t\bar{t}H$ analyses, were performed in a $t\bar{t}H(b\bar{b})$-like phase space common between ATLAS and CMS (cf. Figure 1a and 1b). Very good agreement was observed overall in the $t\bar{t}$ predictions and corresponding scale variations, except for a few angular distributions. However, differences were observed for the `Powheg+Pythia8` $t\bar{t}b\bar{b}$ predictions and associated scale variations, where the CMS setup was found to have a larger variation resulting from using a factor 2 lower scale compared to ATLAS. A good agreement for alternative $t\bar{t}b\bar{b}$ predictions based on the Sherpa event generator [13] was seen which also matches the `Powheg+Pythia8` CMS setup.



(a) **(b)** **(c)**

Figure 1: Modeling comparisons between ATLAS and CMS MC setups for (a) the $t\bar{t}$ process in terms of the b-jets multiplicity and (b) $t\bar{t}b\bar{b}$ process in terms of the jets multiplicity [12] , (c) comparison of the optimised ATLAS nominal $t\bar{t}b\bar{b}$ prediction with different PS settings and Sherpa for the subleading light jet p_T [14].

Detailed studies of the $t\bar{t}b\bar{b}$ MC predictions were carried out by ATLAS [14] following the full Run 2 $t\bar{t}H(b\bar{b})$ and $t\bar{t}b\bar{b}$ measurements, to provide recommendations for the nominal prediction to be adopted in the legacy $t\bar{t}H$ analysis. Different modeling effects were also explored as inputs to help define common conventions for the determination of the uncertainties. Parameter settings of the 4FS `Powheg+Pythia8` prediction were revisited and optimised based on theoretical arguments and on comparisons to data. Figure 1c shows comparisons between the optimised nominal prediction and alternative 4FS setups. The resulting shape differences were found to be small for observables characterising b-jets described by the MEs ($\sim 10\%$), however, they increase to $\sim 20\%$ for observables related to additional light radiation from the PS.

The modeling of the $t\bar{t}b\bar{b}$ production and its associated uncertainties is the key factor driving the sensitivity of the $t\bar{t}H(b\bar{b})$ measurements. Despite significant improvements in the theoretical knowledge of this process, measurements of its differential distributions with enhanced precision are still needed to constrain further the available MC models.

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.

[2] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[3] L. Evans and P. Bryant, eds., *LHC Machine*, *JINST* **3** (2008) S08001.

[4] ATLAS Collaboration, *Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **2022** (2022) .

[5] T. Ježo, J.M. Lindert, N. Moretti and S. Pozzorini, *New NLOPS predictions for $t\bar{t}$ +b-jet production at the LHC*, *EPJC* **78** (2018) .

[6] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159.

[7] CMS Collaboration, *Measurement of $t\bar{t}H$ production in the $H\rightarrow b\bar{b}$ decay channel in* $41.5\,\mathrm{fb}^{-1}$ *of proton-proton collision data at $\sqrt{s}$=13 TeV*, *CMS-PAS-HIG-18-030 (2019)* .

[8] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower monte carlo programs: the POWHEG BOX*, *JHEP* **2010** (2010) .

[9] ATLAS Collaboration, *Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton-proton collisions at $\sqrt{s}$=13 TeV with the ATLAS detector*, *JHEP* **2019** (2019) .

[10] CMS Collaboration, *Measurement of the cross section for $t\bar{t}$ production with additional jets and b jets in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **2020** (2020) .

[11] CMS Collaboration, *Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s}$=13 TeV*, *Phys. Lett. B* **803** (2020) .

[12] LHC Higgs cross-section working group. `https://indico.cern.ch/event/960331/timetable/#8-input-from-lhc-higgs-xs-wg-o`, 2020.

[13] SHERPA Collaboration, *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) .

[14] ATLAS Collaboration, *Studies of monte carlo predictions for the $t\bar{t}b\bar{b}$ process*, *ATL-PHYS-PUB-2022-006 (2022)* .