Review of data-driven methods used to control the normalization of the top quark background contribution in the H –> WW(*) –> $\ell\nu\ell\nu$ analyses at the LHC

Y. Li, X. Ruan, L. Yuan, Z. Zhang

LAL 13-373

Review of data-driven methods used to control the normalization of the top quark background contribution in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analyses at the LHC

Yichen Li,[1,2] Xifeng Ruan,[3] Li Yuan,[4] and Zhiqing Zhang[2,*]

[1] *Nanjing University, Nanjing, China*

[2] *Laboratoire de l'Accélérateur Linéaire, Université Paris-Sud 11, IN2P3/CNRS, Orsay, France*

[3] *School of Physics, University of the Witwatersrand, Johannesburg, South Africa*

[4] *Graduate School of Science, Kobe University, Kobe, Japan*

(Dated: November 9, 2013)

A few data-driven methods used in deriving the normalization of the dominant top quark background contribution in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analyses by the ATLAS and CMS experiments at the LHC are reviewed and compared. Additional information, justification or modification to some of the methods is provided. These methods have also been or can be applied to other analyses such as cross section measurements of the Standard Model WW process and searches for new physics in channels with similar final states.

I. INTRODUCTION

The top quark events from both $t\bar{t}$ and single top quark processes are often one of the dominant backgrounds in the searches for new particles and measurements of Standard Model (SM) cross sections. One good example is the measurement of W^+W^- production and search for the Higgs boson in the same final state decaying to $\ell\nu\ell\nu$ with $\ell = e, \mu$ and τ with its subsequent leptonic decays. In hunting for the Higgs boson, due to the undetected neutrinos in the leptonic WW channels, the transverse mass resolution of the Higgs boson is very limited and no narrow mass peak is expected. It is therefore extremely important to derive the normalization of various background processes from data instead of using the predictions from Monte Carlo (MC) simulation. Several (semi)data-driven methods have been proposed and used. However, the choice of the method is often somewhat arbitrary. This may not make a big difference as far as the search is concerned. Given the recent discovery of the Higgs boson, the situation is different as we are now aiming for a precise measurement of its property in order to check if it is the SM Higgs boson or not. The same is true for the precision measurement of the SM WW cross section. This motivates us to review these methods and discuss the advantage or disadvantage of each method and give recommendation for an optimum choice.

The article is organized as follows. In Sec. II, various methods used in previous publications are introduced. In Sec. III, the results of these methods are discussed and compared. Potential extension for application to other channels or analyses is also discussed.

II. THE METHODS

The analyses of ATLAS and CMS in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ are classified in three different subchannels with zero, one and two or more jets in the final state above a typical transverse momentum ($p_T^{\rm jet}$) threshold of around 25 GeV and within a pseudo-rapidity ($\eta^{\rm jet}$) of about 4.5. Jets are built using the anti-k_T clustering algorithm [1] with a distance parameter of typically $R = 0.4$ (ATLAS) and 0.5 (CMS). The choice is made to optimize the Higgs signal sensitivity in each channel for different background contributions and signal over background ratios. The first two

*Electronic address: zhang@lal.in2p3.fr

channels are sensitive to the gluon-gluon fusion (ggF) production mode of the Higgs boson whereas the latter one to the vector boson fusion (VBF) production mode, whose cross section is about ten times smaller than that of ggF. For convenience, we name them as $0j$, $1j$ and $\geq 2j$ channels in the following with the corresponding numbers of events N^{0j}, N^{1j} and $N^{\geq 2j}$.

A. Jet veto survival probability method

As far as the $0j$ channel is concerned, both the top quark background and other background contribution can be greatly suppressed by a jet veto requirement. In ATLAS, the baseline method used [2, 3] to determine the normalization of the top quark background is the so-called Jet Veto Survival Probability (JVSP) method [4, 5]:

$$N^{\rm Exp}_{\rm top,0j} = (N^{\rm Data}_{\rm all} - N_{\text{non-top}}) \times P_2^{\rm Exp} \tag{1}$$

where $N^{\rm Data}_{\rm all}$ and $N_{\text{non-top}}$ are, respectively, the number of all selected data events and the corresponding non-top background events in an inclusive event sample selected just before the jet veto requirement[1], and $P_2^{\rm Exp}$ is a data-driven estimate of the full jet veto survival probability, standing for the fraction of top events in the zero jet bin over all top events. This expression has an analogy to the Monte Carlo expectation of $N^{\rm MC}_{\rm top,0j} = N^{\rm MC}_{\text{all top}} \times P_2^{\rm MC}$.

The data-driven $P_2^{\rm Exp}$ is derived using

$$P_2^{\rm Exp} \;=\; P_2^{\rm MC} \times \left(\frac{P_1^{\rm Data}}{P_1^{\rm MC}}\right)^2 \tag{2}$$

$$\simeq\; P_2^{\rm MC} \times \left(\frac{P_1^{\text{Data, btag}}}{P_1^{\text{MC, btag}}}\right)^2 \tag{3}$$

where $P_1^{\rm Data(MC)}$ is a single jet veto survival probability in data (MC) and $P_1^{\text{Data (MC), btag}}$ is the corresponding jet veto survival probability determined from a control sample in which there is at least one tagged b-jet selected in a certain phase space of ($P_T^{\text{b-jet}}$, $\eta^{\text{b-jet}}$) which may be different from the previous jet phase space ($P_T^{\rm jet}$, $\eta^{\rm jet}$).

Equation (2) is constructed based on an expected relation

$$P_2 = P_1^2\,. \tag{4}$$

This relation is derived following the consideration that in a top quark event one b-jet may be untagged (P_1) or tagged ($1-P_1$) independent of the other b-jet such that over a sample of N events, a subsample $N_0 = P_1^2 N$ has both b-jets being untagged and another subsample $N_1 = 2P_1(1-P_1)N$ has one of the b-jets untagged and the other tagged. The independence consideration can be checked by comparing the P_1 derived from each of these subsamples. Using a $t\bar{t}$ MC sample generated with the MC@NLO package [6], the two probabilities are found to agree well within 2%. The same numerical agreement is obtained when single top events from tW process [6] is included. In reality there are radiative jets from initial and final state radiations in addition to the b-jets in a top event. In the presence of radiative jets, the physical meaning for P_1 and P_2 remains the same but the relation (4) may be modified as $P_2 = CP_1^2$ with C representing the probability that non-b-jets fall under the P_T threshold. This modification does not affect the validity of Eqs. (2) and (3). Further discussion on the effect of the radiation jets can be found in Appendix A.

Equation (3) has the advantage over Eq.(2) in that a highly pure top quark sample can be selected and $P_1^{\rm Btag}$ can be determined both in data and in MC. In fact, $P_1^{\rm Btag}$ is simply the fraction of events in which no probing jet is reconstructed in addition to the tagged b-jet ($N^{\rm Btag}_{\text{0 prob-jet}}$) over the total number of tagged events ($N^{\rm Btag}_{\rm all}$) after subtracting a small non-top background contribution. For the probing jet reconstruction, it is preferable that the same $P_T^{\rm jet}$ threshold and $\eta^{\rm jet}$ acceptance are used so that jet related experimental systematic uncertainties largely cancel in the ratio $P_2^{\rm MC}/\left(P_1^{\text{MC, btag}}\right)^2$. It is this ratio term which contributes to the experimental and theoretical systematic

[1] If there are additional cuts, they may be moved forward provided there is sufficient data statistics left for the method to apply.

uncertainties of the top quark background estimate. In the application of the method to the ATLAS analyses, the probing jet is defined by requiring a minimum distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 1$ between the probing jet and the tagged b-jet in order to avoid selecting a radiative jet from the b-jet as a probing jet. But it has been checked that the choice is not critical and the result is stable within the statistical uncertainty for a large variation on ΔR between 0.4 and 1.2.

In using Eq.(3) instead of Eq.(2) there is a little price to pay, namely the systematic uncertainty cancellation is slightly worse in $P_2^{\rm MC}/\left(P_1^{\rm MC,\ btag}\right)^2$ than in $P_2^{\rm MC}/\left(P_1^{\rm MC}\right)^2$ (see Appendix B for more detail).

The method was first applied in the cross section measurement of the SM WW process based on the 7 TeV pp collision data with an integrated luminosity of $1.02\,\text{fb}^{-1}$ [2], the quoted relative statistical and systematic uncertainties were 15% and 20%, respectively, for the dominant different flavor channel $e\mu + \mu e$. The systematic uncertainties were dominated by a conservative estimate of the theoretical uncertainties.

For the application to the search of the SM Higgs boson in the WW channel based on the 7 TeV pp collision data corresponding to an integrated luminosity of $4.7\,\text{fb}^{-1}$ [3], the same conservative systematic uncertainties were kept but the statistical uncertainty has been substantially reduced to 6.7%. The similar uncertainty was quoted in the Higgs discovery paper from ATLAS based on $4.7\,\text{fb}^{-1}$ of 7 TeV and $5.3\,\text{fb}^{-1}$ of 8 TeV data [7] and it was reevaluated with a total uncertainty of 13% in the mass and coupling measurement paper from ATLAS based on the full "Run-1" data [8].

B. Template method

In [9], a template method is used to determine the normalization of the top background contribution in the SM WW analysis. In this method, the top estimation is performed in an extended signal region (ESR) just before applying the jet veto requirement. In addition, a control region (CR) is defined as a subset of the ESR, which contains events having at least one b-tagged jet at lower transverse momentum with $20\,\text{GeV} < P_T < 25\,\text{GeV}$. The ESR is similar to the inclusive sample of the JVSP method except that it contains no tagged b-jets above the $P_T^{\rm jet}$ threshold (25 GeV). Also the CR differs in the two methods. The jet multiplicity distribution for top quark events in the ESR, $T_{\rm data}^{\rm ESR}$, is estimated from the jet multiplicity distribution in the CR, $T_{\rm data}^{\rm CR}$. In a first step, the non-top background distribution $T_{\rm MC,\ non\text{-}top}^{\rm CR}$ in the CR is estimated with simulation, scaled by a normalization factor f_n' and then subtracted from the measured $T_{\rm data}^{\rm CR}$ distribution. Subsequently, the resulting distribution is extrapolated bin-by-bin from the CR to the ESR via the MC prediction of the ratio $T_{{\rm MC},i}^{\rm ESR}/T_{{\rm MC},i}^{\rm CR}$ for each jet multiplicity bin i. The method can be summarized by the following equation for each jet multiplicity bin:

$$T_{\rm data}^{\rm ESR} = \frac{T_{\rm MC}^{\rm ESR}}{T_{\rm MC}^{\rm CR}}\left(T_{\rm data}^{\rm CR} - f_n' \times T_{\rm MC,\ non\text{-}top}^{\rm CR}\right), \tag{5}$$

where each symbol T represents a full jet multiplicity distribution. The normalization scale factor f_n' for the non-top background contributions in the CR is determined from events in the ESR by fitting the jet multiplicity distribution observed in data with the templates constructed from the data in the CR for top quark contributions and from simulation for non-top contributions. If the template fit for f_n' is not performed, the method becomes the one to be introduced in Sec. II C. In a final step, the number of top background events in the signal region is estimated using the number of top events in the ESR observed in data scaled by the ratio of top events in the signal region to the number in the ESR in the MC simulation for the $0j$ bin.

The normalization scale factor f_n' may be viewed as an effective normalization for the various non-top background processes. It depends thus on the composition of the background processes. In the method, it is implicitly assumed that the composition is the same between the ESR and the CR. The potential difference and the corresponding systematic uncertainty was neglected. The other dominant experimental and theoretical uncertainties arise from the ratio term $T_{\rm MC}^{\rm ESR}/T_{\rm MC}^{\rm CR}$.

For the 7 TeV pp collision data corresponding to an integrated luminosity of $4.6\,\text{fb}^{-1}$, the value of f_n' was found to be 1.07 ± 0.03 [9]. For the dominant $e\mu + \mu e$ channel, the quoted relative statistical and systematic uncertainties were 26% and 15%, respectively. The large statistical uncertainty is due to the limited number of data events observed in the CR. The systematic uncertainties were dominated by the b-tagging uncertainty.

C. Extrapolation method from control region

In some of the analyses, a CR is defined from which the normalization of the top background is determined. The top background in the SR is derived assuming that the normalization in the SR is same as in the CR:

$$N_{\rm top}^{\rm SR} = N_{\rm top}^{\rm SR,MC} \times \frac{N_{\rm data}^{\rm CR} - N_{\rm non\text{-}top}^{\rm CR}}{N_{\rm top}^{\rm CR,MC}} \,. \tag{6}$$

For the $1j$ bin analysis of the Higgs search in the WW channel from ATLAS [3, 7, 8], the CR corresponds to an event sample with only one reconstructed jet which is b-tagged. For the $\geq 2j$ bin analysis, the CR may be defined with either at least one tagged b-jet [3, 7] or only one tagged b-jet [8].

The advantage of the method is that it is simple. The similar method has also been applied for determining the normalization of the other backgrounds, such as the SM WW background contribution in the search of $H \to WW^{(*)}$. However the experimental and theoretical uncertainties arising from the ratio term $N_{\rm top}^{\rm SR,MC}/N_{\rm top}^{\rm CR,MC}$ are in general large. For instance, the total quoted uncertainties for the $1j$ and $\geq 2j$ channels in [8] were about 30% and 40%, respectively.

D. In-situ b-tagging efficiency based method

In the analyses of CMS [10, 11], the normalization of the top quark background is also estimated from data by counting the number of top-tagged ($N_{\rm tagged}$) events and applying the corresponding top-tagging efficiency. The top-tagging efficiency ($\epsilon_{\rm top\ tagged}$) is measured with a control sample dominated by $t\bar{t}$ and tW events, which is selected by requiring a b-tagged jet. The residual number of top events ($N_{\rm not\ tagged}$) in the signal region is given by:

$$N_{\rm not\ tagged} = \frac{N_{\rm tagged}}{\epsilon_{\rm top\ tagged}} \times (1 - \epsilon_{\rm top\ tagged}) \,. \tag{7}$$

In general, the efficiency appearing in the brackets can be different from the one in the denominator as we will see in Sec. III.

For the jet category definition, CMS has used a slightly different threshold $P_T > 30\,{\rm GeV}$. No details were given in the publications on how the control sample was exactly defined for each jet bin, though some information was provided for the $0j$ and $1j$ channels in thesis [12]. The choice of the control sample depends on the number of jets per event and is usually not unique (see Sec. III for further discussions).

In the search for the standard model Higgs boson decaying to W^+W^- in the fully leptonic final state in pp collisions at $\sqrt{s} = 7\,{\rm TeV}$ corresponding to an integrated luminosity of $4.6\,{\rm fb}^{-1}$, the quoted uncertainty was about 25% in the $0j$ category and about 10% for the other categories [10]. In the VBF analysis based on the full "Run-1" data [11], the quoted uncertainties were about 27% and 18% at $\sqrt{s} = 7\,{\rm TeV}$ and $8\,{\rm TeV}$, respectively. The main uncertainty comes from the statistical uncertainty in the control sample and from the systematic uncertainties related to the measurement of $\epsilon_{\rm top\ tagged}$.

III. COMPARISON AND DISCUSSIONS

In all the analyses, one normalizes the number of data-driven top quark background events derived from the various methods over the corresponding MC one to get a normalization factor (NF):

$${\rm NF_{top}} \equiv \frac{N_{\rm top,\ data}^{\rm Exp}}{N_{\rm top,\ MC}} \,, \tag{8}$$

where $N_{\rm top,\ MC}$ can be either the number of top MC events derived by applying the same formula as for data ($N_{\rm top,\ MC}^{\rm Exp}$) or the number of top MC events counted directly in the corresponding signal region ($N_{\rm top,\ MC}^{\rm Count}$). In all the methods

except for the in-situ b-tagging efficiency based method, the two numbers are identical by definition whereas in the latter method, they can be different in general and the difference provides a measure of the non-closure of the method.

Among the three methods used for the $0j$ channel, the result obtained from the JVSP method has the smallest statistical uncertainty of below 3% for the full "Run-1" data whereas the corresponding statistical uncertainty from the other two methods is about a factor of $2-3$ larger due to the small b-tagged data control sample at low P_T. The experimental systematic uncertainty of about 6% from the JVSP is also moderate and smaller than that of the template method because of a better cancellation in the probability ratio $P_2^{\rm MC}/\left(P_1^{\rm MC,\ btag}\right)^2$ than in the event ratio $T_{\rm MC}^{\rm ESR}/T_{\rm MC}^{\rm CR}$. The theoretical uncertainties quoted for the JVSP method of about 10% were due to either the limited MC statistics or LO MC generators used in the uncertainty evaluation. The uncertainties have been reevaluated with higher statistical NLO MC samples with a total theoretical uncertainty of about 4% which includes model uncertainties (MC@NLO vs. Powheg [13] and Pythia [14] vs. Herwig [15]), uncertainties of renormalization and factorization scale variations, of parton distribution functions, of relative variation of the single top contribution by $\pm 30\%$ and of the $t\bar{t}$ and Wt interference effect (evaluated by comparing two different schemes: diagram removal vs. diagram subtraction [16]). Therefore the final total uncertainty is about 8% for the JVSP method with the "Run-1" data. In the in-situ b-tagging efficiency based method, the top quark event tagging efficiency is determined directly from data, therefore one of the main systematic uncertainties arises from a potential bias in the efficiency determination with a control sample. This bias is evaluated with a MC closure test. The other possible important systematic source is associated to the non-top background subtraction which depends of course on the b-tagging algorithms used. But it is generally expected that the rate of the fake b-tagging and the corresponding uncertainty at low P_T may be relatively important.

For the $1j$ and $\geq 2j$ channels, the statistical uncertainty of the extrapolation method is in general smaller than that of the in-situ b-tagging efficiency based method. The latter method performs nevertheless better in terms of the experimental systematic uncertainties. However since the control sample used to determine the top event tagging efficiency may not be unique, some discussion is needed to see if one choice of control sample is better than the other and has less bias not only in terms of the global efficiency but eventually in terms of different kinematic distributions.

Indeed, for the $1j$ analysis, the signal region corresponds to an event sample which has one reconstructed jet which is not b-tagged. The top background contribution in data can be estimated by using Eq.(7) where $N_{\rm tagged}$ is the number of tagged events in one-jet data sample. The corresponding top-tagging efficiency is determined from a control sample with two reconstructed jets. The sub-leading P_T jet is b-tagged and the leading P_T jet is used to measure the b-tagging efficiency according to thesis [12]. This choice of the control sample may introduce a kinematic bias. This is shown in Fig. 1 (left) comparing the b-jet P_T spectrum in the one-jet sample with that of the leading and sub-leading b-jets from the two-jet sample. A random b-tagging could reduce the bias in the P_T distribution as

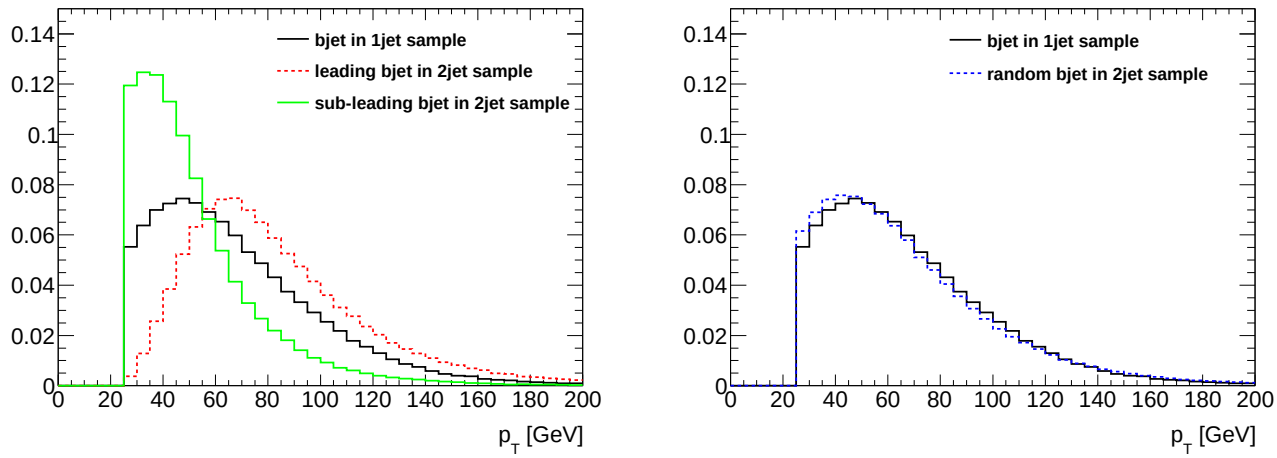

FIG. 1: Left: comparing the shape of the P_T jet spectrum in the b-tagged $1j$ sample with that of leading and sub-leading b-jets in the $2j$ sample. Right: comparing the shape of the P_T jet spectrum in the b-tagged $1j$ sample with that of a randomly tagged b-jet in the $2j$ sample.

shown in Fig. 1 (right). In addition, the b-jet purity and thus the b-tagging efficiency could be different between the one-jet and two-jet samples. To correct for these potential biases, one could introduce a MC based correction factor

$\epsilon_{\rm MC}/\epsilon_{\rm MC\ tagged}$ to Eq.(7) for the $1j$ analysis to read:

$$N^{\rm SR\ 1j}_{\rm not\ tagged} = \frac{N_{\rm tagged}}{\epsilon_{\rm top\ tagged} \times \frac{\epsilon_{\rm MC}}{\epsilon_{\rm MC\ tagged}}} \times \left(1 - \epsilon_{\rm top\ tagged} \times \frac{\epsilon_{\rm MC}}{\epsilon_{\rm MC\ tagged}}\right) . \tag{9}$$

In this way, the MC closure is guaranteed by construction but one pays a price by introducing additional experimental and theoretical uncertainties through the MC correction factor. Nevertheless, the uncertainties are expected to be small or moderate as some cancellation is expected in the ratio. The formulae for various efficiencies in Eq.(9) are explicitly given in Appendix C. There is another practical advantage in introducing the MC correction factor, without it, the top tagging efficiency $\epsilon_{\rm top\ tagged}$ has to be determined with a sample selected with the same selection cuts as those used for $N_{\rm tagged}$. Both of these, though correlated, contribute to the statistical uncertainty of the top background evaluation. With the MC correction, one could use a different and statistically much larger sample by removing some of the selection cuts to determine the top tagging efficiency such that its contribution to the statistical uncertainty becomes sub-leading or negligible with respect to that of $N_{\rm tagged}$.

For the $\geq 2j$ VBF analysis, a similar equation can be defined:

$$N^{\rm SR\ \geq 2j}_{\rm not\ tagged} = \frac{N_{\rm tagged}}{\epsilon_{\rm top\ tagged} \times \frac{\epsilon_{\rm MC}}{\epsilon_{\rm MC\ tagged}}} \times \left(1 - \epsilon'_{\rm top\ tagged} \times \frac{\epsilon'_{\rm MC}}{\epsilon'_{\rm MC\ tagged}}\right) . \tag{10}$$

Here we have used different notations for the top tagging efficiencies appearing outside and inside of the brackets. The one outside corrects for the inefficiency of the top tagging in order to get the total number of top events. This efficiency depends on the tagging sample used. In the $\geq 2j$ analysis, one has the choice of selecting a control sample with only one tagged b-jet or with at least one tagged b-jet. Once we have the full top sample, we know that for the inclusive $\geq 2j$ VBF analysis, we have two types of top events to veto: one type with one tagged b-jet and the other type with two tagged b-jets. Therefore this efficiency is independent of the choice of the previous tagging sample. This complication is not there in the $1j$ analysis since there is only one b-jet to tag and to veto [2]. Explicit formulae are given in Appendix D.

For the $0j$ channel, the formula is similar:

$$N^{\rm SR\ 0j}_{\rm not\ tagged} = \frac{N_{\rm tagged}}{\epsilon_{\rm top\ tagged} \times \frac{\epsilon_{\rm MC}}{\epsilon_{\rm MC\ tagged}}} \times \left(1 - \epsilon'_{\rm top\ tagged} \times \frac{\epsilon'_{\rm MC}}{\epsilon'_{\rm MC\ tagged}}\right) . \tag{11}$$

The only difference is that the tagged sample is selected from events having at least one low P_T b-jet below the threshold of $P_T^{\rm jet}$ in the $0j$ channel.

These methods can be applied to all those analyses studying the similar final state namely dilepton with opposite charge and with the top background being one of the dominant backgrounds. A few non exhaustive examples are listed below:

- Search for direct slepton and gaugino production in final states with two leptons and missing transverse momentum with the ATLAS detector in pp collision at $\sqrt{s} = 7\,\text{TeV}$ [17].
- Exclusive search for supersymmetry with same-flavor dilepton final states with the ATLAS detector [18].
- Search for heavy neutrinos and right-handed W bosons in events with two leptons and jets in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector [19].
- Search for heavy neutrinos and W_R bosons with right-handed couplings in a left-right symmetric model in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ [20].
- Search for narrow resonances in dilepton mass spectra in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ [21].
- Search for new phenomena in the $WW \to \ell\nu\ell\nu$ final state in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector [22].

---

[2] This is true only when the same jet reconstruction and b-jet veto P_T threshold is used. Otherwise, one may have more than one b-jet to veto.

In most of these analyses the top background was estimated from MC simulation and therefore can be improved by applying one of the data-driven methods discussed here.

IV. SUMMARY

In the search for the Higgs boson in the $WW^{(*)} \to \ell\nu\ell\nu$ channel at the LHC, the top background has been a dominant background source and a number of data-driven methods have been developed to determine its normalization from data. These methods are reviewed and compared with more details shown here than those given in the ATLAS and CMS publications. We have also extended the in-situ b-tagging efficiency based method by introducing a MC correction factor so that potential kinematic bias can be taken into account. The uncertainty of the top estimation varies from one method to others and may differ by several factors in terms of the total precision for a given data sample. It is recommended to try more than one method for comparison. An optimum choice of the method is important in view of precision measurements with larger data samples to come. These methods can also be applied to other analyses studying the same final states both for the cross section measurements and for search for new resonances or new physics.

Acknowledgments

The authors are grateful to colleagues in ATLAS and CMS in particular A.J. Armbruster, B. Di Micco, D. Froidevaux, G. Gomez Ceballos, C. Hays, X. Janssen, J. Jovicevic, B. Malaescu, C. Mills, G. Salamanna and P. Savard for their contribution, clarification and stimulating discussion. Li thanks S. Chen for support and encouragement. Ruan and Zhang thank B. Mellado for early collaboration.

Appendix A: More on b-jets and radiative jets in the JVSP method

Doing matching between a truth jet and a b parton with a distance of $\Delta R < 0.4$ to define a b-jet, the P_T spectrum of b-jets within a pseudo-rapidity acceptance of 2.5 is shown in Fig. 2(a) in comparison with that of the non-b-jets within an acceptance of 4.5. The non-b-jets include both unmatched jets and those b-jets beyond the b-tagging acceptance of 2.5. The number of jets in the figure is normalized to the total number of $t\bar{t}$ events. The corresponding number of

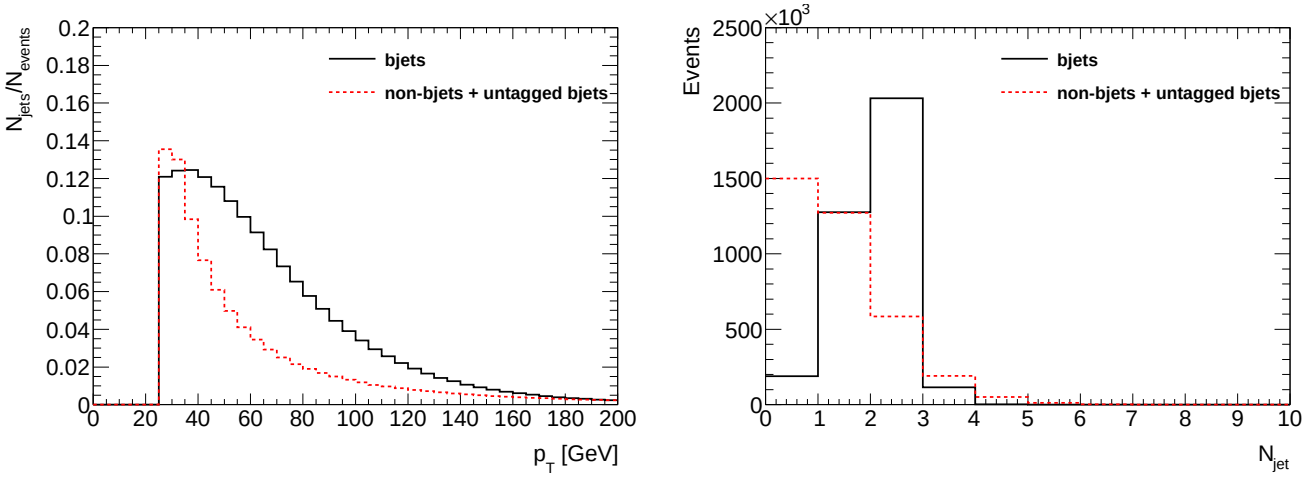

FIG. 2: Left: the transverse momentum spectrum of b-jets in comparison with that of non-b-jets for $P_T > 25\,\text{GeV}$ in a $t\bar{t}$ event sample generated with MC@NLO. The pseudo-rapidity acceptance used for the b-jets and non-b-jets are 2.5 and 4.5, respectively. Right: the number of b-jets in comparison with that of non-b-jets per event.

jets for $P_T > 25\,\text{GeV}$ per event is shown in Fig. 2(b). Close to the P_T threshold of 25 GeV, there are slightly more

non-b-jets than b-jets but overall b-jets dominate in the sample. The fraction of events in the zero jet bin in Fig. 2(b) defines the jet veto survival probability. From the shapes of the $N_{\rm jet}$ distribution, it is clear that the b-jet veto survival probability is more sensitive to all those systematic variations such as jet energy scale and resolution uncertainties which result in jet bin migration, whereas the non-b-jet veto survival probability varies much more slowly for the same systematic uncertainties. In other words, radiative jets, although affect the jet multiplicity distribution, have little effect on the variation of P_2.

Appendix B: Relation between $P_1^{\rm Btag}$ and P_1

In addition to the jet veto survival probability P_1, let us define ϵ as the b-jet tagging (including acceptance) efficiency. Therefore the sample with at least one tagged b-jet $N_{\rm all}^{\rm Btag}$ can be written as

$$N_{\rm all}^{\rm Btag} = 2(1-P_1)\epsilon \times P_1 + 2(1-P_1)\epsilon \times (1-P_1)(1-\epsilon) + (1-P_1)\epsilon \times (1-P_1)\epsilon \tag{B1}$$

where the first term corresponds to the case where only one jet is present and b-tagged (i.e. $(1-P_1)\epsilon$) and the other jet is not present (P_1) and the factor of 2 represents the two possible combinations, the second term corresponds to the case where one of the jets is tagged ($(1-P_1)\epsilon$) and the other jet is present but untagged ($(1-P_1)(1-\epsilon)$), and finally the last term corresponds to the case where both jets are tagged. The first term also means no probing jet can be reconstructed, thus $P_1^{\rm Btag}$ can be written as

$$P_1^{\rm Btag} \equiv \frac{N_{0\ \text{prob-jet}}^{\rm Btag}}{N_{\rm all}^{\rm Btag}} \tag{B2}$$

$$= \frac{2(1-P_1)\epsilon \times P_1}{N_{\rm all}^{\rm Btag}} \tag{B3}$$

$$= \frac{P_1}{1-\frac{(1-P_1)\epsilon}{2}}\,. \tag{B4}$$

Equation (B4) shows that $P_1^{\rm Btag}$ is proportional to P_1 with, in the denominator, a correction which may have a typical value of around 0.25, depending on the selection cuts and probing jet definition. This explains why the systematic uncertainty cancellation is expected in $P_2/\left(P_1^{\rm Btag}\right)^2$ but is not as good as in $P_2/\left(P_1\right)^2$.

Appendix C: Explicit formula for b-jet tagging efficiency determination in the $1j$ analysis

In the $1j$ analysis, the $2j$ control sample with at least one tagged b is used to determine the top tagging efficiency $\epsilon_{\text{top tagged}}$. Within the sample, the numbers of events with one and two tagged b-jets, N_1^{2j} and N_2^{2j}, can be expressed in terms of the b-quark tagging efficiency $\epsilon_{\rm btag}$ and the total number of $2j$ events N^{2j} as:

$$N_1^{2j} = 2\epsilon_{\rm btag}\left(1-\epsilon_{\rm btag}\right)N^{2j}\,, \tag{C1}$$

$$N_2^{2j} = \left(\epsilon_{\rm btag}\right)^2 N^{2j}\,. \tag{C2}$$

From these relations, one obtains

$$\epsilon_{\text{top tagged}} \equiv \epsilon_{\rm btag} = \frac{2N_2^{2j}}{N_1^{2j}+2N_2^{2j}}\,. \tag{C3}$$

This formula applies to both data and MC and the equivalence sign reflects the fact that in the $1j$ channel, the top event tagging efficiency and b-quark tagging efficiency are identical. The MC $1j$ tagging efficiency $\epsilon_{\rm MC}$ in Eq.(9) is simply

$$\epsilon_{\rm MC} = \frac{N_{1,\ \rm MC}^{1j}}{N_{0,\ \rm MC}^{1j}+N_{1,\ \rm MC}^{1j}}\,, \tag{C4}$$

with $N^{1j}_{0,\,\mathrm{MC}}$ and $N^{1j}_{1,\,\mathrm{MC}}$ being the number of events with zero and one tagged b-quark in the $1j$ MC sample, respectively. Therefore one has

$$\frac{\epsilon_{\mathrm{MC}}}{\epsilon_{\mathrm{MC\ tagged}}} = \frac{N^{1j}_1(N^{2j}_{1,\,\mathrm{MC}} + 2N^{2j}_{2,\,\mathrm{MC}})}{2N^{2j}_{2,\,\mathrm{MC}}(N^{1j}_{0,\,\mathrm{MC}} + N^{1j}_{1,\,\mathrm{MC}})}\,. \tag{C5}$$

Note in Eqs.(C1) and (C2), it is implicitly assumed that the two b-jets are uncorrelated. If there is any correlation, this is taken into account in the MC correction factor.

Appendix D: Different choices of control samples in the $\geq 2j$ analysis

Given the inclusive nature of the $\geq 2j$ VBF analysis, there are a few choices in selecting event samples for tagging and for determining the top tagging efficiency. One example is to select events with at least one tagged b-jet. In this case, one has (a super index $\geq 2j$ is implicitly implied in all variables):

$$N_{\mathrm{tagged}} = N_1 + N_2\,, \tag{D1}$$

$$\epsilon_{\mathrm{top\ tagged}} \equiv \epsilon'_{\mathrm{top\ tagged}} = 2\epsilon_{\mathrm{btag}}(1-\epsilon_{\mathrm{btag}}) + (\epsilon_{\mathrm{btag}})^2\,, \tag{D2}$$

$$\epsilon_{\mathrm{MC}} \equiv \epsilon'_{\mathrm{MC}} = \frac{N_{1,\mathrm{MC}} + N_{2,\mathrm{MC}}}{N_{0,\,\mathrm{MC}} + N_{1,\,\mathrm{MC}} + N_{2,\,\mathrm{MC}}}\,. \tag{D3}$$

Equation (D2) shows the connection between the top event tagging efficiency $\epsilon_{\mathrm{top\ tagged}}$ with the per b-quark jet tagging efficiency ϵ_{btag}. The first and second terms on the right hand side of Eq.(D2) correspond to event samples with one and two tagged b-jets, respectively. Using Eq.(C3) in Eq.(D2), one obtains:

$$\frac{\epsilon_{\mathrm{MC}}}{\epsilon_{\mathrm{MC\ tagged}}} = \frac{(N_{1,\,\mathrm{MC}} + 2N_{2,\,\mathrm{MC}})^2}{4N_{2,\,\mathrm{MC}}(N_{0,\,\mathrm{MC}} + N_{1,\,\mathrm{MC}} + N_{2,\,\mathrm{MC}})}\,. \tag{D4}$$

Another choice is to select events with only one tagged b-jet. In this case one has

$$N_{\mathrm{tagged}} = N_1\,, \tag{D5}$$

$$\epsilon_{\mathrm{top\ tagged}} = 2\epsilon_{\mathrm{btag}}(1-\epsilon_{\mathrm{btag}})\,, \tag{D6}$$

$$\epsilon'_{\mathrm{top\ tagged}} = 2\epsilon_{\mathrm{btag}}(1-\epsilon_{\mathrm{btag}}) + (\epsilon_{\mathrm{btag}})^2\,, \tag{D7}$$

$$\epsilon_{\mathrm{MC}} = \frac{N_{1,\mathrm{MC}}}{N_{0,\,\mathrm{MC}} + N_{1,\,\mathrm{MC}} + N_{2,\,\mathrm{MC}}}\,, \tag{D8}$$

$$\epsilon'_{\mathrm{MC}} = \frac{N_{1,\mathrm{MC}} + N_{2,\mathrm{MC}}}{N_{0,\,\mathrm{MC}} + N_{1,\,\mathrm{MC}} + N_{2,\,\mathrm{MC}}}\,, \tag{D9}$$

$$\frac{\epsilon_{\mathrm{MC}}}{\epsilon_{\mathrm{MC\ tagged}}} \equiv \frac{\epsilon'_{\mathrm{MC}}}{\epsilon'_{\mathrm{MC\ tagged}}} = \frac{(N_{1,\,\mathrm{MC}} + 2N_{2,\,\mathrm{MC}})^2}{4N_{2,\,\mathrm{MC}}(N_{0,\,\mathrm{MC}} + N_{1,\,\mathrm{MC}} + N_{2,\,\mathrm{MC}})}\,. \tag{D10}$$

These formulae also apply to the $0j$ channel except that the tagged b-jet is at lower P_T.

[1] M. Cacciari and G.P. Salam and G. Soyez, JHEP 04 (2008) 063, arXiv:0802.1189 [hep-ph].
[2] ATLAS Collaboration, Phys. Lett. B712 (2012) 289, arXiv:1203.6232 [hep-ex].
[3] ATLAS Collaboration, Phys. Lett. B716 (2012) 62, arXiv:1206.0756 [hep-ex].
[4] B. Mellado, X. Ruan and Z. Zhang, Phys. Rev. D84 (2011) 096005, arXiv:1101.1383 [hep-ph].
[5] X. Ruan, *Search for the Standard Model Higgs boson in the $WW^{(*)}$ channel and drift time measurement in the liquid argon calorimeter at ATLAS*, joint PhD thesis, Univ. Paris-Sud and IHEP, LAL Preprint 12-439, CERN-THESIS-2012-188.
[6] S. Frixione and B.R. Webber, JHEP 06 (2002) 029, arXiv:hep-ph/0204244.
[7] ATLAS Collaboration, Phys. Lett. B716 (2012) 1, arXiv:1207.7214 [hep-ex].

[8] ATLAS Collaboration, Phys. Lett. B726 (2013) 89, arXiv:1307.1427 [hep-ex].
[9] ATLAS Collaboration, Phys. Rev. D87 (2013) 112001, arXiv:1210.2979 [hep-ex].
[10] CMS Collaboration, Phys. Lett. B710 (2012) 91, arXiv:1202.1489 [hep-ex].
[11] CMS Collaboration, Eur. Phys. J. C73 (2013) 2469, arXiv:1304.0213 [hep-ex].
[12] Si Xie, *Search for the Standard Model Higgs Boson Decaying to Two W Bosons at CMS*, PhD thesis, MIT, CERN-THESIS-2012-068.
[13] S. Alioli, P. Nason, C. Oleari and E. Re, JHEP 0904 (2009) 002.
[14] T. Sjostrand, S. Mrenna and P.Z. Skands, JHEP 0605 (2006) 026.
[15] G. Corcella, I. Knowles, G. Marchesini, S. Maretti, K. Odagiri et al., JHEP 0101 (2001) 010.
[16] S. Frixione, E. Laenen, P. Motylinski, C. White and B.R. Webber, JHEP 07 (2009) 029.
[17] ATLAS Collaboration, Phys. Lett. B718 (2013) 879, arXiv:1208.2884 [hep-ex].
[18] M. Böhler, *Exclusive search for supersymmetry with same-flavor dilepton final states with the ATLAS detector*, PhD thesis, Hamburg University, DESY-THESIS-2012-022, CERN-THESIS-2012-245.
[19] ATLAS Collaboration, Eur. Phys. J. C72 (2012) 2056, arXiv:1203.5420 [hep-ex].
[20] CMS Collaboration, Phys. Rev. Lett. 109 (2012) 261802, arXiv:1210.2402 [hep-ex].
[21] CMS Collaboration, Phys. Lett. B714 (2012) 158, arXiv:1206.2402 [hep-ex].
[22] ATLAS Collaboration, Phys. Lett. B718 (2013) 860, arXiv:1208.2880 [hep-ex].