# Work Project Report

CERN Summer Student Programme

September 18, 2014

# A data driven estimation of the $t\bar{t}$ background for the measurement of the $W$ production cross section in association with jets with the ATLAS detector

Bastian Bernarding[a]

[a]*Heidelberg University*

**Abstract**

This report presents a data driven estimation of the $t\bar{t}$ background for the measurement of the $W$ production cross section in association with jets. The data set which was accumulated from proton-proton collisions at a center of mass energy of $\sqrt{s} = 8$ TeV at the Large Hadron Collider in 2012 is used for the measurement. The total number of $t\bar{t}$ events as a function of the jet multiplicity in the signal region of the muon channel is determined by performing a maximum likelihood fit of a $t\bar{t}$ template together with signal and other background Monte Carlo simulations to the data. Eight different $t\bar{t}$ templates obtained from data have been tested. Furthermore the method determines a shape estimate of the $t\bar{t}$ background for some kinematic distributions. The results are compared to the prediction of the next-to-leading-order Monte Carlo generator POWHEG.

# 1 Introduction

A precise knowledge of the $W$ production cross section in association with jets provides the possibility to test one of the most important parts of the Standard Model of Particle Physics, the theory of quantum chromodynamics (QCD). Furthermore it is crucial for some analyses studying the properties of the Higgs Boson or searching for new physics to know this production cross section to a very high precision in order to estimate background processes efficiently.

Since the top quark decays predominantly to a $W$ boson and a bottom quark ($t \to Wb$), $t\bar{t}$ events are important background processes to the $W+$ jets measurement. The most important contribution comes from those $t\bar{t}$ events where one of the $W$ bosons decays hadronically and the second one decays leptonically. For high jet multiplicities ($N_{jet} \geq 3$) $t\bar{t}$ events represent the dominant background contribution to $W$ events in association with jets. This analysis uses a data driven method to estimate the $t\bar{t}$ background, which reduces the systematic uncertainties compared to using Monte Carlo (MC) simulations. [1, 2]

# 2 LHC And The ATLAS Detector

In 2012 proton proton collisions at a center of mass energy of $\sqrt{s} = 8$ TeV were performed at the Large Hardon Collider (LHC). The full data set of 2012 corresponds to an integrated luminosity of 20.1 fb$^{-1}$. The ATLAS detector [3] is designed as a multi-purpose detector with a symmetric cylindrical geometry. The central collision point is surrounded by the inner tracker which consists of pixel and silicon microstrip detectors and a transition radiation tracker. It provides precision tracking and momentum measurements of charged particles for pseudorapidities $|\eta| < 2.5$. The inner tracker is followed by a superconducting solenoid providing a 2 T magnetic field, a calorimeter system using liquid argon, scintillator tiles or iron as active materials and a muon spectrometer consisting of superconducting toroids and a system of trigger and tracking chambers. The calorimeter system can be divided into the electromagnetic and the hadronic calorimeter measuring the energy of particles. The muon system provides another momentum measurement for muons which is independent of the momentum measurement of the inner detector.

# 3 Event Selection

The data used for this analysis was taken during the LHC 2012 run of proton proton collisions. An amount of data corresponding to an integrated luminosity of 19.5 fb$^{-1}$ is used for the $t\bar{t}$ background estimation. The most important selection criteria are explained in this chapter.

The analysis is performed in the muon channel. $W \to \mu\nu$ events are selected by using a single muon trigger. The transverse momentum threshold of the trigger is 24 GeV. Events are required to have at least one reconstructed vertex with at least three associated tracks.

Muons are required to have a transverse momentum $p_{\mathrm{T}} > 25$ GeV and a pseudorapidity of $|\eta| < 2.4$. In addition some vertex matching and isolation criteria are applied. The event is rejected if there is more than one muon passing these criteria.

$W$ candidates are selected by requiring the event to have a missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ greater than 25 GeV and a transverse mass $m_{\mathrm{T}}$ greater than 40 GeV.

Jets are reconstructed by using the anti-$k_T$ algorithm [4] with a distance parameter R = 0.4. A jet is required to have a transverse momentum $p_{\mathrm{T}} > 30$ GeV and a rapidity $|y| < 4.4$. To reject jets from pile up the jet vertex fraction (JVA) defined as the ratio of the sum of the transverse momentum of the tracks associated both to the primary vertex and the jet and the sum of the transverse momentum of all tracks associated to the jet must be greater than 0.5 for all jets with $p_{\mathrm{T}} > 50$ GeV and $|\eta| < 2.4$. In addition,

| Template Name | B-Jet Requirement | Working Point (MV1) |
|---|---|---|
| ttbar_template_1pbjet_60 | $\geq 1$ | 0.60 |
| ttbar_template_1pbjet_70 | $\geq 1$ | 0.70 |
| ttbar_template_1pbjet_80 | $\geq 1$ | 0.80 |
| ttbar_template_1pbjet_85 | $\geq 1$ | 0.85 |
| ttbar_template_2pbjet_60 | $\geq 2$ | 0.60 |
| ttbar_template_2pbjet_70 | $\geq 2$ | 0.70 |
| ttbar_template_2pbjet_80 | $\geq 2$ | 0.80 |
| ttbar_template_2pbjet_85 | $\geq 2$ | 0.85 |

Table 1: Definition of the eight different $t\bar{t}$ templates obtained from data.

an overlap removal for jets and muons or electrons is performed. Jets lying within a cone of $\Delta R = 0.5$ around a muon or an electron are rejected.

B-jets are reconstructed by using the MV1 tagging algorithm. Different working points (0.6, 0.7, 0.8 and 0.85) have been tested during the analysis. Moreover, b-jets must have a transverse momentum of $p_{\mathrm{T}} > 20$ GeV and a pseudorapidity of $|\eta| < 2.5$ due to the acceptance of the tracking detectors.

# 4 Background Estimation Method

The number of $t\bar{t}$ events is determined by fitting a $t\bar{t}$ template together with a second template containing MC simulations of the $W \to \mu\nu$ signal and other backgrounds to the data in a discriminant distribution for each jet multiplicity. The $t\bar{t}$ templates provide the shape of the $t\bar{t}$ events in the discriminant distribution. The discriminant itself should show a good separation of the $t\bar{t}$ event shape from the signal event shape. Finally, the correct normalisation of the $t\bar{t}$ template and therefore the absolute number of $t\bar{t}$ background events for each jet multiplicity can be obtained from the maximum likelihood fit. Since the $t\bar{t}$ template is obtained from data, the described method is a data driven approach. This chapter explains the definition of the different $t\bar{t}$ templates, the choice of the discriminant and the details of the fits.

## 4.1 $t\bar{t}$ Templates

Since the top quark decays with a probability close to 100% into a $W$ boson and a bottom quark, the $t\bar{t}$ templates are constructed by including an additional b-jet requirement to the signal selection. At least one or at least two b-jets are required. These two different requirements and the four different working points of the MV1 tagging algorithm result into eight different $t\bar{t}$ templates which are summarised in table 1. Notice that each $t\bar{t}$ template is a subsample of the signal data sample. The selection of the signal and the templates are not orthogonal. Therefore the $t\bar{t}$ templates are statistically correlated to the signal data sample. One possibility to avoid this correlation is applying a b-jet veto in the signal selection. This approach will be tested in future analyses.

## 4.2 Transformed Aplanarity

The transformed aplanarity is chosen to be the discriminant which should separate the $t\bar{t}$ events from the $W$ + jets signal events, since it was also used in the top cross section paper to separate $W$ + jets from top events [5]. The transformed aplanarity is defined as $A_{\mathrm{trans}} = \exp(-8 \cdot \text{aplanarity})$, where the aplanarity $A$

is defined as $A = 3\lambda_3$. $\lambda_3$ is the smallest eigenvalue of the full momentum tensor of the event $M_{\mathrm{xyz}}$

$$M_{\mathrm{xyz}} = \begin{pmatrix} p_{x_i}^2 & p_{x_i}p_{y_i} & p_{x_i}p_{z_i} \\ p_{y_i}p_{x_i} & p_{y_i}^2 & p_{y_i}p_{z_i} \\ p_{z_i}p_{x_i} & p_{z_i}p_{y_i} & p_{z_i}^2 \end{pmatrix}, \tag{1}$$

where the sum over $i$ runs over all jets of the event as defined in section 3. The transformed aplanarity measures the part of the transverse momentum in or out of the plane defined by the two leading jets. A transformed aplanarity of 1 corresponds to a planar event, whereas a transformed aplanarity near 0 describes a spherical event [6]. The basic principle behind the idea to use the transformed aplanarity to separate $W$ + jets events from $t\bar{t}$ events is that $t\bar{t}$ events tend to be more isotropic and therefore more spherical than $W$ + jets events [5].

### 4.3 Maximum Likelihood Fit

The fits to the transformed aplanarity distribution are performed as described in section 4 for each jet multiplicity using TFractionFitter [7] and RooFit [8]. The fit range is chosen to be $0 \leq A_{\mathrm{trans}} \leq 0.85$, because jet energy scale (JES) uncertainties and pile up dependence have a large influence on the distribution for $A_{\mathrm{trans}} > 0.85$. It would be reasonable to split the template with the signal and the QCD background simulations into two templates and allow the normalisation of the signal template to be slightly modified during the fit routine. This would account for the uncertainties of the signal MC simulation and could improve the fit. This approach was tested during the analysis. An improvement of the fit could not be noticed. Therefore the normalisation of the template containing the signal and the QCD background processes has been fixed to its initial value during the fit routine. The $\chi^2_{\mathrm{red}}$ values are used to evaluate the goodness of the fits.

## 5 Choice Of A $t\bar{t}$ Template For Further Analysis

The $t\bar{t}$ template used for further analyses is identified using the following criteria: statistics, template contamination from other processes, goodness of the fits and bias in some kinematic distributions.

Since the $t\bar{t}$ template is taken from data, it does not only consist of $t\bar{t}$ events. A tight selection results into a very pure template, but the statistics decrease at the same time. Therefore a good compromise between high statistics and a small contamination must be found. The statistics of the templates is evaluated by the number of events in the fit range (table 3). The templates requiring at least two b-jets show as expected a smaller number of events in the fit range. This might result into larger uncertainties in the fit for high jet multiplicities. Nevertheless, no template can be rejected by looking at the statistics.

The contamination of the $t\bar{t}$ templates from the signal and other QCD backgrounds is estimated using MC simulations of these processes and applying the same selection criteria as for the different $t\bar{t}$ templates. Table 4 and table 5 show the signal and the QCD background processes included in the fit routine as well as the total number of events of these processes for the different $t\bar{t}$ template selections and the contamination of the $t\bar{t}$ templates. The two templates requiring at least one b-jet at high working points (0.80 and 0.85) show large contaminations from the signal (around 27 % and 10 %) and $Wc$ processes (around 10 %) and are therefore not used for further analyses. All $t\bar{t}$ templates with a two b-jet requirement show in general a higher purity.

The selection criteria used to get the different $t\bar{t}$ templates could result into a bias in some kinematic distributions. This bias has to be small because the presented data driven estimate will be used to subtract the $t\bar{t}$ background from all measurement distributions. A bias in one of these distribution would have an impact on the accuracy of this approach. Therefore the bias is investigated in the rapidity and the transverse momentum distribution of the jets and the $H_{\mathrm{T}}$ distribution by comparing the distributions

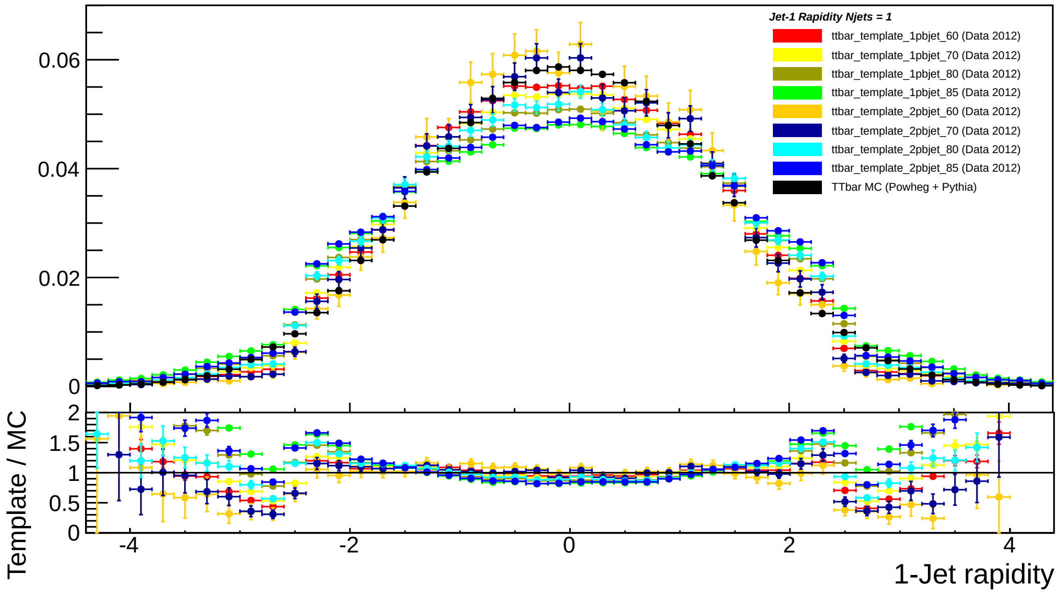



Figure 1: Comparison of the jet rapidity distribution of the leading jet for one jet events from the eight different $t\bar{t}$ templates and the MC simulation. The lower plot shows the ratio of the distribution from the templates and the one obtained from the POWHEG simulation.

from the templates to a $t\bar{t}$ Monte Carlo simulation, since these distributions have shown the largest bias in previous analyses. The distributions are normalised to the area in order to allow a comparison. The jet $p_{\mathrm{T}}$ (fig. 8) and the $H_{\mathrm{T}}$ (fig. 9) distribution show a large difference between the templates and the MC simulation. These differences could result from bulk QCD processes that are included in the templates but not in the MC simulation. They are dominant at lower energies and could therefore explain the higher peak and due to the normalisation also the flatter tale of the templates in these two distributions. The influence of the bulk QCD backgrounds on the jet rapidity distribution (fig. 1) is expected to be smaller. Therefore this distribution is mainly used to investigate the bias of the templates. Under the templates with the one b-jet criterion the ones with the highest working points show the largest bias. This can be explained with the fact that these distributions also have the largest contamination from other processes. The templates with a two b-jet requirement and high working points (0.85 and 0.8) show a non negligible bias of around 20 % in the central region although they have a higher purity. These templates are excluded from further analyses. The two b-jet selection criterion itself seems to have an influence on the jet rapidity distribution.

The $\chi^2_{\mathrm{red}}$ values of the fits can be found in table 6 and table 7. Higher values at smaller jet multiplicities can result again from bulk QCD processes which are not subtracted during the fit routine. No $t\bar{t}$ template is rejected because of the goodness of the fit.

For further analyses the template 'ttbar_template_1pbjet_60' is chosen, because it shows sufficient statistics, a small contamination and a small bias in the jet rapidity distribution. Notice that this study would also justify to use the following templates, 'ttbar_template_1pbjet_70', 'ttbar_template_2pbjet_60' and 'ttbar_template_2pbjet_70'.

# 6 Results

The number of predicted $t\bar{t}$ events in the signal region of the muon channel as a function of the jet multiplicity is shown in fig. 2 and table 2. The given uncertainties for the data driven method include the

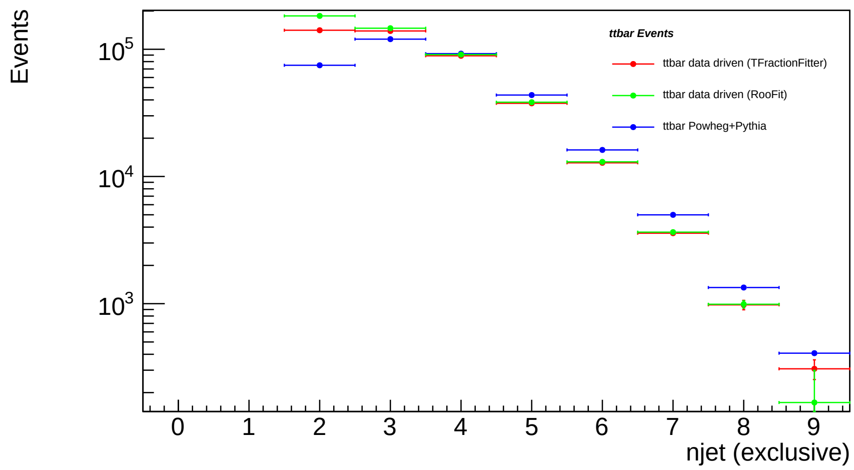



Figure 2: Comparison of the predicted number of $t\bar{t}$ events as a function of the jet multiplicity between the POWHEG MC simulation and the data driven estimation.

| $N_{\text{jets}}$ | POWHEG | Number Of Predicted $t\bar{t}$ Events<br>Data Driven (TFractionFitter) | Data Driven (RooFit) |
|---|---|---|---|
| 2 | 74862.5 ± 273.61 | 141155 ± 375.71 | 182945 ± 427.73 |
| 3 | 120136 ± 346.61 | 139295 ± 1406.56 | 146437 ± 1102.51 |
| 4 | 92416 ± 304.0 | 88884.2 ± 890.11 | 90858.6 ± 702.02 |
| 5 | 43672.1 ± 208.98 | 37591.8 ± 552.23 | 38358.6 ± 434.65 |
| 6 | 16173.1 ± 127.18 | 12808.5 ± 315.74 | 13031.3 ± 247.34 |
| 7 | 4994.3 ± 70.68 | 3582.11 ± 163.92 | 3651.03 ± 128.63 |
| 8 | 1339.7 ± 36.61 | 978.89 ± 83.24 | 989.79 ± 64.89 |
| 9 | 408.17 ± 20.21 | 307.57 ± 54.31 | 167.14 ± 131.70 |

Table 2: Prediction of the number of $t\bar{t}$ events in the signal region of the muon channel for each jet multiplicity. The uncertainties for the data driven methods include the statistical uncertainties and the uncertainties on the fits.

statistical uncertainty and the uncertainty on the fit. At high jet multiplicities the data driven approach predicts a smaller number of $t\bar{t}$ events than the MC simulation. For $N_{\text{jet}} = 2$ and $N_{\text{jet}} = 3$ the estimate of the data driven method is larger compared to the MC prediction. This can be seen as a result of the bulk QCD background processes which are not subtracted from the data in the fit routine. Therefore the fit seems to assign a part of these events to the $t\bar{t}$ background. The sum of the simulated signal, all simulated backgrounds and the $t\bar{t}$ background from the data driven estimation in the number of jets distribution is still lower than the data for these jet multiplicities (fig. 3 and fig. 4), which is expected because of the bulk QCD backgrounds not included in the analysis.

This could be seen as an argument for the robustness of the fitting method because the shape estimates

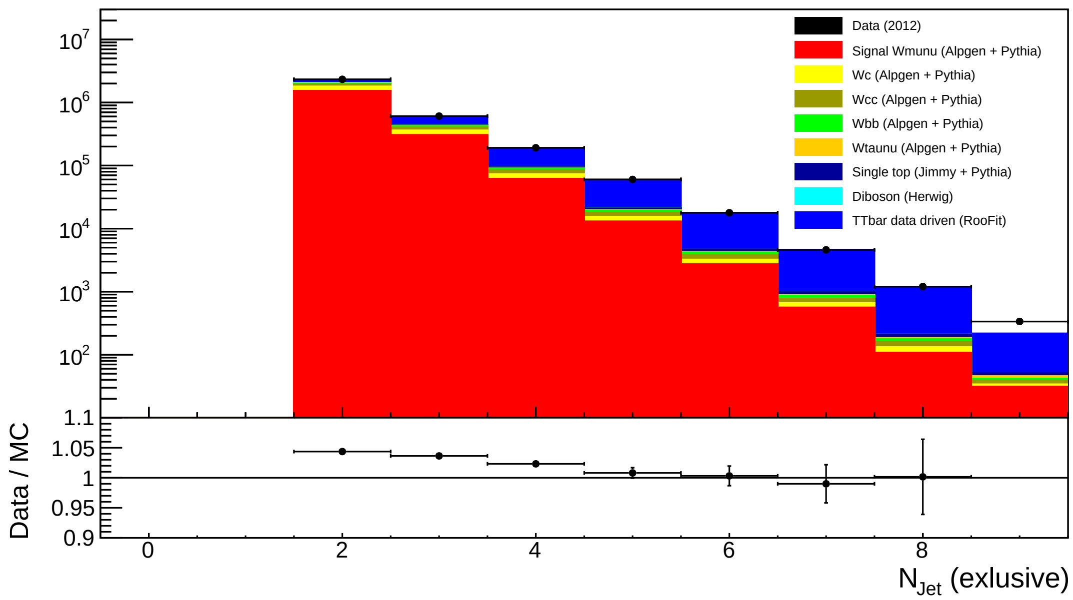



Figure 3: Number of jets distribution for data and for the sum of the simulated signal and simulated backgrounds plus the $t\bar{t}$ background from the data driven estimation with RooFit.

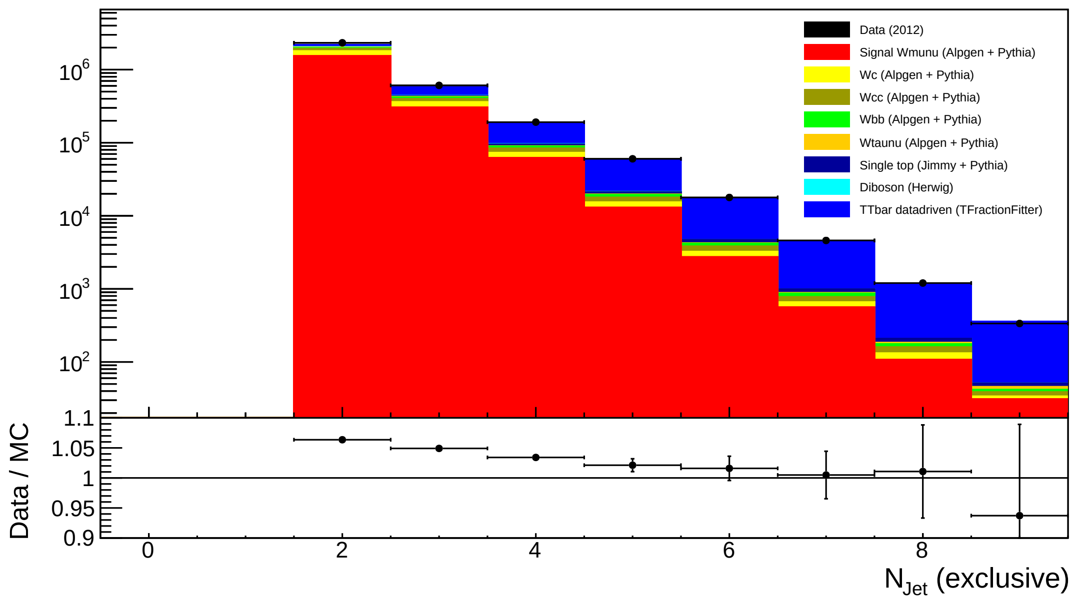
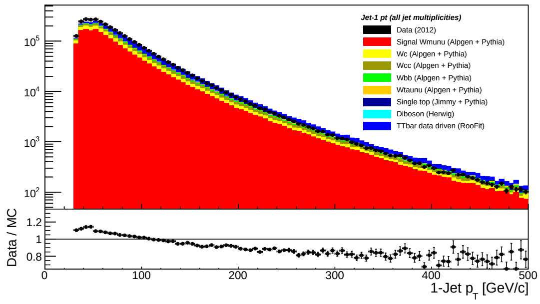


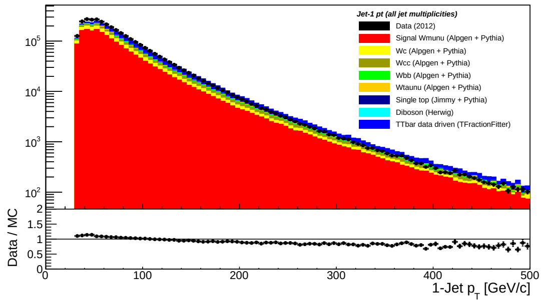

Figure 4: Number of jets distribution for data and for the sum of the simulated signal and simulated backgrounds plus the $t\bar{t}$ background from the data driven estimation with TFractionFitter.

seem to have a bigger influence on the fit result than the actual number of events. This underlines the significance of a good $t\bar{t}$ shape estimate and therefore of the purity of the used template. Furthermore, fig. 3 and fig. 4 show a very good agreement between the data and the predicted signal and backgrounds for higher multiplicities, whereas the results from RooFit seem to provide a slightly better prediction. The $t\bar{t}$ background estimate for $N_{\text{jet}} = 9$ has to be excluded from further analyses because of low statistics.

These results are used to make a shape estimate of the $t\bar{t}$ background for other kinematic distributions, such as the $H_{\text{T}}$ distribution, the transverse momentum of the leading jet and the rapidity of the leading jet for an arbitrary number of jets (fig. 5, fig. 6 and fig. 7 ). The first two distributions show again more events in the data at low energies due to the missing bulk QCD processes in the background simulation. In the rapidity distribution the sum of the predicted signal and backgrounds is slightly below the data, which could also be a result of missing QCD background contributions.

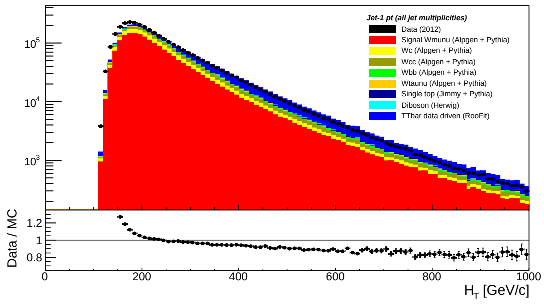



(a)

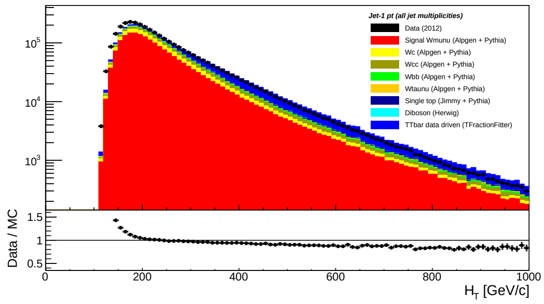



(b)

Figure 5: Transverse momentum of the leading jet for an arbitrary number of jets for data and MC simulation of the signal and the QCD backgrounds together with the $t\bar{t}$ estimate from the data driven method for RooFit (a) and TFractionFitter (b).

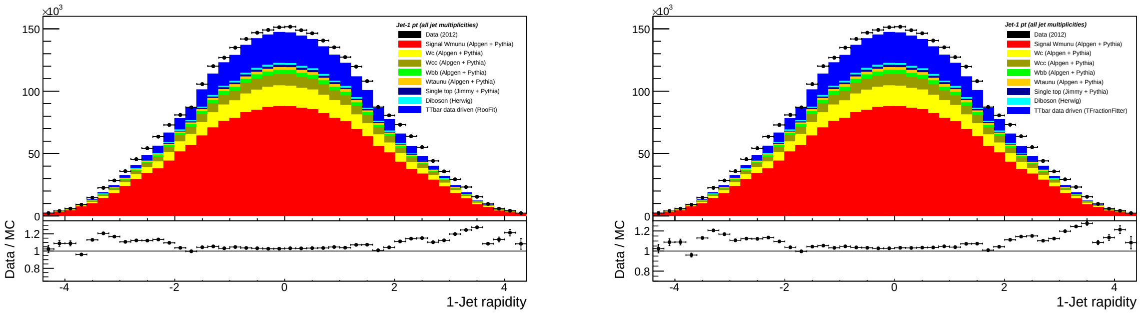



(a)




(b)

Figure 6: $H_{\text{T}}$ distribution for an arbitrary number of jets for data and MC simulation of the signal and the QCD backgrounds together with the $t\bar{t}$ estimate from the data driven method for RooFit (a) and TFractionFitter (b).




(a) (b)

Figure 7: Leading jet rapidity for an arbitrary number of jets for data and MC simulation of the signal and the QCD backgrounds together with the $t\bar{t}$ estimate from the data driven method for RooFit (a) and TFractionFitter (b).

## 7 Conclusion

A data driven estimate for the $t\bar{t}$ background for the $W$ + jets cross section measurement is presented in this note. The number of $t\bar{t}$ events in the signal region of the muon channel has been predicted for each jet multiplicity. Shape estimates for some kinematic distributions have been made from these results. The estimation of the $t\bar{t}$ background at low jet multiplicities is influenced by some bulk QCD processes which are not included in the fit routine. The comparison with the $t\bar{t}$ MC simulation (POWHEG) shows that the data driven method predicts a smaller number of $t\bar{t}$ events at high jet multiplicity. Especially in this region the data driven approach describes the observed data in a better way compared to the MC prediction (compare fig. 11 and fig. 3, fig. 4). The presented data driven method could reduce the uncertainty from the $t\bar{t}$ background estimation on the $W$ cross section.

# A Choice Of A $t\bar{t}$ Template

| Template Name | Number Of Events In Fit Region |
|---|---|
| ttbar_template_1pbjet_60 | 223483 ± 472.74 |
| ttbar_template_1pbjet_70 | 274059 ± 523.506 |
| ttbar_template_1pbjet_80 | 382953 ± 618.832 |
| ttbar_template_1pbjet_85 | 503438 ± 709.534 |
| ttbar_template_2pbjet_60 | 59824 ± 244.589 |
| ttbar_template_2pbjet_70 | 85988 ± 293.237 |
| ttbar_template_2pbjet_80 | 134034 ± 366.107 |
| ttbar_template_2pbjet_85 | 193927 ± 440.371 |

Table 3: Number of events in the fit range combined for all jet multiplicities to evaluate the statistics of the $t\bar{t}$ templates.

| Template Name | Process | Contamination | Number Of Events In The Fit Range |
|---|---|---|---|
| | $W \to \mu\nu$ | 3.33 ± 0.05 % | 444.98 ± 104.75 |
| | $Wc$ | 6.01 ± 0.06 % | 13440.9 ± 109.20 |
| | $Wcc$ | 2.59 ± 0.04 % | 5788.49 ± 72.68 |
| ttbar_template_1pbjet_60 | $Wbb$ | 5.10 ± 0.05 % | 11392.8 ± 99.93 |
| | $W \to \tau\nu$ | 0.07 ± 0.02 % | 150.67 ± 29.77 |
| | Single top | 4.66 ± 0.03 % | 10413.6 ± 60.65 |
| | Diboson | 0.47 ± 0.01 % | 1057.7 ±13.57 |
| | $W \to \mu\nu$ | 6.43 ± 0.06 % | 17629.9 ± 149.07 |
| | $Wc$ | 8.12 ± 0.06 % | 22249.6 ± 140.37 |
| | $Wcc$ | 3.58 ± 0.04 % | 9820.07 ± 94.75 |
| ttbar_template_1pbjet_70 | $Wbb$ | 4.88 ± 0.05% | 13379.8 ± 108.09 |
| | $W \to \tau\nu$ | 0.13 ± 0.02 % | 360.01 ± 40.79 |
| | Single top | 4.35 ± 0.03 % | 11920.2 ± 64.99 |
| | Diboson | 0.58 ± 0.01% | 1593.31 ± 16.86 |
| | $W \to \mu\nu$ | 17.01 ± 0.08 % | 65121.7 ± 259.34 |
| | $Wc$ | 10.27 ± 0.06 % | 39336.5 ± 186.12 |
| | $Wcc$ | 4.77 ± 0.04 % | 18265.9 ± 129.70 |
| ttbar_template_1pbjet_80 | $Wbb$ | 4.21 ± 0.04 % | 16108.7 ± 118.15 |
| | $W \to \tau\nu$ | 0.39 ± 0.03 % | 1483.93 ± 97.65 |
| | Single top | 3.53 ± 0.02 % | 13504.1 ± 69.30 |
| | Diboson | 0.78 ± 0.01 % | 2985.06 ± 23.37 |
| | $W \to \mu\nu$ | 27.12 ± 0.09 % | 136513 ± 365.40 |
| | $Wc$ | 10.94 ± 0.05 % | 55092.2 ± 219.89 |
| | $Wcc$ | 5.35 ± 0.04 % | 26927.9 ± 157.89 |
| ttbar_template_1pbjet_85 | $Wbb$ | 3.61 ± 0.03 % | 18175.8 ± 125.03 |
| | $W \to \tau\nu$ | 0.62 ± 0.03 % | 3096.66 ± 144.27 |
| | Single top | 2.87 ± 0.02 % | 14439.6 ± 71.76 |
| | Diboson | 0.94 ± 0.01 % | 4713.81 ± 29.52 |

Table 4: Contamination of the different $t\bar{t}$ templates from the signal and other QCD background processes obtained from MC.

| Template Name | Process | Contamination | Total Number Of Events In The Fit Range |
|---|---|---|---|
| ttbar_template_2pbjet_60 | $W \to \mu\nu$ | 0.28 ± 0.04% | 167.35 ± 21.99 |
| | $Wc$ | 0.24 ± 0.02 % | 145.33 ± 10.79 |
| | $Wcc$ | 0.33 ± 0.03 % | 198.04 ± 13.44 |
| | $Wbb$ | 3.22 ± 0.08 % | 1924.8 ± 41.71 |
| | $W \to \tau\nu$ | 0.04 ± 0.04 % | 23.82 ± 20.21 |
| | Single top | 3.15 ± 0.05 % | 1882.42 ± 24.71 |
| | Diboson | 0.19 ± 0.01 % | 110.52 ± 3.93 |
| ttbar_template_2pbjet_70 | $W \to \mu\nu$ | 0.64 ± 0.05 % | 552.33 ± 36.89 |
| | $Wc$ | 0.61 ± 0.03 % | 526.62 ± 20.79 |
| | $Wcc$ | 0.70 ± 0.03 % | 608.54 ± 23.61 |
| | $Wbb$ | 3.25 ± 0.06 % | 2798.28 ± 50.13 |
| | $W \to \tau\nu$ | 0.04 ± 0.03 % | 31.80 ± 20.98 |
| | Single top | 3.29 ± 0.04 % | 2827.45 ± 30.49 |
| | Diboson | 0.21 ± 0.01 % | 181.22 ± 5.11 |
| ttbar_template_2pbjet_80 | $W \to \mu\nu$ | 3.17 ± 0.06 % | 4245.86 ± 77.11 |
| | $Wc$ | 2.38 ± 0.04 % | 3184.74 ± 51.45 |
| | $Wcc$ | 1.75 ± 0.04 % | 2350.83 ± 46.63 |
| | $Wbb$ | 3.47 ± 0.05 % | 4654.64 ± 64.03 |
| | $W \to \tau\nu$ | 0.08 ± 0.02 % | 103.14 ± 26.66 |
| | Single top | 3.47 ± 0.04 % | 4645.88 ± 39.63 |
| | Diboson | 0.31 ± 0.01 % | 411.24 ± 8.11 |
| ttbar_template_2pbjet_85 | $W \to \mu\nu$ | 8.94 ± 0.08 % | 17344.8 ± 138.97 |
| | $Wc$ | 4.62 ± 0.05 % | 8951.37 ± 86.62 |
| | $Wcc$ | 2.92 ± 0.04 % | 5669.94 ± 72.66 |
| | $Wbb$ | 3.46 ± 0.05 % | 6705.05 ± 76.28 |
| | $W \to \tau\nu$ | 0.17 ± 0.03 % | 329.04 ± 39.11 |
| | Single top | 3.39 ± 0.03 % | 6575.79 ± 47.64 |
| | Diboson | 0.47 ± 0.01 % | 917.82 ± 12.62 |

Table 5: Contamination of the different $t\bar{t}$ templates from the signal and other QCD background processes obtained from MC.

| Template Name | $N_{\text{jet}}$ | $\chi^2_{\text{red}}$ (TFractionFitter) | $\chi^2_{\text{red}}$ (RooFit) |
|---|---|---|---|
| ttbar_template_1pbjet_60 | 2 | 1.31 | 8.44 |
| | 3 | 0.55 | 2.86 |
| | 4 | 0.37 | 1.24 |
| | 5 | 0.26 | 0.70 |
| | 6 | 0.22 | 0.49 |
| | 7 | 0.18 | 0.33 |
| | 8 | 0.12 | 0.22 |
| | 9 | 0.26 | 0.34 |
| ttbar_template_1pbjet_70 | 2 | 1.22 | 8.10 |
| | 3 | 0.44 | 2.43 |
| | 4 | 0.25 | 0.93 |
| | 5 | 0.20 | 0.54 |
| | 6 | 0.17 | 0.37 |
| | 7 | 0.15 | 0.28 |
| | 8 | 0.13 | 0.22 |
| | 9 | 0.17 | 0.16 |
| ttbar_template_1pbjet_80 | 2 | 1.13 | 7.85 |
| | 3 | 0.96 | 3.23 |
| | 4 | 0.69 | 1.73 |
| | 5 | 0.24 | 0.60 |
| | 6 | 0.15 | 0.33 |
| | 7 | 0.14 | 0.24 |
| | 8 | 0.07 | 0.12 |
| | 9 | 0.14 | 0.12 |
| ttbar_template_1pbjet_85 | 2 | 1.106 | 7.76 |
| | 3 | 2.17 | 5.41 |
| | 4 | 1.73 | 3.63 |
| | 5 | 0.38 | 0.88 |
| | 6 | 0.16 | 0.34 |
| | 7 | 0.12 | 0.22 |
| | 8 | 0.07 | 0.12 |
| | 9 | 0.12 | 0.05 |

Table 6: $\chi^2_{\text{red}}$ values for the templates requiring at least one b-jets for all fits.

| Template Name | $N_{\text{jet}}$ | $\chi^2_{\text{red}}$ (TFractionFitter) | $\chi^2_{\text{red}}$ (RooFit) |
|---|---|---|---|
| ttbar_template_2pbjet_60 | 2 | 2.09 | 11.94 |
| | 3 | 2.49 | 9.88 |
| | 4 | 2.10 | 7.83 |
| | 5 | 0.85 | 3.01 |
| | 6 | 0.66 | 3.08 |
| | 7 | 0.41 | 1.19 |
| | 8 | 0.56 | 1.42 |
| | 9 | 0.53 | 0.98 |
| ttbar_template_2pbjet_70 | 2 | 1.87 | 10.76 |
| | 3 | 2.14 | 7.73 |
| | 4 | 1.75 | 5.49 |
| | 5 | 0.68 | 2.06 |
| | 6 | 0.61 | 1.78 |
| | 7 | 0.40 | 0.95 |
| | 8 | 0.34 | 0.86 |
| | 9 | 0.38 | 0.61 |
| ttbar_template_2pbjet_80 | 2 | 1.52 | 9.30 |
| | 3 | 1.22 | 4.74 |
| | 4 | 0.81 | 2.40 |
| | 5 | 0.48 | 1.28 |
| | 6 | 0.36 | 0.82 |
| | 7 | 0.26 | 0.53 |
| | 8 | 0.19 | 0.44 |
| | 9 | 0.25 | 0.30 |
| ttbar_template_2pbjet_85 | 2 | 1.31 | 8.46 |
| | 3 | 0.47 | 2.67 |
| | 4 | 0.29 | 1.08 |
| | 5 | 0.28 | 0.76 |
| | 6 | 0.22 | 0.49 |
| | 7 | 0.20 | 0.36 |
| | 8 | 0.15 | 0.26 |
| | 9 | 0.17 | 0.16 |

Table 7: $\chi^2_{\text{red}}$ values for the templates requiring at least two b-jets for all fits.

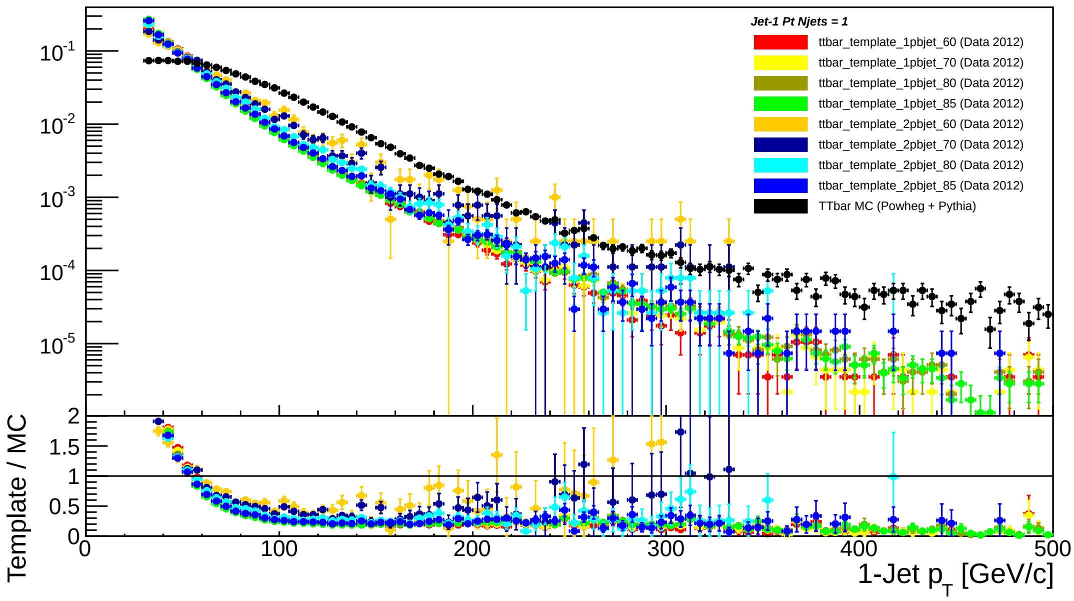



Figure 8: Comparison of the transverse momentum distribution of the leading jet for one jet events from the eight different $t\bar{t}$ templates and the MC simulation. The lower plot shows the ratio of the distribution from the templates and the one obtained from the POWHEG simulation.

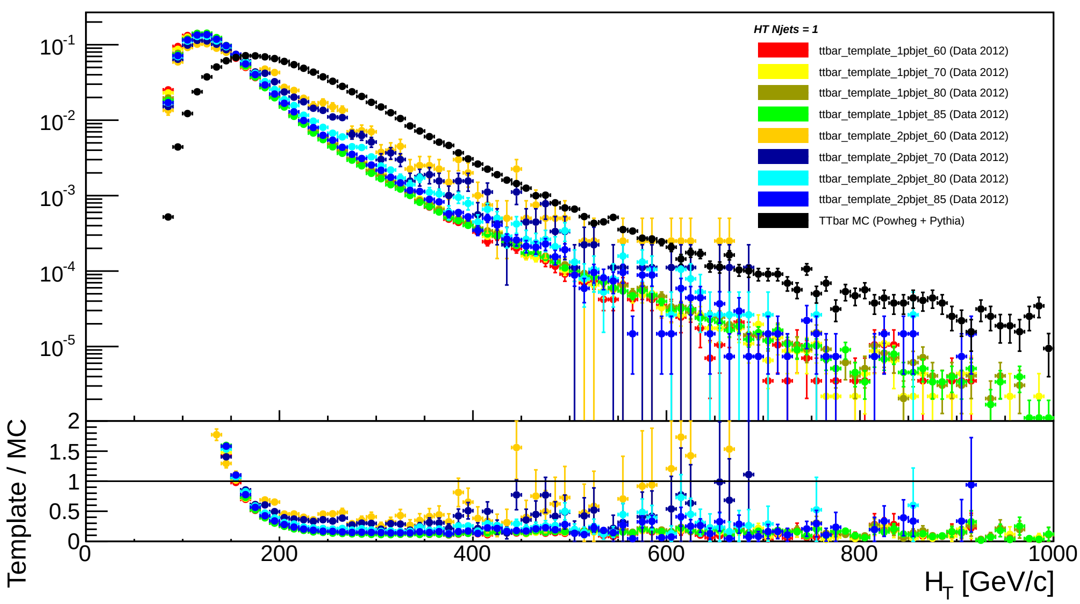



Figure 9: Comparison of the $H_{\mathrm{T}}$ distribution for one jet events from the eight different $t\bar{t}$ templates and the MC simulation. The lower plot shows the ratio of the distribution from the templates and the one obtained from the POWHEG simulation.

## B Fit Routine

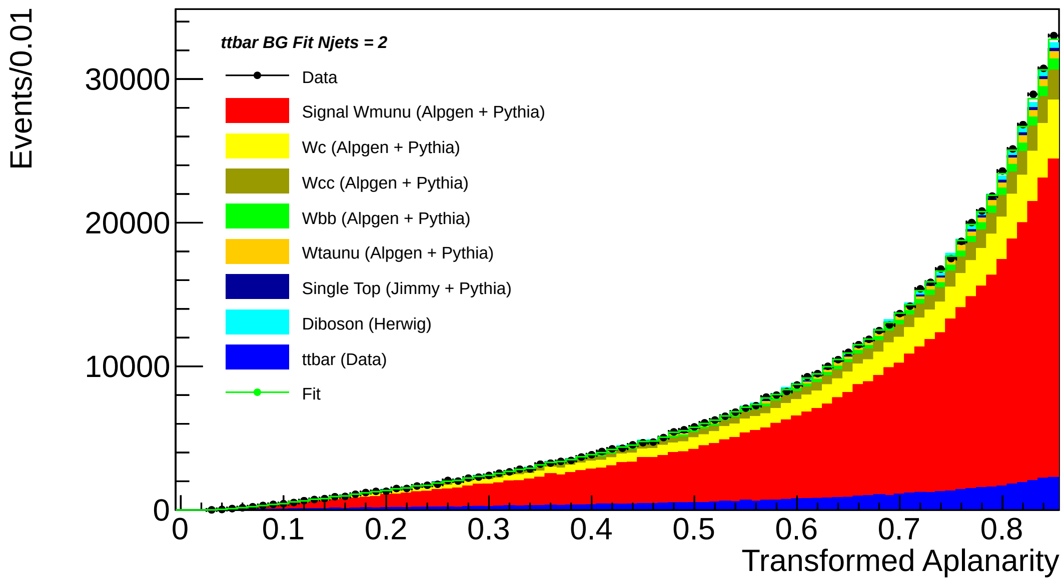



Figure 10: Example of the fit for $N_{\text{jet}} = 2$ and the processes included in the fit routine. The normalisation of the template including the signal and the QCD backgrounds has been fixed.

## C MC Signal and Background Prediction

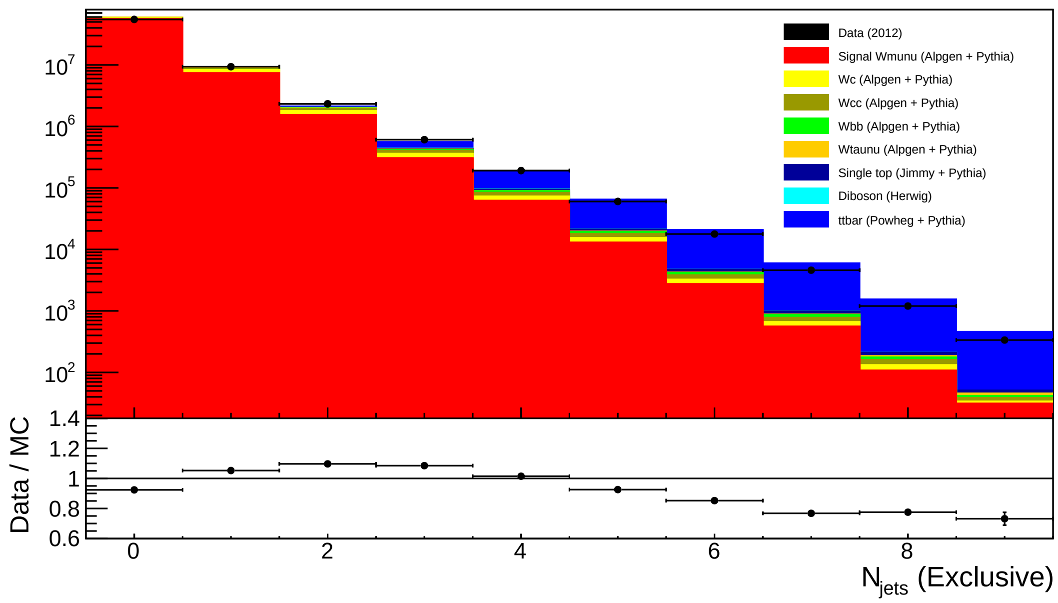



Figure 11: Prediction of pure MC simulations of the signal and the listed backgrounds compared to the data.

## References


[1] ATLAS Collaboration, *A measurement of the ratio of W to Z production cross sections in association with jets with the ATLAS detector*, Tech. Rep. arXiv:1408.6510. CERN-PH-EP-2014-200, CERN, Geneva, Aug, 2014.

[2] ATLAS Collaboration, *Measurements of the W production cross sections in association with jets with the ATLAS detector*, Tech. Rep. CERN-PH-EP-2014-199, CERN, Geneva, 2012.

[3] ATLAS Collaboration, G. Aad et al., *The ATLAS experiment at the CERN large hadron collider*, JINST **3** (2008) S08003.

[4] G. S. M. Cacciari, G. P. Salam, *The anti-$k_t$ jet clustering algorithm*, J. High Energy Phys. **04** no. arXiv:0802.1189. LPTHE-07-03, (2008).

[5] G. Aad, B. Abbott, J. Abdallah, A. Abdelalim, A. Abdesselam, O. Abdinov, B. Abi, M. Abolins, H. Abramowicz, H. Abreu, et al., *Measurement of the top quark pair production cross-section with ATLAS in the single lepton channel*, Physics Letters B **711** no. arXiv:1201.1889 CERN-PH-EP-2011-201, (2012) 244–263.

[6] A. collaboration et al., *Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at sqrt (s)= 7 TeV*, Eur. Phys. J. no. arXiv:1206.2135 CERN-PH-EP-2012-119, (2012).

[7] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Computer Physics Communications **77** no. 2, (1993) 219–228.

[8] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, arXiv:physics/0306116 (2003).