ATLAS NOTE

29th July 2016

Review of the ATLAS Open Data Dataset

The ATLAS Collaboration

ATL-OREACH-PUB-2016-001
03 August 2016

Abstract

The ATLAS collaboration is releasing to the public for educational purposes 1fb^{-1} of real data at a centre-of-mass energy of 8 TeV from the 2012 data taking period. This set of real data is accompanied by matching simulated data of several Standard Model processes and a Beyond the Standard Model signal. Analysis tools are provided to make analysis of the data easily accessible. The purpose of the data and tools released is to enable users to experience the analysis of particle physics data in a simplified environment, for example, in lab courses or as an extension of physics masterclasses.

This document summarises the properties of the ATLAS open data dataset and the analysis tools. In addition, example analyses intended as starting points for further analysis work by users are shown and their results reviewed.

Contents

1 Introduction

The ATLAS collaboration is releasing an official dataset, open to the public for educational use only, following the guidelines of the ATLAS Data Access Policy [1]. The dataset consists of real data with an integrated luminosity (1.0 ± 0.0019) fb^{-1} and a centre-of-mass energy of 8 TeV with matching simulated data. The dataset is intended to provide the means for doing hands-on particle physics exercises in the context of higher education, for example laboratory courses or introductory exercises for undergraduate students. The released data may also prove beneficial for the production of teaching materials, for lectures, and public talks. Furthermore, it may be used by people with data analysis experience but not necessarily a physics background as a test dataset for studying and developing analysis techniques. The Kaggle Higgs Boson Machine Learning Challenge [2] has demonstrated the viability of this application scenario.

The released data is provided in a simplified format to reduce the complexities of a full-scale analysis, decrease the processing time, and facilitate code development. Analysis code is written in Python and several example analyses are available as a starting point for further work. The technical details about the dataset are discussed in the following sections. Section 1.1 details the preselection of the datasets and explains the simplifications that have been made in comparison with a physics analysis. Section 2 describes the example analyses, which are shown afterwards.

The example analyses include a single W boson, Z boson, and top quark pair production analysis, all of which have sufficiently high event yields to study the processes in detail. Analyses for processes with lower production cross sections, namely WZ, ZZ, and $H \to WW$, are used to illustrate the statistical limitations of the dataset. Finally, a Z' analysis is included to allow searches for new physics, again with an emphasis on the educational character of the exercise.

1.1 Datasets

The ATLAS open data dataset is comprised of real data recorded with the ATLAS detector in 2012 and matching simulated data. Both real and simulated data are subjected to a loose event preselection to reduce processing time by reducing the overall number of events that have to be analysed. The preselection consists of a set of object selection criteria listed in Table 1. A further event selection is applied to these selected objects, defined by the following criteria:

- Corrupted event protection;
- Single lepton trigger satisfied;
- Veto on events containing bad jets. Bad jets are jets not associated to energy deposits in the calorimeters from particles originating from the primary pp collision. They arise from various sources, ranging from LHC beam conditions and cosmic-ray showers;
- Primary vertex cut ($N_{\text{tracks}} > 4$);
- At least one preselected lepton with $p_{\text{T}} > 25$ GeV.

The selected events are available in a simplified data format reducing the information content of the original data analysis format used in ATLAS. The resulting format is a TTree with 45 branches as detailed in Table 3. The layout is optimised towards simplicity to reduce the complexities encountered in a full-scale analysis, emphasising the educational character of the dataset. The framework used by the ATLAS top analysis group, called “AnalysisTop", was used to derive the simplified structure.

The set of real data has an integrated luminosity $(1.0007 \pm 0.019)\,\text{fb}^{-1}$ and a centre-of-mass energy of 8 TeV. Events from the Egamma and Muon streams from runs 207490, 207532, 207582, 207589, 207749, 207772, 207845, 207865, 207934, 207982, 208126, 208184, 208189, and 208258 are selected. All runs

electrons	muons	jets
reconstruction author 1||3	Muid combined	antiKt4LCTopo
medium++ quality	tight quality	jet cleaning (veto BadLooseMinus)
$p_{\text{T}} > 5$ GeV	$p_{\text{T}} > 5$ GeV	$p_{\text{T}} > 25$ GeV
$\|\eta\| < 2.47$ w/o crack	$\|\eta\| < 2.5$	$\|\eta\| < 2.5$
Object Quality is Good	MCP Hit requirement.	
$\|z_0\| < 2.0$ mm	$\|z_0\| < 2.0$ mm	
not Converted		

Table 1: Preselection requirements for electrons, muons and jets, as applied in the ATLAS top group analysis framework. The electron reconstruction algorithm is either calorimeter-based (author value 1) or both calorimeter and track based (author value 3). The term “w/o crack” refers to the so-called crack region located at $1.37 < |\eta| < 1.52$ where detector performance is degraded. Electron candidates found in this region are discarded. In addition, the term “not converted” denotes that electrons which have been identified as originating from photon conversion are not considered further. The Muid muon reconstruction algorithm package is used and muons are required to satisfy the set of requirements on detector hits issued by the Muon Combined Performance (MCP) group. Jet reconstruction is carried out by the anti-k_t clustering algorithm, with clusters identified by energy in calorimeter cells. Jet cleaning is applied. Jets are labelled as bad if they are not associated to energy deposits in the calorimeters from particles originating from pp collisions.

belong to period D of the 2012 data taking and form the input for the preselection resulting in the dataset summarised in Table 4.

The simulated datasets used in the data release are shown in Table 5 and Table 6. The same preselection as for real data is applied. A reduction procedure has also been applied to samples with very high initial statistics. The aim of the procedure is to lower the processing time by reducing the number of preselected events in the sample while retaining enough statistics for meaningful comparisons between real and simulated data. The number of reduced events and the resulting luminosities are listed in Table 5 and Table 6. In cases where very large datasets were available, e.g. $Z \rightarrow ee$, only a subset of the full dataset was processed.

An important aspect of the samples is that they were prepared specifically for educational purposes. To this end, precision has been traded for simplicity of use. The simplifications are:

- No facilities to estimate systematic uncertainties have been included as these quickly introduce large complexities. This is of special importance as some variables may show discrepancies when only considering the statistical uncertainties, especially in high statistics analyses.
- Scale factors implementing corrections for electrons and muons are calculated using the preselection strategy of the AnalysisTop framework. This object selection does not have to coincide with the actual object selection defined by the user. Therefore, discrepancies may arise due to non-matching object definitions.
- The b-tagging scale factor is computed for a specific working point for a specific b-tagging algorithm (MV1@70% efficiency). The user, however, is free to specify the b-tagging weight used for tagging jets. This introduces a potential mismatch between real and simulated data because the working point and algorithm considered in the scale factor calculation differ from the ones being actually applied.
- No QCD samples were prepared as they would have been insufficient in statistics while introducing a large set of additional samples. The contributing effects of QCD may be countered using strict object definitions. However, analyses such as the W boson analysis may still suffer from the omission of these samples.
- The description of the W boson properties in simulated W+jets events is not ideal. The AnalysisTop framework provides scale factors to correct for these issues. These corrections are only available for samples produced with ALPGEN but not for those produced with SHERPA. However, using ALPGEN would have introduced a prohibitively large number of samples. Therefore, SHERPA was used although no corrections for the W boson modelling are provided for it.
- The simulated data take into account the pile-up and vertex position profile of the whole 2012 data taking although the real data is taken from a small list of runs from period D. This introduces a certain mismatch regarding the number of vertices and the primary vertex position.

1.2 Analysis Tools

The ATLAS open data dataset is accompanied by a set of analysis tools written in Python interfaced with ROOT [3]. These tools implement the protocols needed for reading the files, writing out histograms and plotting results. Ease of use and a clear structure of the tools is emphasised. Several example analyses are

electrons & muons	jets
$p_{\mathrm{T}} > 25$ GeV	$p_{\mathrm{T}} > 25$ GeV
ptconerel30 < 0.15	Jet Vertex Fraction cut
etconerel20 < 0.15	

Table 2: Standard selection on objects applied in the ATLAS open data tools. This standard selection is intended as a starting point for customised object selections implemented by the user. Unless stated otherwise this standard selection is applied in each of the example analyses. Jet Vertex Fraction (JVF) is a measure used to suppress jets from pp collisions additional to the primary pp collision. It uses information on the track to vertex association to evaluate which fraction of tracks associated with the jet stems from the primary vertex. Requirements are placed on the relative transverse energy isolation (etconerel20) and the relative transverse momentum isolation (ptconerel30).

provided and are intended to be starting points for further development. Full documentation on the tools is provided as a gitbook in an online resource.

2 Example Analyses

Example analyses and their results are shown in the following sections alongside the selection criteria specific to the analysis at hand. In most analyses a standard object selection, detailed in Table 2 is applied on top of the preselection detailed in Table 1 as described in Section 1.1. This standard object selection is intended as a starting point for a more optimised object selection and serves primarily as a common ground for the subsequent event selections of the individual analyses. In real data, the event is required to satisfy quality constraints defined in the Good Run List (GRL) to ensure only high quality data is used for physics measurements.

The purpose of these example analyses is to showcase the abilities and limitations of the real and simulated datasets included in the ATLAS open data release. These analyses are grouped as follows:

- Three **high statistics Standard Model analyses** have been implemented: a selection of events with one W boson decaying to leptons, a selection of a Z boson decaying to a lepton pair, and a selection of top quark pairs resulting in the final state $\ell\nu jjjj$. These analyses are intended to show that the general description of the data for these important Standard Model processes is sound. They also enable the study of Standard Model observables, such as the mass of the Z boson. Observable discrepancies between data and simulation are due to the simplified nature of the ntuples.
- Three **low statistics Standard Model analyses** are presented showing the limitations of the ATLAS open data dataset with respect to rarer processes. They are a WZ analysis, a ZZ analysis, and a $H \to WW$ analysis. Although it is still possible to obtain results in these analyses and achieve educational objectives, the statistical limitations prohibit more meaningful analyses. This point is particularly important as it demonstrates that the proposed datasets are intended for educational purposes only.
- A $\boldsymbol{Z' \to t\bar{t}}$ **analysis** serves as an example for a beyond the Standard Model (BSM) analysis. Multiple samples of simulated data containing Z' signal events are provided to implement a simplified analysis for searching for new physics.

The analysis plots in the following sections contain the ratio of real data to simulated Monte-Carlo data, to give an understanding of the quality of simulated data modelling. These are labelled Data/MC. In case two leptons are present in the final state they are ordered by transverse momentum with the leading one labelled “leading” and the subleading one labelled “trailing”.

The list of example analyses is not exhaustive. Further processes that may be explored include WW production, dileptonic top quark pair production, single top production, and many others.

2.1 W Analysis

This analysis is intended to provide an example for a high statistics analysis using the ATLAS open data dataset. Furthermore it tests the description of the real data by the simulated W boson data, which is the most limited process in terms of available Monte-Carlo statistics. An interesting variable to study would be the ratio W^+/W^- and its dependence on the pseudorapidity of the selected lepton. This would be a direct extension of the physics examined in the W-path of the ATLAS Masterclasses [4].

This analysis implements the criteria for single W boson events with the W boson decaying to leptons. It is based loosely on the charge asymmetry measurement carried out at $\sqrt{s} = 7$ GeV [5]. The standard object selection criteria (see Table 2) are applied. The event selection criteria are:

- Single electron or muon trigger is satisfied;
- Event in real data passes the Good Run List;
- Event has a good vertex ($N_{\text{tracks}} > 4$);
- Exactly one good lepton[1] with $p_{\text{T}} > 25$ GeV;
- $E_{\text{T}}^{\text{miss}} > 30$ GeV;
- $M_{\text{T}}^{W} > 30$ GeV.

The W analysis is potentially prone to QCD contributions as there is only one lepton present which may come from non-prompt sources mimicking the desired final state. Therefore, potential disagreements must always be understood as a sign that the QCD contributions are not taken into account. QCD samples are not provided as these have very low statistics after a selection while having a large file size.

The distributions of the transverse mass[2] as well as the missing transverse momentum shown in Figure 1 are affected by the omission of QCD contributions, which predominantly populate the low missing transverse momentum and low transverse mass regions. A comparison of results obtained here to those of W+jets analyses considering the impact of QCD processes supports this explanation [6].

The histograms depicting the vertex information in Figure 2 show the expected disagreement between simulated and real data. The pile-up treatment in simulated data considered the whole 2012 run period whilst the real data is taken only from period D of the 2012 data taking.

The overall description of the lepton kinematics by the simulated data is good as can be seen in Figure 3. The figure also depicts the type of lepton expressed using the absolute value of the PDG id [7]. Electrons/positrons have a PDG id of 11/-11 whereas muons/antimuons are denoted with a PDG id of 13/-13. Less well described are the tracking and isolation variables shown in Figure 4. Here, the rise of the ratio between data and simulation at higher isolation values suggests that QCD contributions are missing. In this region QCD processes would contribute by either the misidentification of a jet as a lepton or by a hadron decay to leptons inside a jet. These so-called non-prompt leptons are not well isolated resulting in higher values for the isolation variables shown.

Figure 5 depicts the kinematics, jet vertex fraction, and the MV1 b-tagging weight of the selected jets. Here, a slightly larger normalisation offset between real and simulated data is observed. This again may be attributed to missing QCD contributions as the one jet bin would most likely be populated by dijet

[1] When describing selections, lepton refers to an electron or muon candidate.

[2] The transverse mass is defined as: $m_{\text{T}} = \sqrt{2p_{\text{T}}^{\ell}E_{\text{T}}^{\text{miss}} \times \left[1 - \cos\left(\Delta\phi\left(\ell, E_{\text{T}}^{\text{miss}}\right)\right)\right]}$.

events, where one jet is either misidentified as the lepton or supplies a non-prompt lepton and the other counts towards the jet multiplicity. Apart from the normalisation issue the variables are reasonably well described by the simulated data.

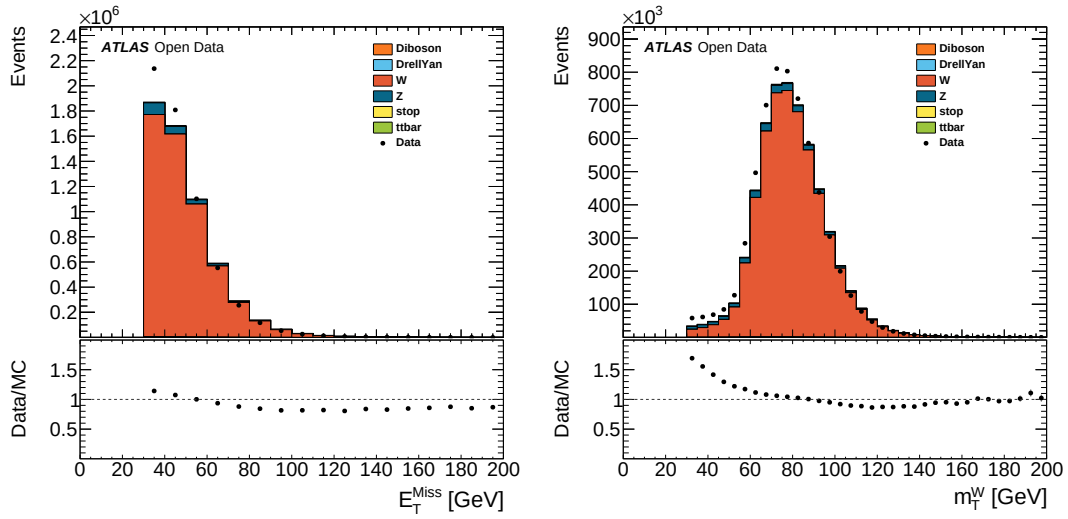

Figure 1: W Analysis: Event variable histograms. The variables plotted are the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ and the transverse mass of the W boson candidate m_{T}^{W}.

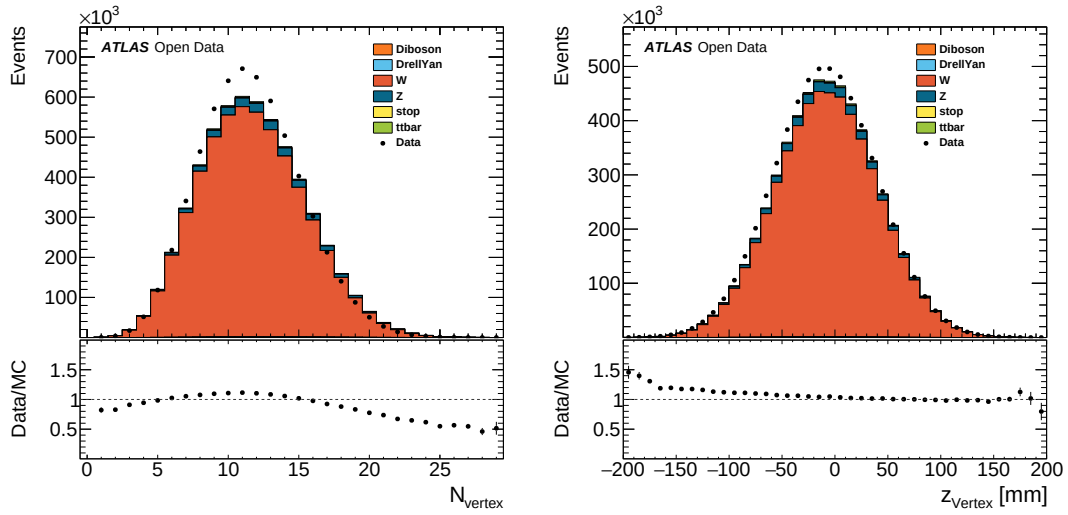

Figure 2: W Analysis: Vertex histograms. The number of vertices N_{vertex} and the z coordinate of the primary vertex z_{vertex} are shown.

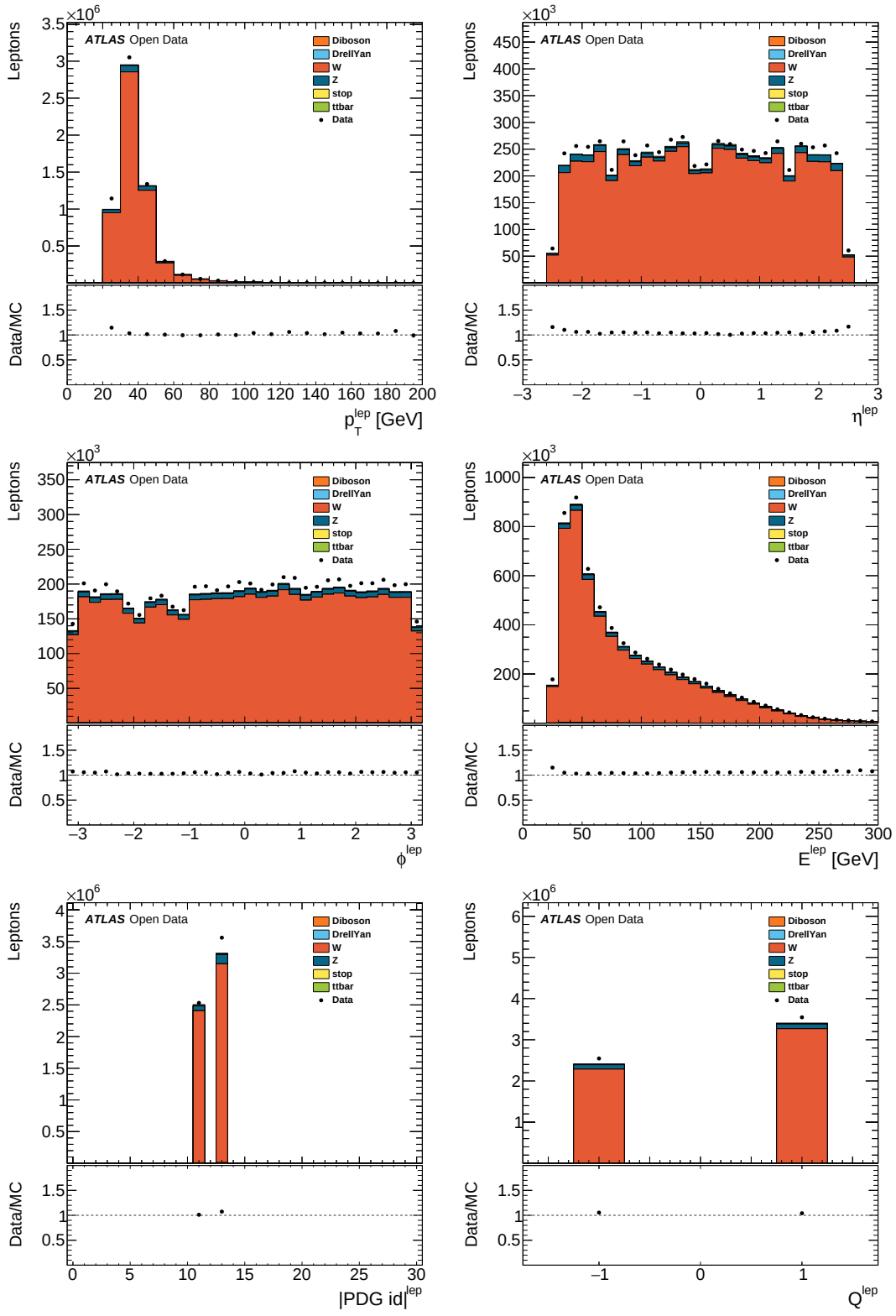

Figure 3: *W* Analysis: Leading lepton properties. From upper left to lower right are shown: transverse momentum p_{T}, pseudorapidity η, azimuthal angle ϕ, energy E, absolute value of the PDG id |PDG id|, and charge Q.

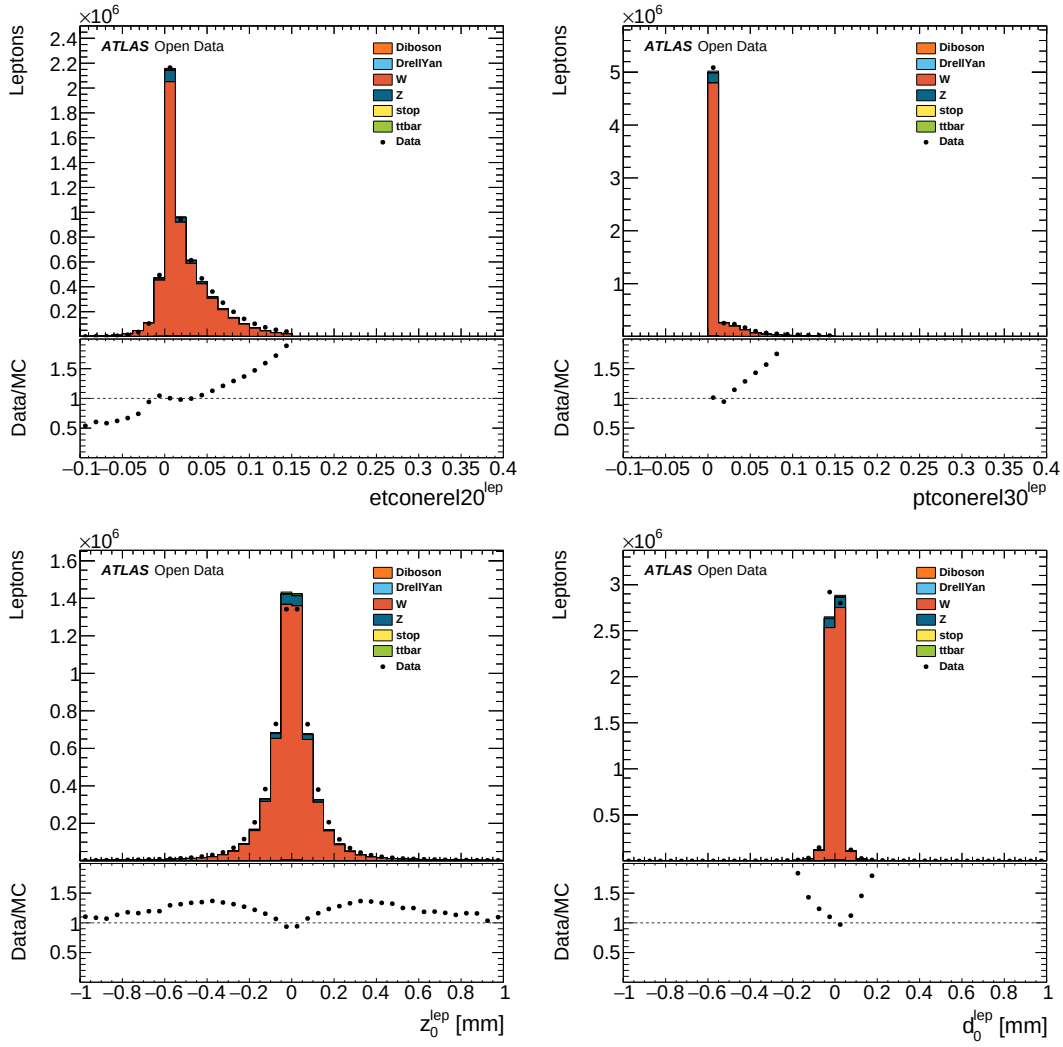

Figure 4: *W* Analysis: Leading lepton isolation and tracking information. From upper left to lower right are shown: relative transverse energy isolation (etconerel20), relative transverse momentum isolation (ptconerel30), longitudinal impact parameter z_0, and transverse impact parameter d_0.

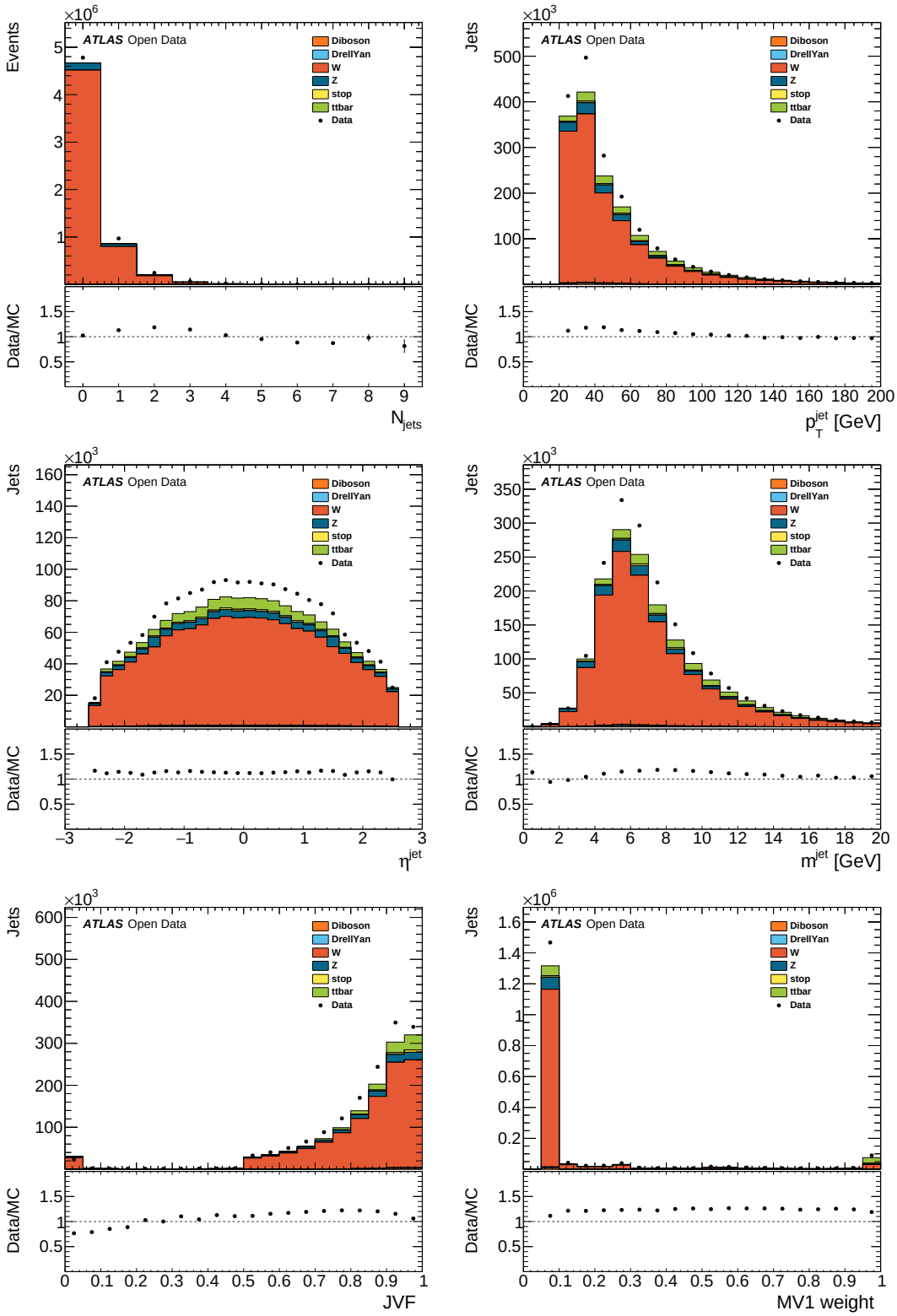

Figure 5: *W* Analysis: Jet properties. From upper left to lower right are shown: Jet multiplicity N_{jets}, transverse momentum p_{T}, pseudorapidity η, mass m, jet vertex fraction (JVF) and MV1 b-tagging weight of the selected jets.

2.2 Z Analysis

Many analyses selecting leptons suffer from Z+jets as a contributing background due to its large production cross section. It is therefore vital to check the correct modelling of this process by the simulated data. In addition, the exercises of the ATLAS Z-path Masterclasses [8] may be extended by this example analysis.

The Z boson analysis implemented here considers Z boson decays into an electron positron or muon antimuon pair. The standard object selection criteria (see Table 2) are applied. The event selection criteria are:

- Single electron or muon trigger is satisfied;
- Event in real data passes the Good Run List;
- Event has a good vertex ($N_{\text{tracks}} > 4$);
- Exactly two good leptons with $p_{\text{T}} > 25$ GeV;
- Leptons have opposite charge;
- Leptons have same flavour;
- $|m_{\ell\ell} - m_Z| < 20$ GeV with $m_Z = 91.18$ GeV.

The modelling of the lepton kinematics by the simulated data is very good, as can be seen in Figures 6 and Figure 8. The isolation and tracking information shown in Figure 7 and Figure 9 are also described well. Figure 10 summarises the jet information. The jet multiplicity shows a slight disagreement in the higher jet bins, which would be covered by systematic uncertainties. This leads to a slight disagreement in normalisation for the jet histograms. Nonetheless, the description of the real data by the simulated data is very good when comparing shapes.

The histograms depicting the vertex information in Figure 11 show the expected disagreement between simulated and real data. The pile-up treatment in simulated data considers the whole 2012 run period whilst the real data is taken only from period D of the 2012 data taking. Figure 12 depicts the invariant mass of the reconstructed Z boson candidate which shows excellent agreement between real and simulated data. The poor modelling of $E_{\text{T}}^{\text{miss}}$ is due to the complexities of simulating missing transverse momentum in the absence of an actual neutrino from the hard scattering.

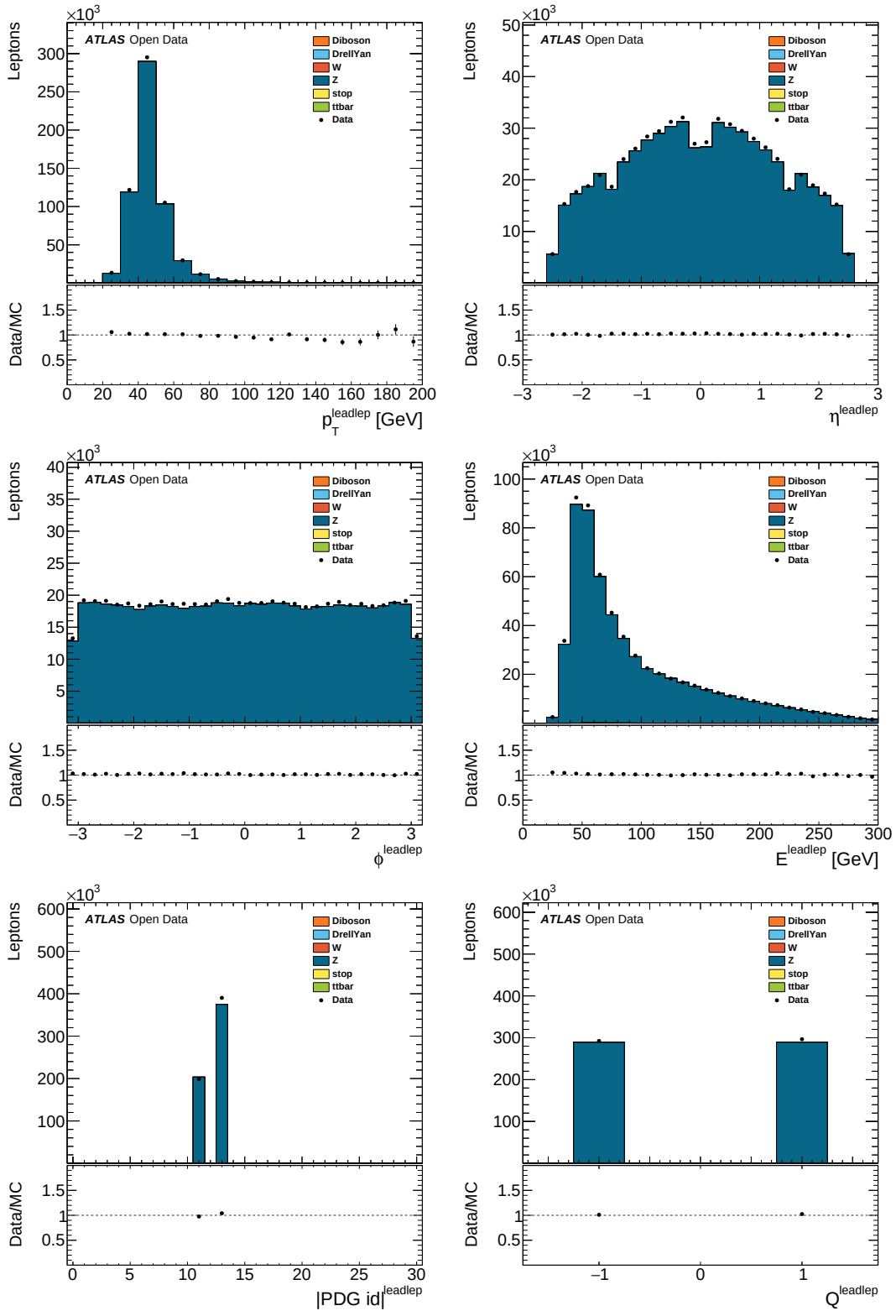

Figure 6: Z Analysis: Leading lepton properties. From upper left to lower right are shown: transverse momentum p_{T}, pseudorapidity η, azimuthal angle ϕ, energy E, absolute value of the PDG id |PDG id|, and charge Q.

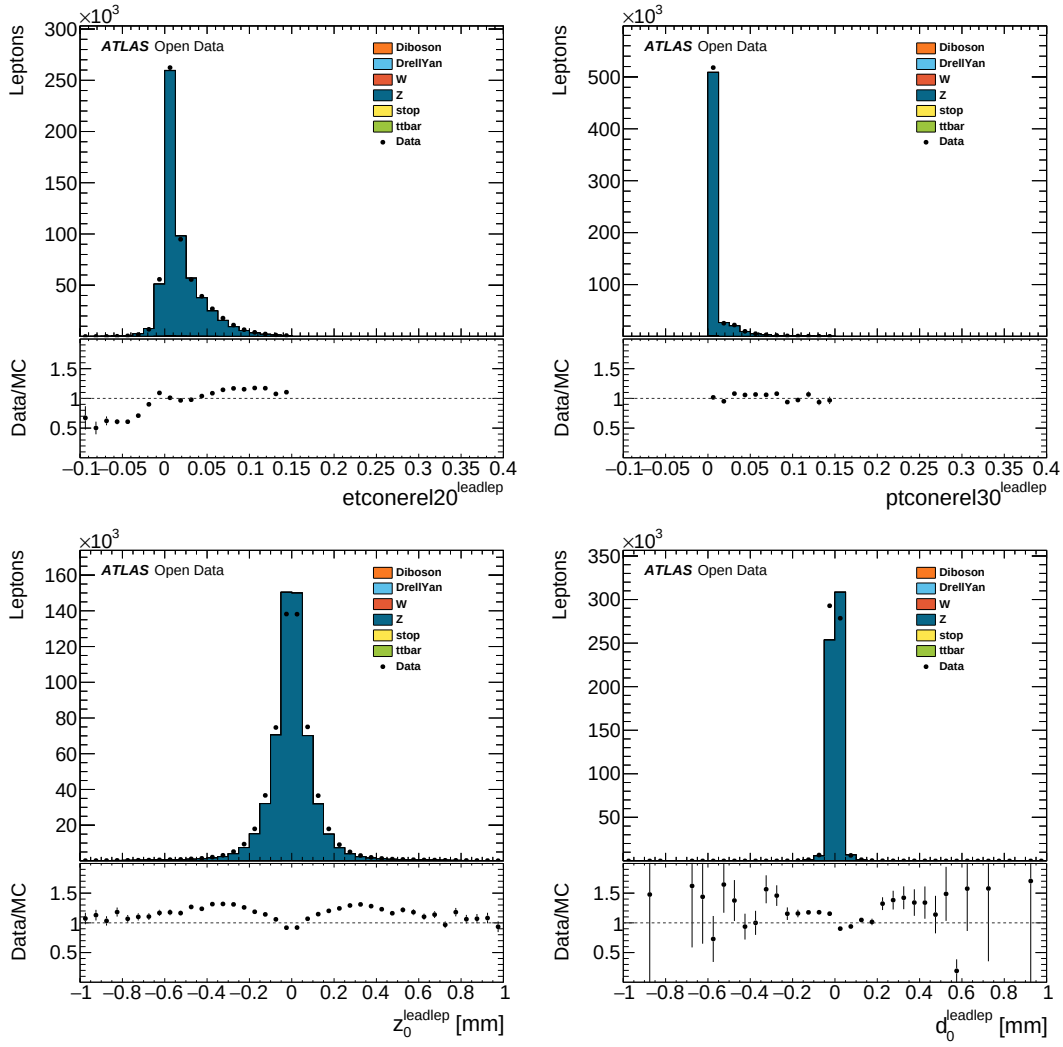

Figure 7: Z Analysis: Leading lepton isolation and tracking information. From upper left to lower right are shown: relative transverse energy isolation (etconerel20), relative transverse momentum isolation (ptconerel30), longitudinal impact parameter z_0, and transverse impact parameter d_0.

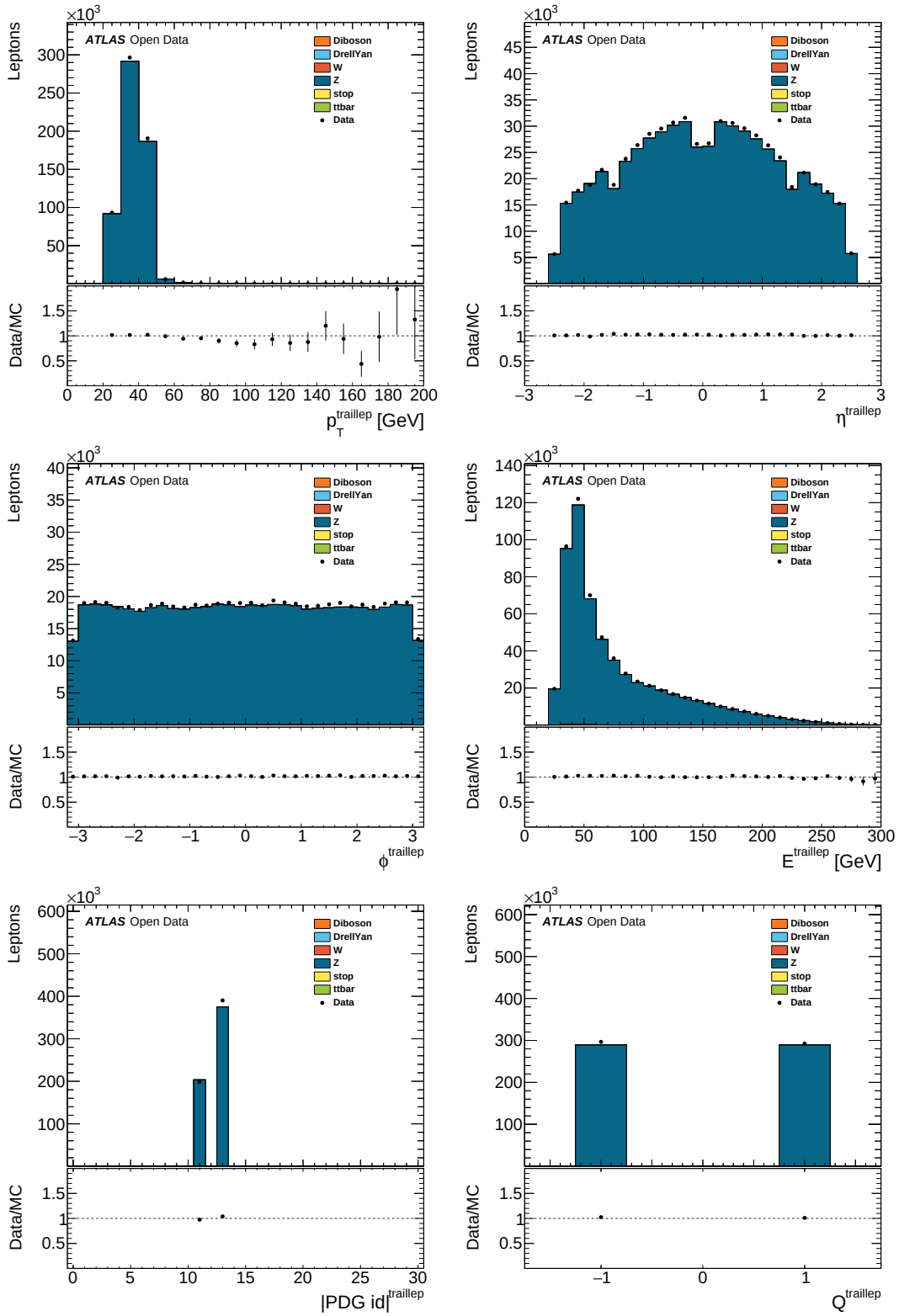

Figure 8: Z Analysis: Trailing lepton properties. From upper left to lower right are shown: transverse momentum p_{T}, pseudorapidity η, azimuthal angle ϕ, energy E, absolute value of the PDG id |PDG id|, and charge Q.

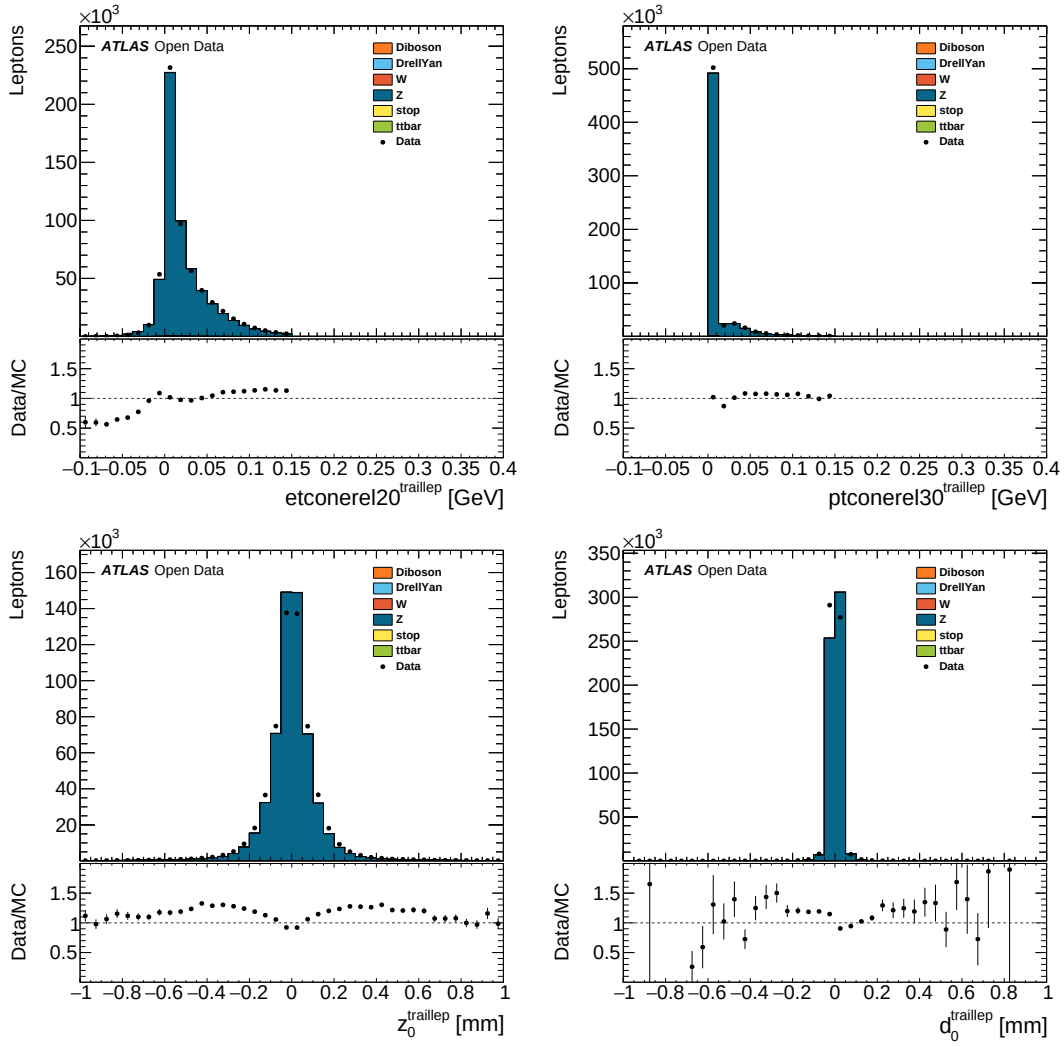

Figure 9: Z Analysis: Trailing lepton isolation and tracking information. From upper left to lower right are shown: relative transverse energy isolation (etconerel20), relative transverse momentum isolation (ptconerel30), longitudinal impact parameter z_0, and transverse impact parameter d_0.

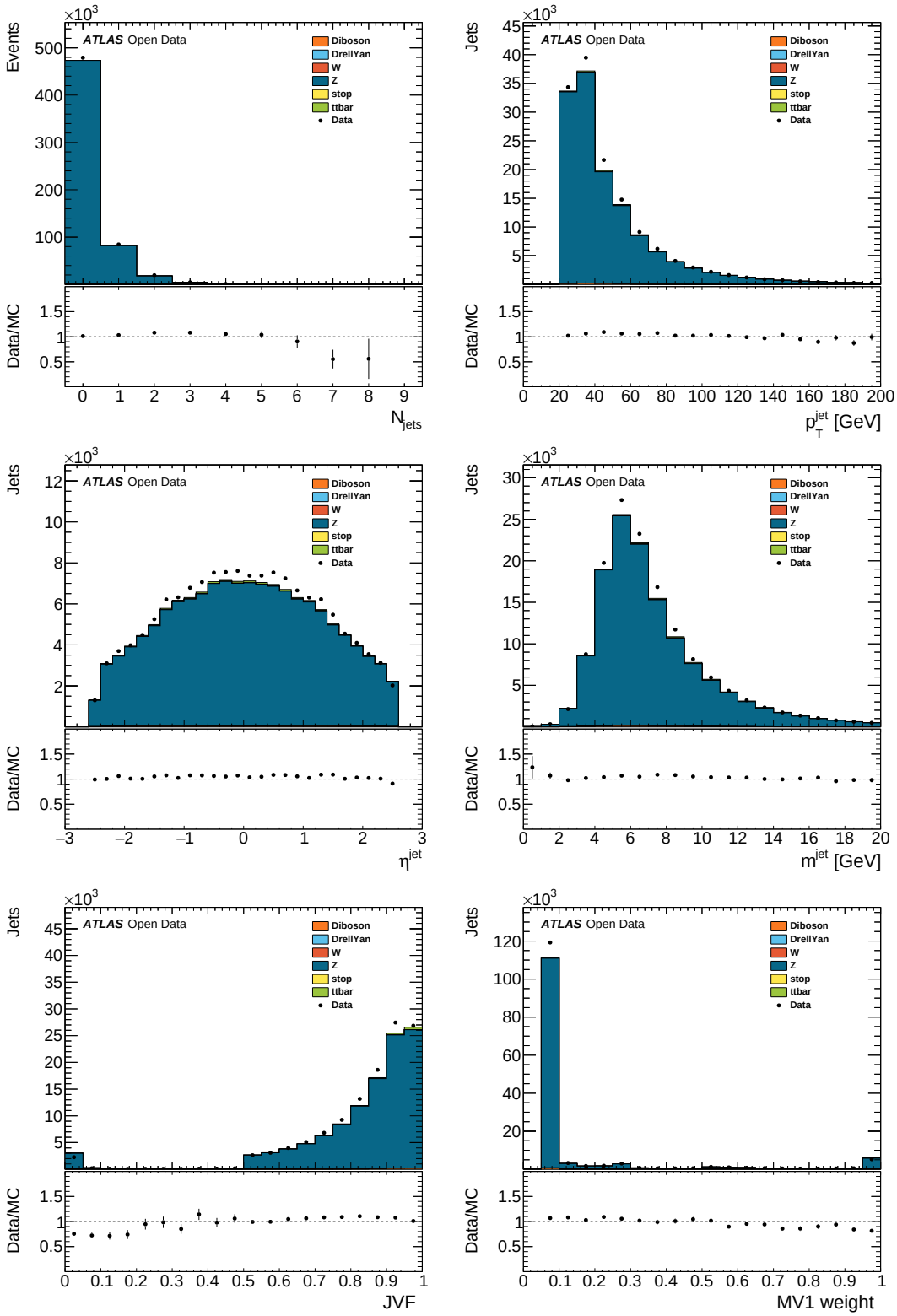

Figure 10: Z Analysis: Jet properties. From upper left to lower right are shown: Jet multiplicity N_{jets}, transverse momentum p_{T}, pseudorapidity η, mass m, jet vertex fraction (JVF) and MV1 b-tagging weight of the selected jets.

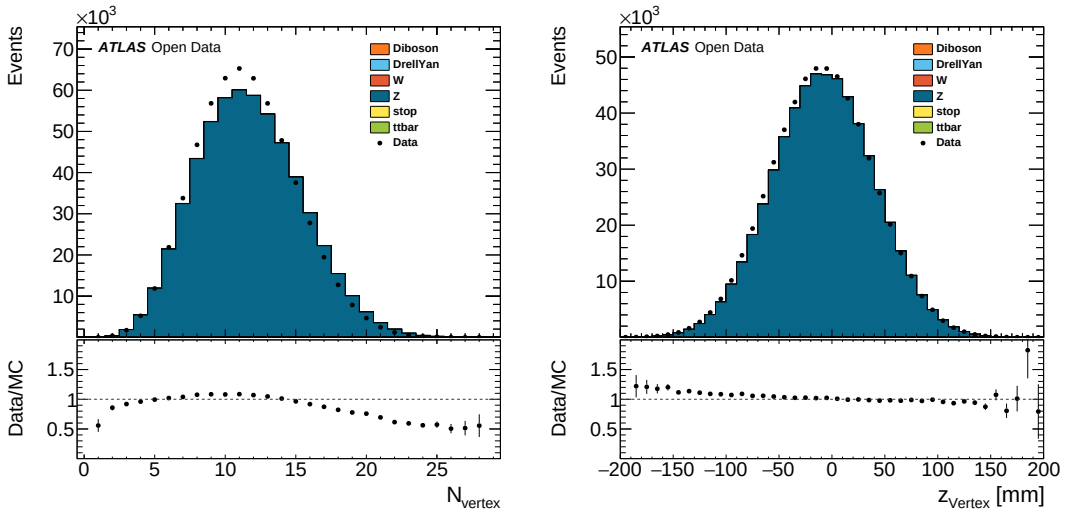

Figure 11: Z Analysis: Vertex histograms. The number of vertices N_{vertex} and the z coordinate of the primary vertex z_{vertex} are shown.

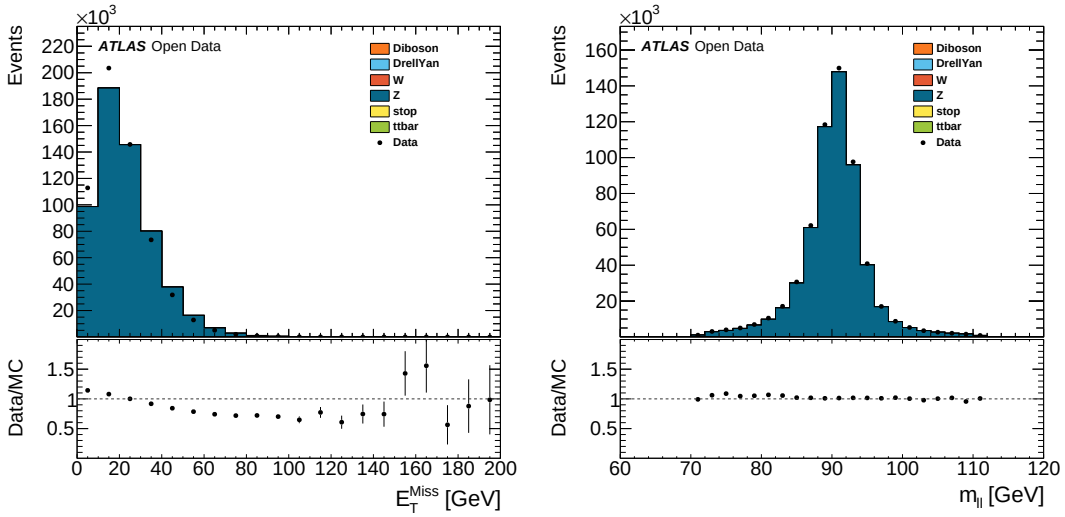

Figure 12: Z Analysis: Event variable histograms. The variables plotted are the missing transverse momentum $E_{\text{T}}^{\text{miss}}$ and the invariant mass of the Z boson candidate $m_{\ell\ell}$.

2.3 Top Quark Pair Analysis

The LHC is a top quark factory and studying and understanding top quark physics is one of the major goals of the ATLAS physics programme. This understanding is crucial for studying rarer processes as top quark pair production is a background to virtually all processes having leptons and multiple jets in their final states. Top quark pair production can be studied in the ATLAS open data dataset in both the semileptonic and dileptonic final state. Statistics are expected to be sufficient for producing detailed distributions and exploring advanced techniques like the reconstruction of the top quark pair system.

This analysis mimics a standard top quark pair selection in the semileptonic channel. The standard object selection criteria (see Table 2) are applied. The event selection is defined as:

- Single electron or muon trigger is satisfied;
- Event in real data passes the Good Run List;
- Event has a good vertex ($N_{\text{tracks}} > 4$);
- Exactly one good lepton with $p_{\text{T}} > 25$ GeV;
- At least four good jets;
- At least two b-tagged jets (MV1@70%);
- $E_{\text{T}}^{\text{miss}} > 30$ GeV;
- $m_{\text{T}}^{W} > 30$ GeV.

There is a small disagreement in normalisation of approximately 5 % between simulated and real data. This may be attributed to the fact that the b-tagging scale factor is not applied despite two b-tags being required. A survey of the shapes of all presented histograms reveals no obvious discrepancies.

Figure 13 depicts the lepton kinematics, type, and charge of the leptons with the simulated data reproducing the real data well. Tracking and isolation related variables (see Figure 14) are not as well described but show reasonable agreement in the high statistics regions. Figure 15 summarises the jet properties. Jet multiplicity, jet kinematics, jet vertex fraction, and MV1 b-tagging weight are well described by the simulated data.

The histograms depicting the vertex information in Figure 16 show the expected disagreement between simulated and real data. The pile-up treatment in simulated data considers the whole 2012 run period whilst the real data is taken only from period D of the 2012 data taking. The transverse mass of the W boson candidate and the missing transverse momentum are shown in Figure 17. Both exhibit a flat ratio between simulated and real data, indicating that no apparent mismodelling is present.

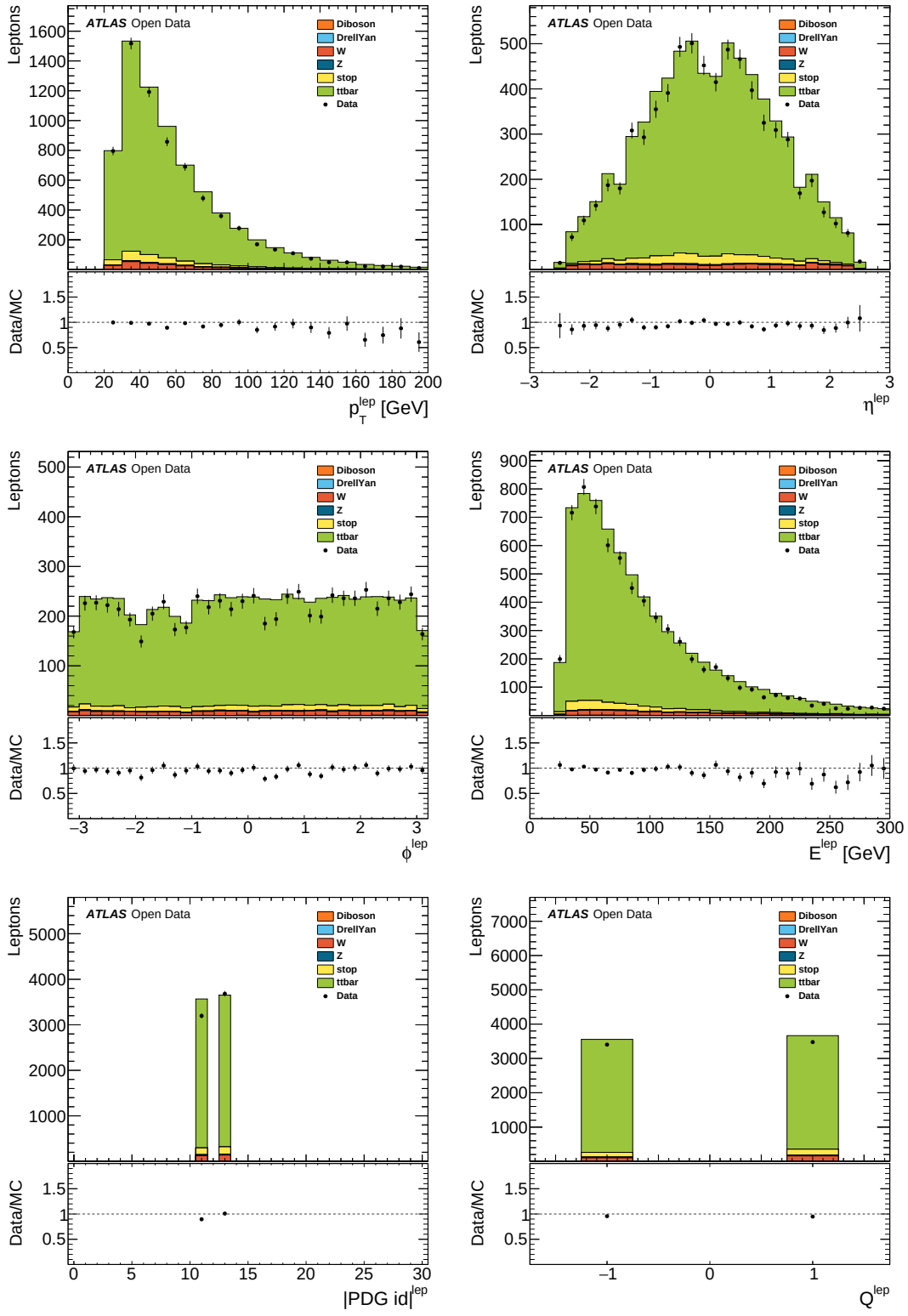

Figure 13: $t\bar{t}$ Analysis: Leading lepton properties. From upper left to lower right are shown: transverse momentum p_{T}, pseudorapidity η, azimuthal angle ϕ, energy E, absolute value of the PDG id |PDG id|, and charge Q.

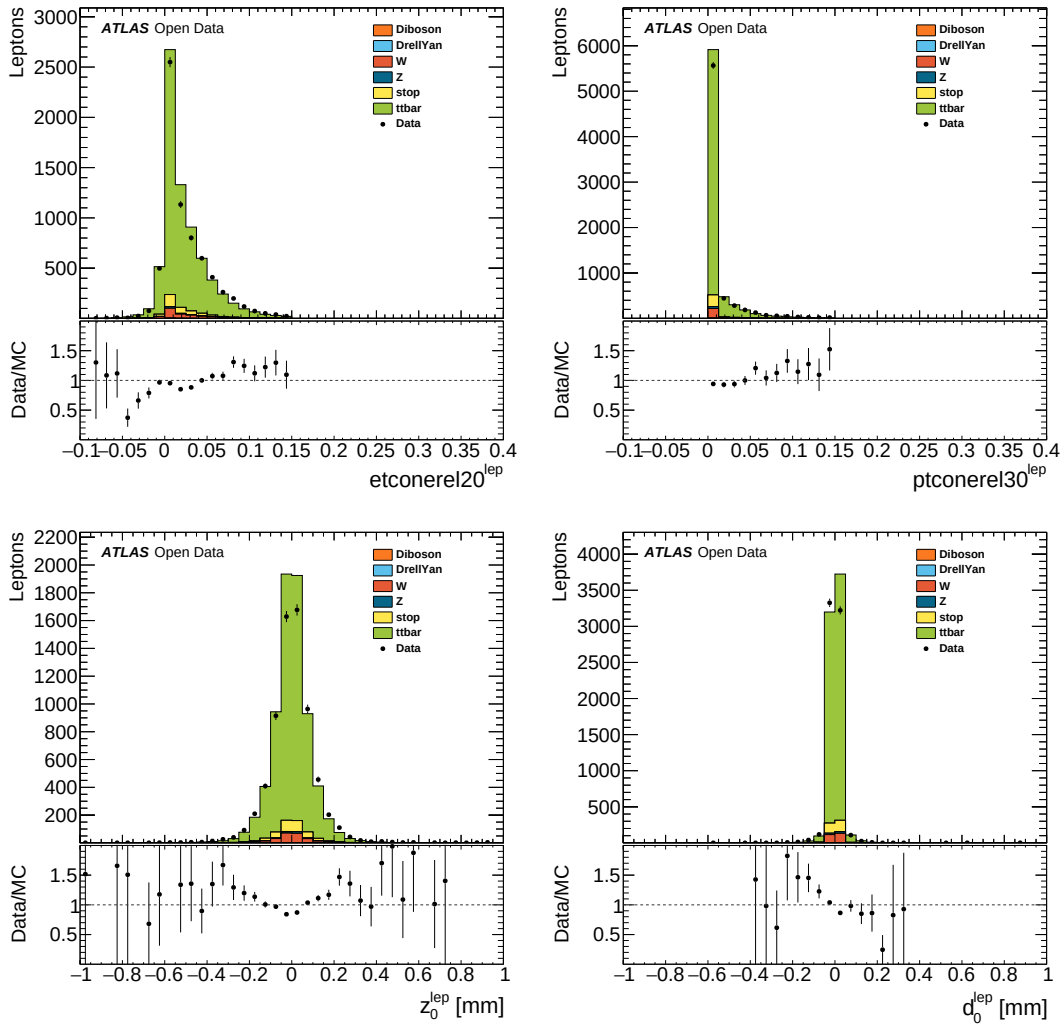

Figure 14: $t\bar{t}$ Analysis: Leading lepton isolation and tracking information. From upper left to lower right are shown: relative transverse energy isolation (etconerel20), relative transverse momentum isolation (ptconerel30), longitudinal impact parameter z_0, and transverse impact parameter d_0.

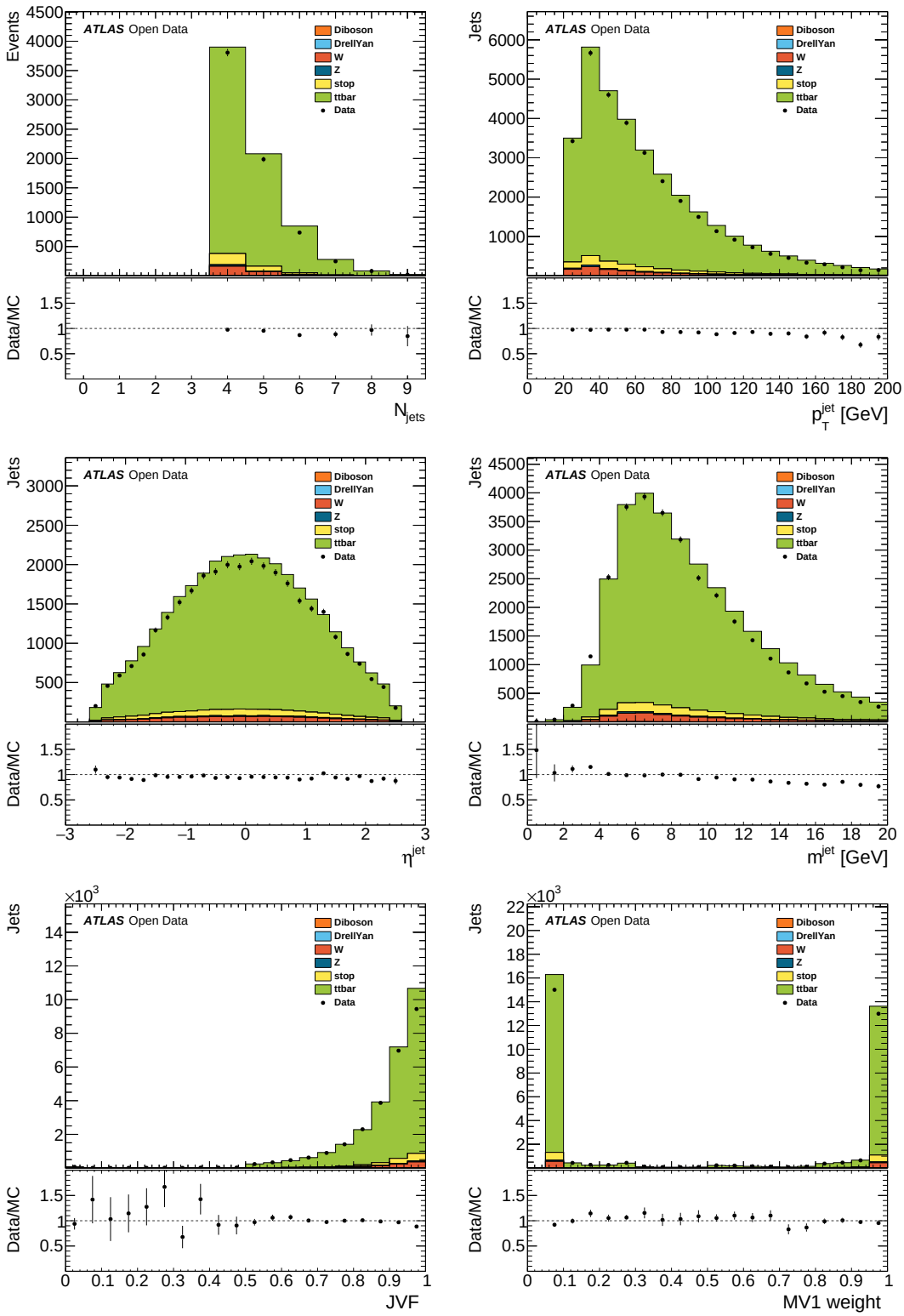

Figure 15: $t\bar{t}$ Analysis: Jet properties. From upper left to lower right are shown: Jet multiplicity N_{jets}, transverse momentum p_{T}, pseudorapidity η, mass m, jet vertex fraction (JVF) and MV1 b-tagging weight of the selected jets.

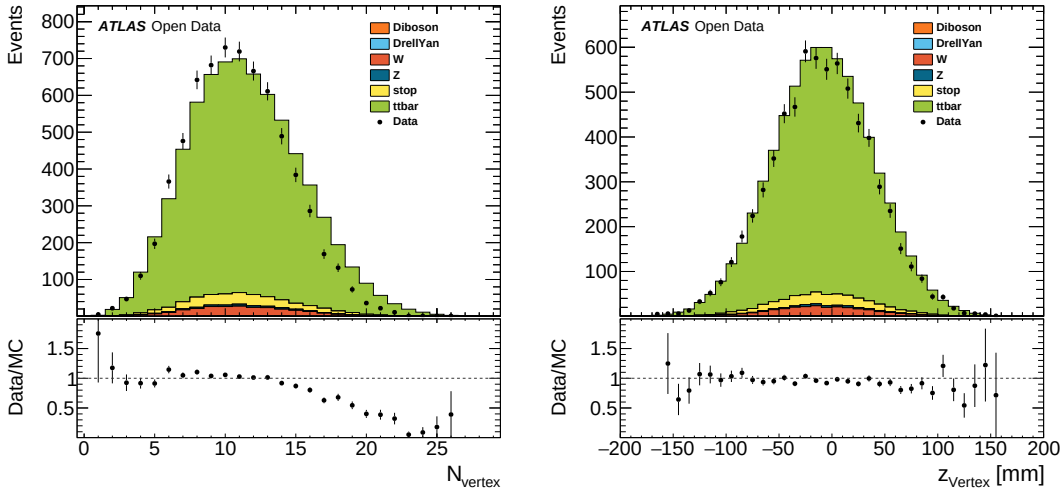

Figure 16: $t\bar{t}$ Analysis: Vertex histograms. The number of vertices N_{vertex} and the z coordinate of the primary vertex z_{vertex} are shown.

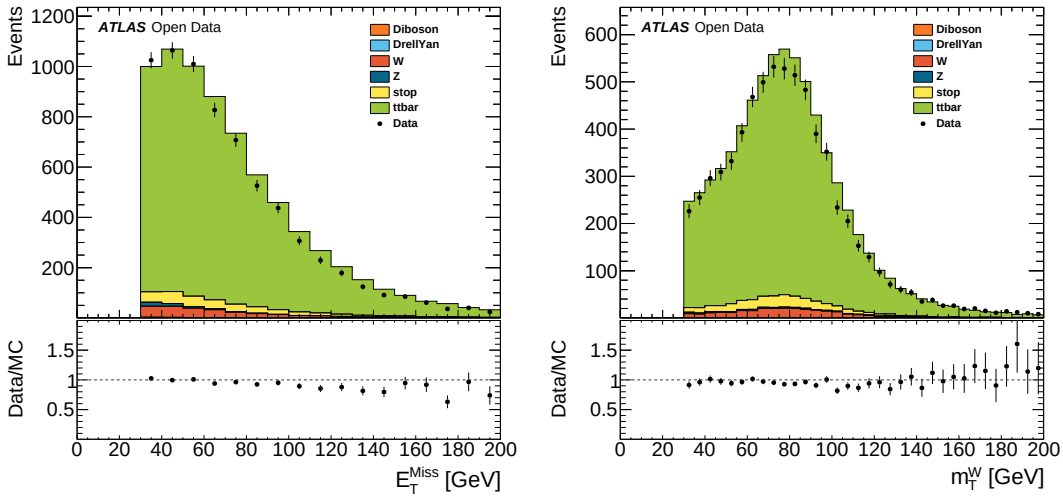

Figure 17: $t\bar{t}$ Analysis: Event variable histograms. The variables plotted are the missing transverse momentum $E_{\text{T}}^{\text{miss}}$ and the transverse mass of the W boson candidate m_{T}^{W}.

2.4 WZ Analysis

Diboson physics is an important part of the physics programme of ATLAS as it is a probe for electroweak physics. It enables tests of key predictions of the electroweak theory like the self-coupling of the electroweak gauge bosons. The WZ analysis was chosen as an example analysis for the ATLAS open data tools. It is one of the most abundantly produced diboson processes and has a clean final state consisting of three charged leptons and a neutrino. Reconstructing the WZ system and studying its properties is possible, but introduces a slight challenge due to the neutrino which may be seen as a interesting educational challenge. The available statistics in the ATLAS open data dataset allows for a rediscovery of the WZ process in a lab course.

This analysis is abridged from the WZ analysis in the fully leptonic channel as it is carried out by ATLAS using the 2012 dataset [9]. The selected phase space is the one used for extracting the production cross section of WZ. Although events fitting the WZ final state are present, their number is too small to draw stringent conclusions in terms of real to simulated data agreement. The standard object selection criteria (see Table 2) are applied. The event selection criteria are:

- Single electron or muon trigger is satisfied;
- Event in real data passes the Good Run List;
- Event has a good vertex ($N_{\text{tracks}} > 4$);
- Exactly three good leptons with $p_{\text{T}} > 25$ GeV;
- WZ candidate is chosen by finding the Z boson candidate closest to the nominal Z mass;
- $|m_{\ell\ell} - m_Z| < 10$ GeV with $m_Z = 91.18$ GeV;
- $m_{\text{T}}^W > 30$ GeV.

The WZ signal is shown independently from the other diboson processes (WW and ZZ) in the figures in this section. The overall description of the real data by the simulated data is reasonably good given the statistical limitations. The kinematics of the three leptons are summarised in Figure 18 and show good agreement. The isolation and tracking variables depicted in Figure 19 are equally well reproduced by the simulated data. Due to the low statistics no jet histograms have been included as the expected yields would be too low for a meaningful comparison. The vertex information depicted in Figure 20 as well as the invariant mass of the Z boson candidate and the transverse mass of the W boson candidate shown in Figure 21 do not show any major mismodelling.

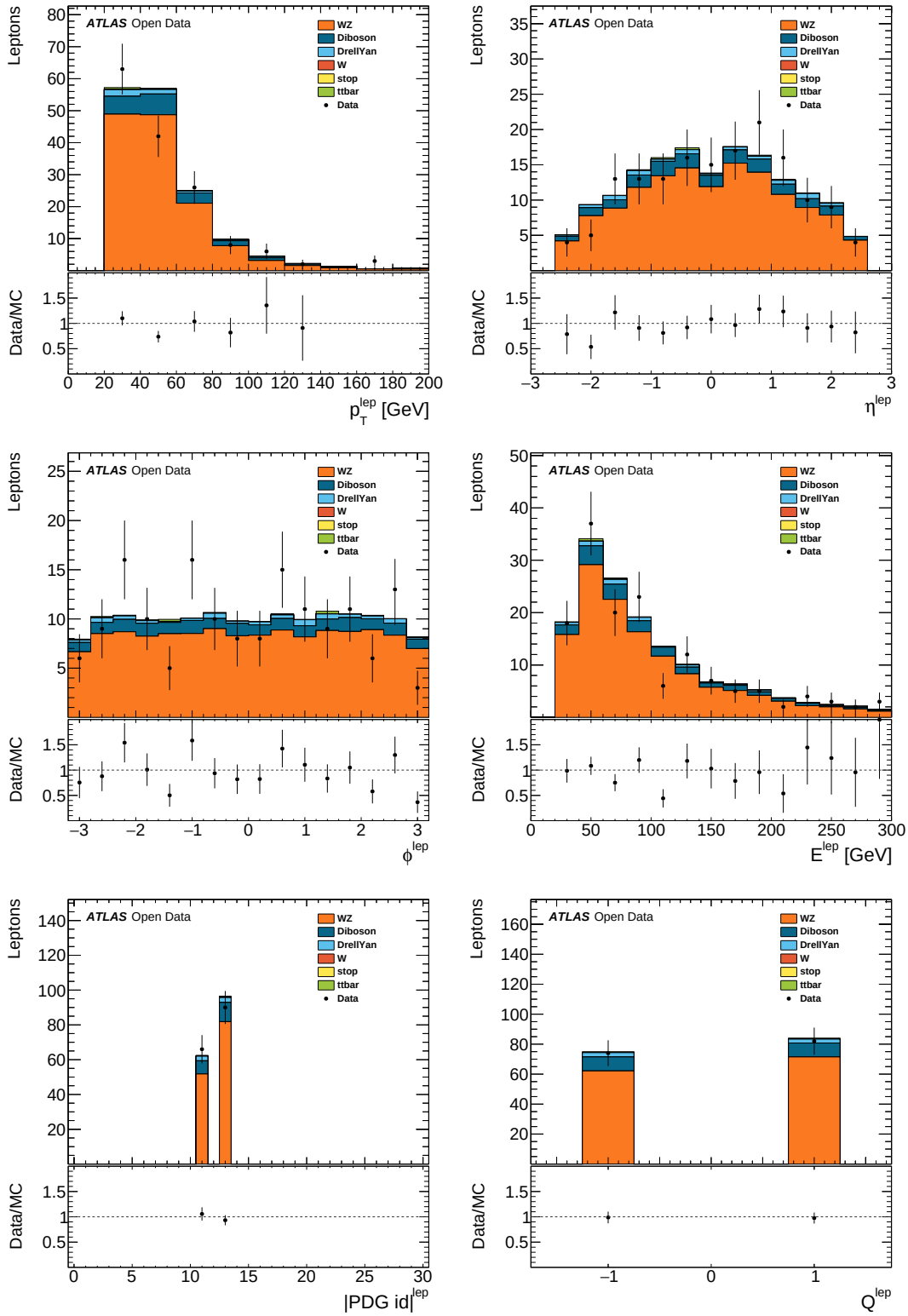

Figure 18: WZ Analysis: Lepton properties. From upper left to lower right are shown: The transverse momentum p_{T}, pseudorapidity η, azimuthal angle ϕ, energy E, absolute value of the PDG id |PDG id|, and charge Q of the three leptons in the selected events.

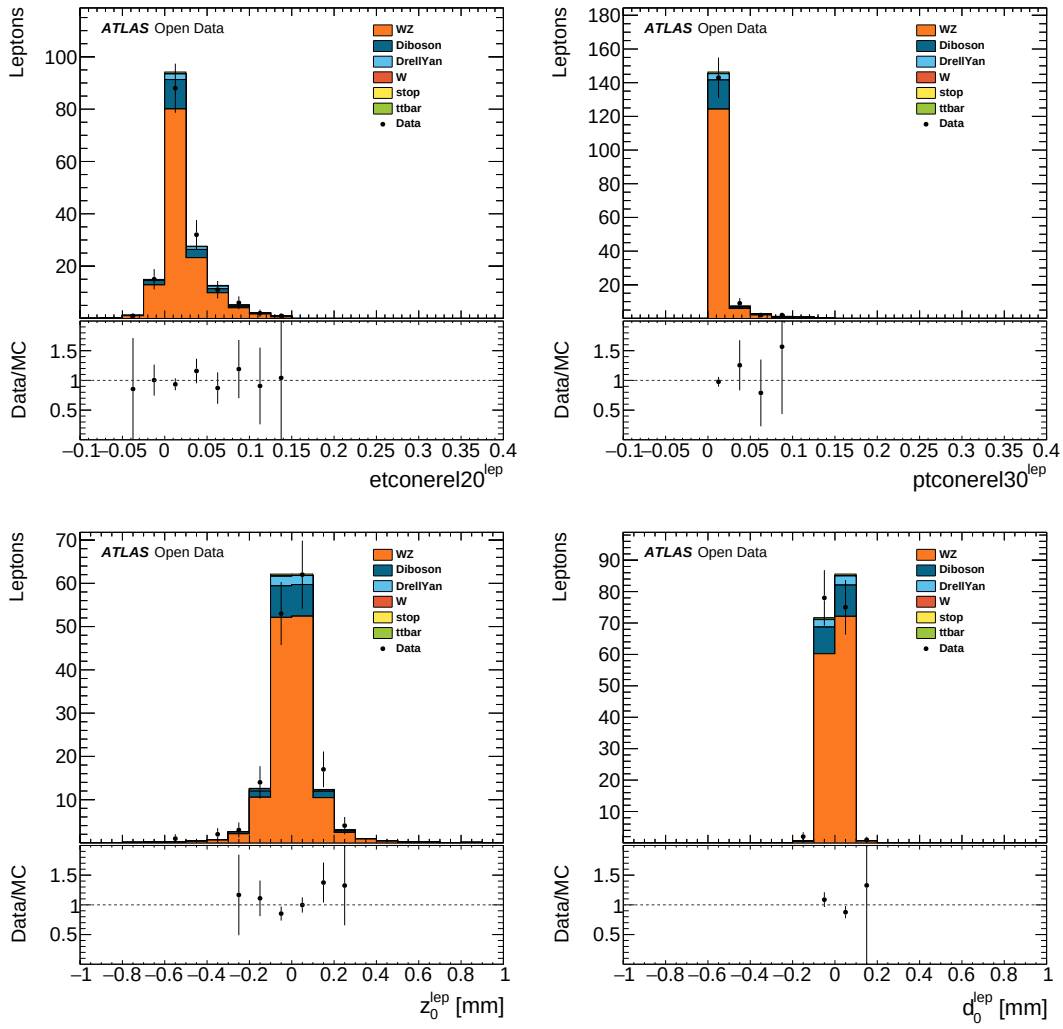

Figure 19: WZ Analysis: Lepton isolation and tracking information. From upper left to lower right are shown: relative transverse energy isolation (etconerel20), relative transverse momentum isolation (ptconerel30), longitudinal impact parameter z_0, and transverse impact parameter d_0 of the three leptons in the selected events.

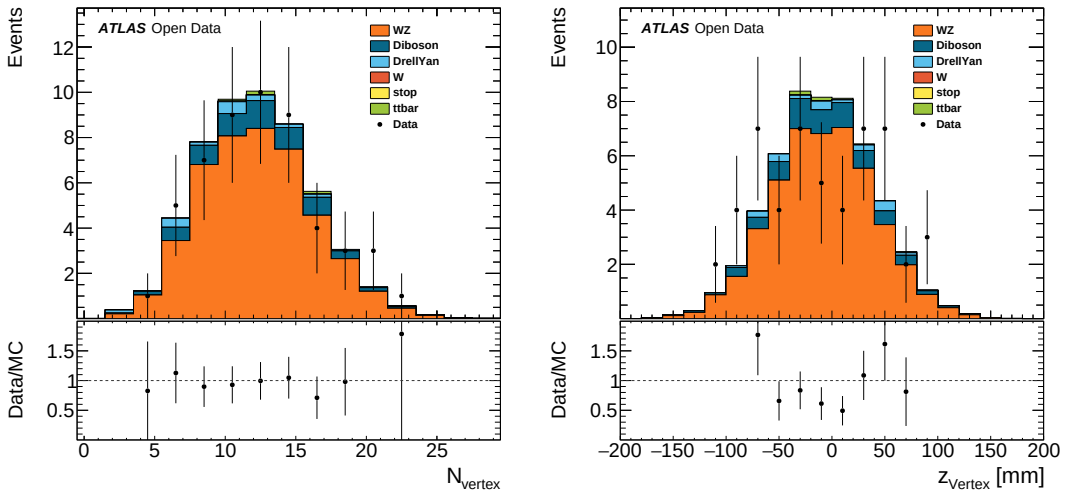

Figure 20: WZ Analysis: Vertex histograms. The number of vertices N_{vertex} and the z coordinate of the primary vertex z_{vertex} are shown.

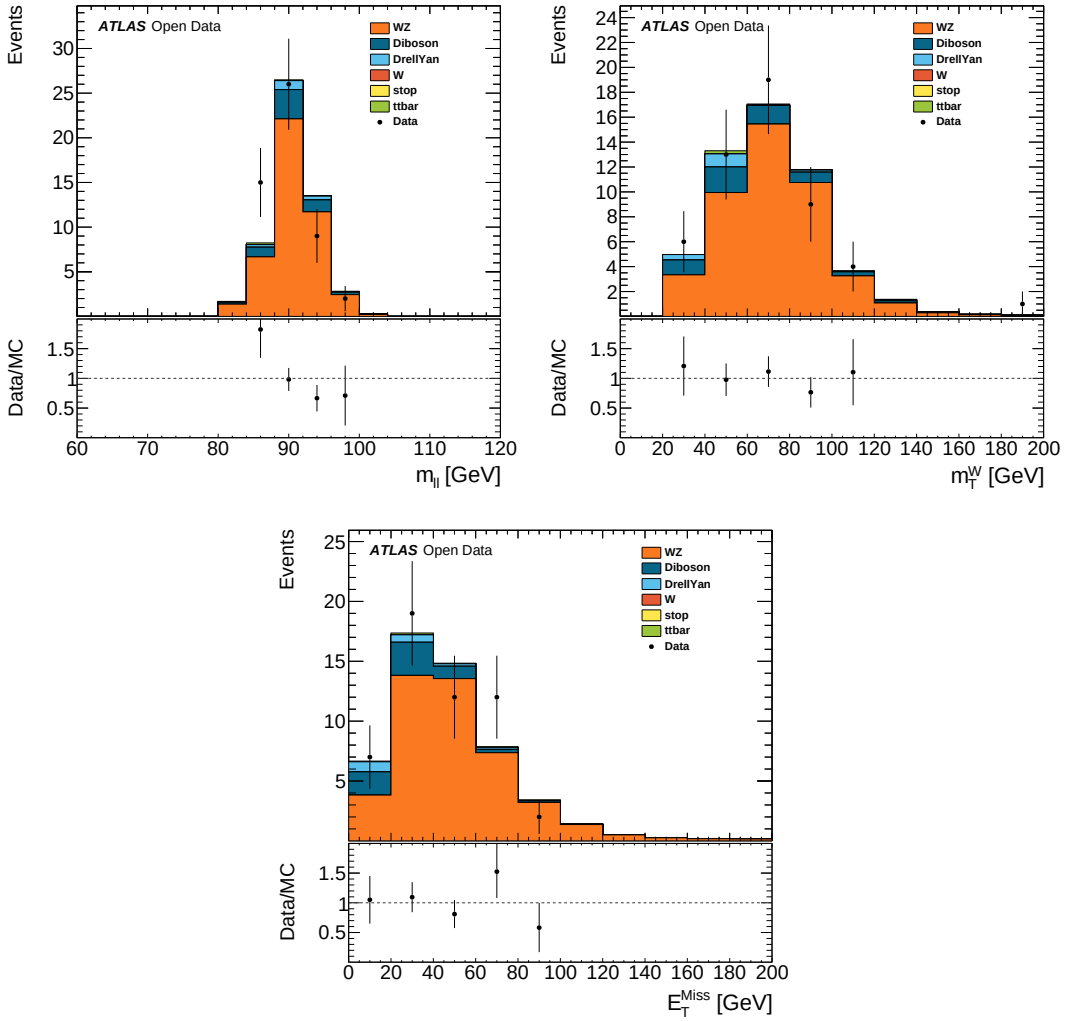

Figure 21: WZ Analysis: Event variable histograms. The variables plotted are invariant mass of the Z boson candidate $m_{\ell\ell}$, the transverse mass of the W boson candidate m_{T}^{W}, and the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$.

2.5 ZZ Analysis

The production of ZZ with subsequent decay to leptons is the dominant Standard Model process with four charged prompt leptons in the final state. Its low production cross section results in a very low yield for the ATLAS open data dataset and its highlights the statistical limitations. Although some events can be selected, the low event yield prohibits detailed analysis and conclusions drawn are rather qualitative in nature.

The ZZ analysis implemented in the ATLAS open data tools selects events where both Z bosons decay to leptons. It is based on the ZZ production cross section measurement carried out at $\sqrt{s} = 7$ GeV [10]. The standard object selection criteria (see Table 2) are applied with a loosened lepton p_{T} requirement of $p_{\mathrm{T}} > 10$ GeV. The event selection criteria are:

- Single electron or muon trigger is satisfied;
- Event in real data passes the Good Run List;
- Event has a good vertex ($N_{\mathrm{tracks}} > 4$);
- Exactly four good leptons with $p_{\mathrm{T}} > 10$ GeV;
- Two Z candidates built from lepton pairs of same flavour and opposite charge minimising the total deviation of both candidates from the Z boson mass;
- $|m_{\mathrm{ZCand1}} - m_Z| + |m_{\mathrm{ZCand2}} - m_Z| < 20$ GeV with $m_Z = 91.18$ GeV.

The ZZ signal is shown independently from the other diboson processes (WW and WZ) in the figures in this section. The event yields in this analysis are particularly low, as can be seen in Figures 22 to 25, and no stringent statement regarding the quality of the description of the real data by the simulated data can be made. The histograms are included solely for illustrative purposes.

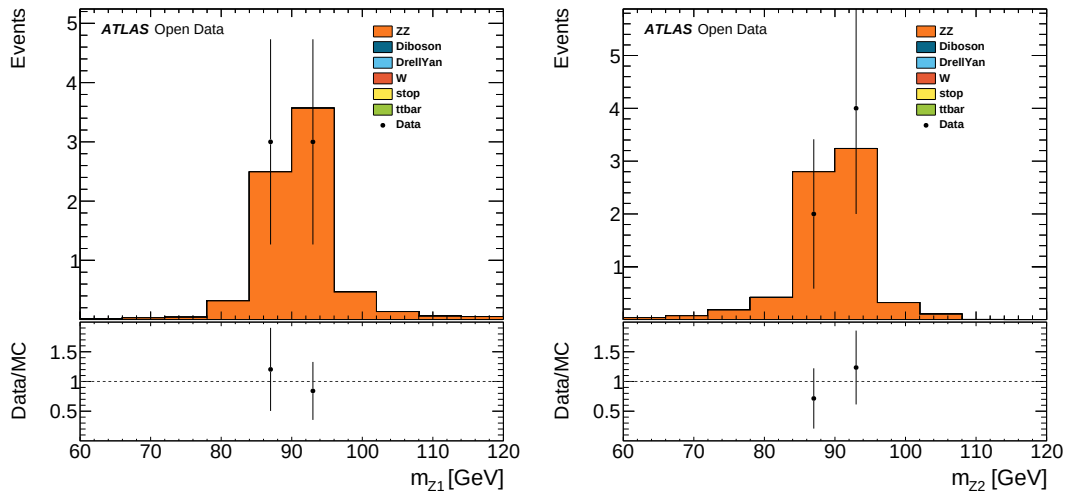

Figure 22: ZZ Analysis: Event variable histograms. The variables plotted are invariant masses of the two Z boson candidates m_{z1} and m_z2.

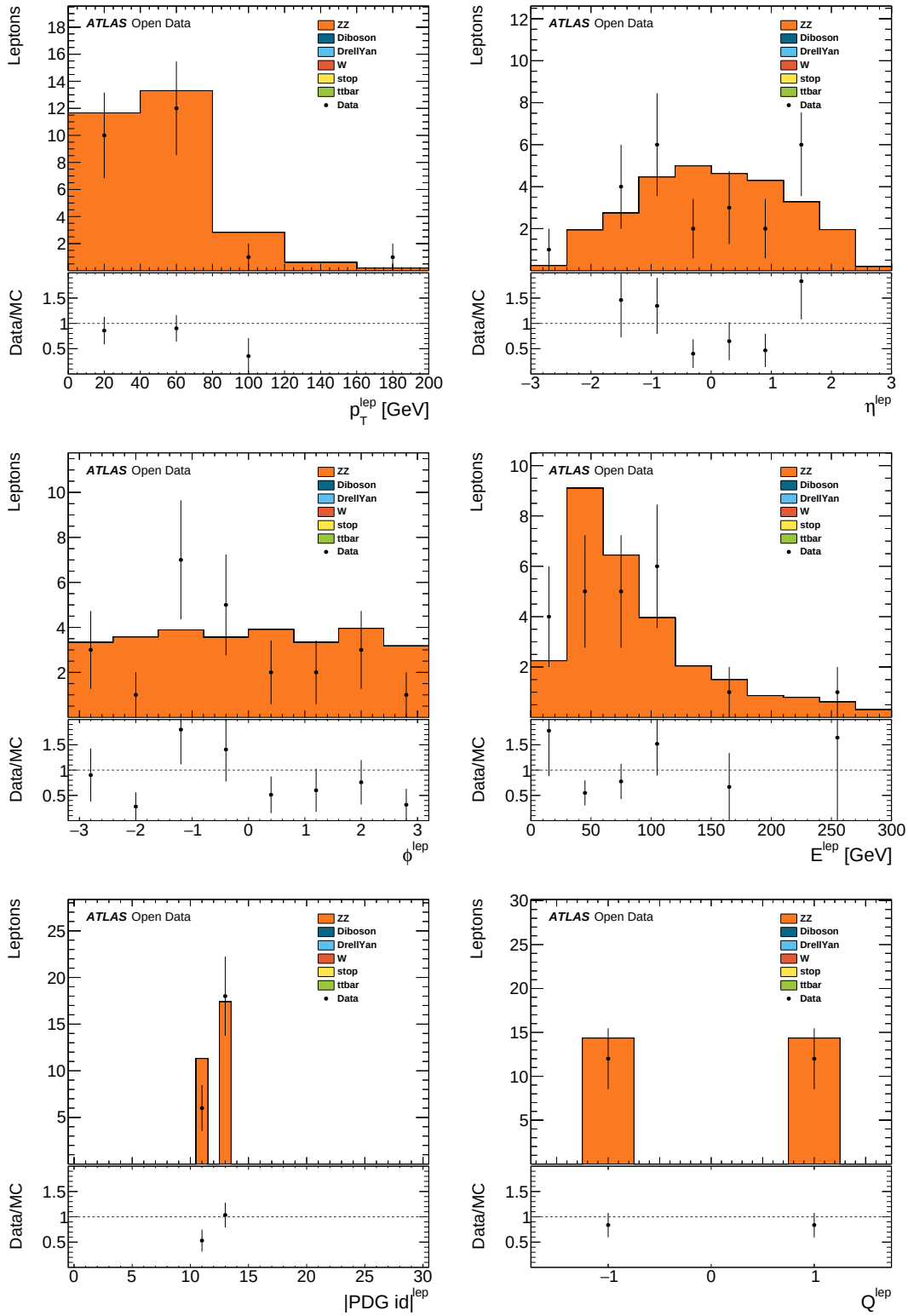

Figure 23: ZZ Analysis: Lepton properties. From upper left to lower right are shown: Transverse momentum p_{T}, pseudorapidity η, azimuthal angle ϕ, energy E, absolute value of the PDG id |PDG id|, and charge Q of the four leptons in the selected events.

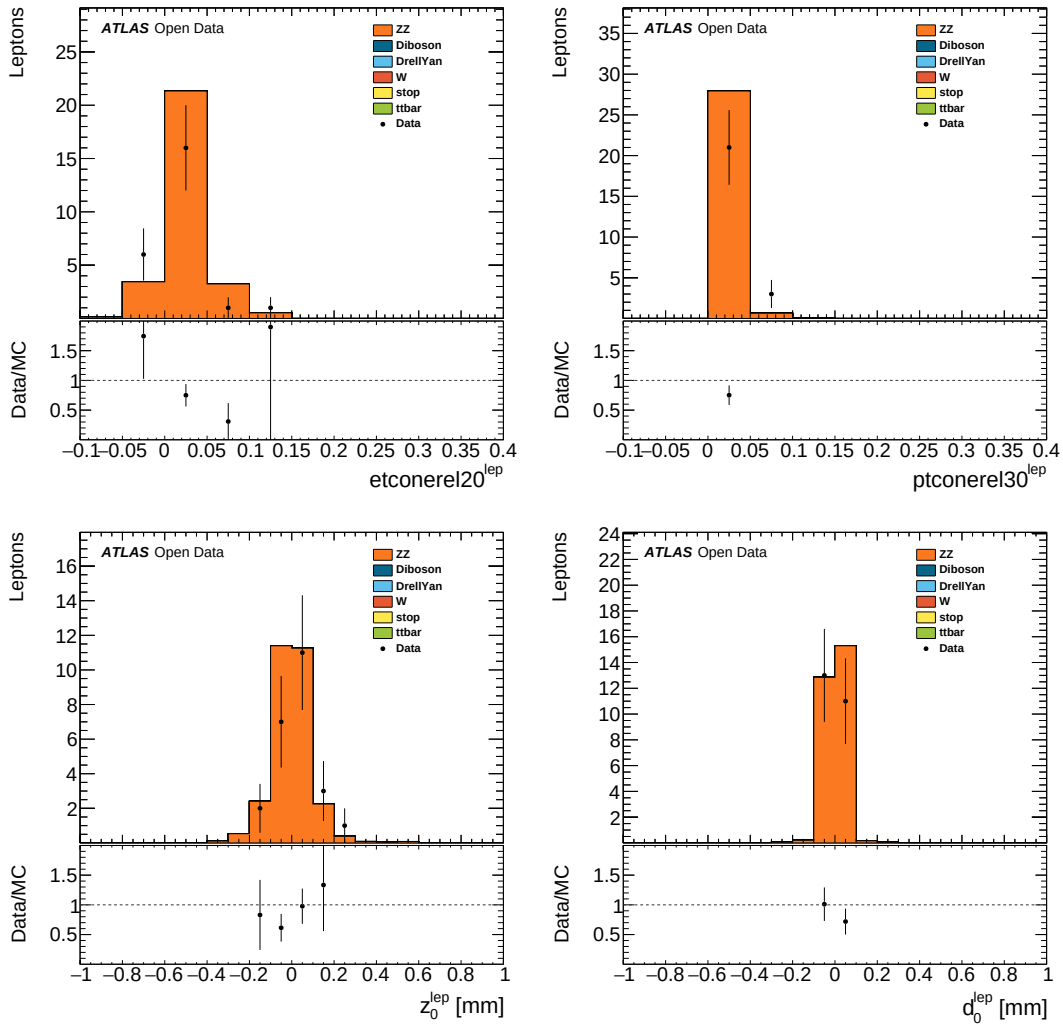

Figure 24: ZZ Analysis: Lepton isolation and tracking information. From upper left to lower right are shown: relative transverse energy isolation (etconerel20), relative transverse momentum isolation (ptconerel30), longitudinal impact parameter z_0, and transverse impact parameter d_0 of the four leptons in the selected events.

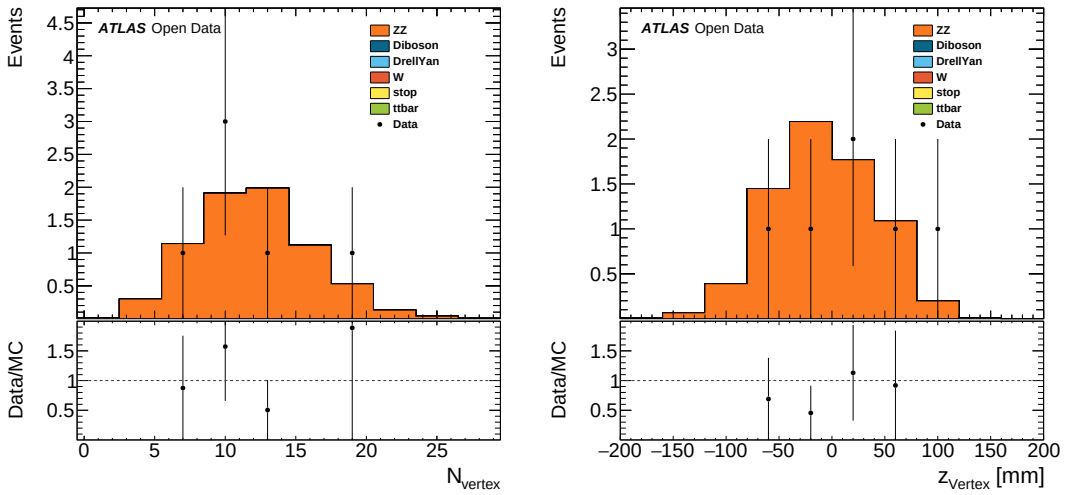

Figure 25: ZZ Analysis: Vertex histograms. The number of vertices N_{vertex} and the z coordinate of the primary vertex z_{vertex} are shown.

2.6 $H \rightarrow WW$ Analysis

The discovery of the Higgs in 2012 was one of the milestones of the LHC physics programme. The $H \rightarrow WW$ analysis is of special interest as it was one of the earliest analyses in ATLAS with sizeable Higgs contributions. In addition, the analysis presented here was used to preselect a part of the data used by the ATLAS W-path Masterclasses [4]. Thus, it represents a natural extension of the exercises performed there.

This analysis implements the criteria for the selection of the zero jet bin of the $H \rightarrow WW$ analysis with both W bosons decaying to leptons [11]. The released data will enable users to develop an understanding of a Higgs analysis. However, they will not be able to derive definitive statements about its existence or properties due to the very limited statistics.

The standard object selection criteria (see Table 2) are applied. The event selection criteria are:

- Single electron or muon trigger is satisfied;
- Event in real data passes the Good Run List;
- Event has a good vertex ($N_{\text{tracks}} > 4$);
- Exactly two good leptons with $p_{\text{T}} > 25$ GeV;
- Leptons have opposite charge;
- No jets with $p_{\text{T}} > 25$ GeV;
- If leptons have same flavour:
  - $m_{\ell\ell}^{\text{vis}} > 12$ GeV;
  - $\mid m_{\ell\ell}^{\text{vis}} - m_Z \mid > 15$ GeV;
  - $E_{\text{T}}^{\text{miss}} > 40$ GeV;
- Else:
  - $m_{\ell\ell} > 10$ GeV;
  - $E_{\text{T}}^{\text{miss}} > 20$ GeV;
- $p_{\text{T},\ell\ell} > 30$ GeV;
- $\Delta\phi(\ell\ell, E_{\text{T}}^{\text{miss}}) > \pi/2$;
- $m_{\ell\ell} < 55$ GeV;
- $\Delta\phi(\text{leadlep}, \text{traillep}) < 1.8$ radians.

The overall normalisation of the selected $H \rightarrow WW$ events looks reasonable. The results shown include the ratio between the Higgs signal hypothesis and the total background represented by the stacked contributions estimated via simulated data. The Higgs signal shape is drawn in front of the stack of Standard Model backgrounds.

Given the low statistics no precise statements about the description of the real data by the simulation can be made. All results shown exhibit good agreement between measured and simulated data. Figure 26 exhibits various event variables relevant for the selection of $H \rightarrow WW$ events. None of them show any

apparent discrepancy. The lepton kinematics for the leading and trailing leptons are shown in Figure 27 and Figure 29. Additional isolation and tracking information is accessible via Figure 28 and Figure 30. No significant disagreements are visible in any of the lepton histograms. The histograms depicting the vertex information in Figure 31 show the expected disagreement between simulated and real data. The pile-up treatment in simulated data considered the whole 2012 run period whilst the real data is taken only from period D of the 2012 data taking.

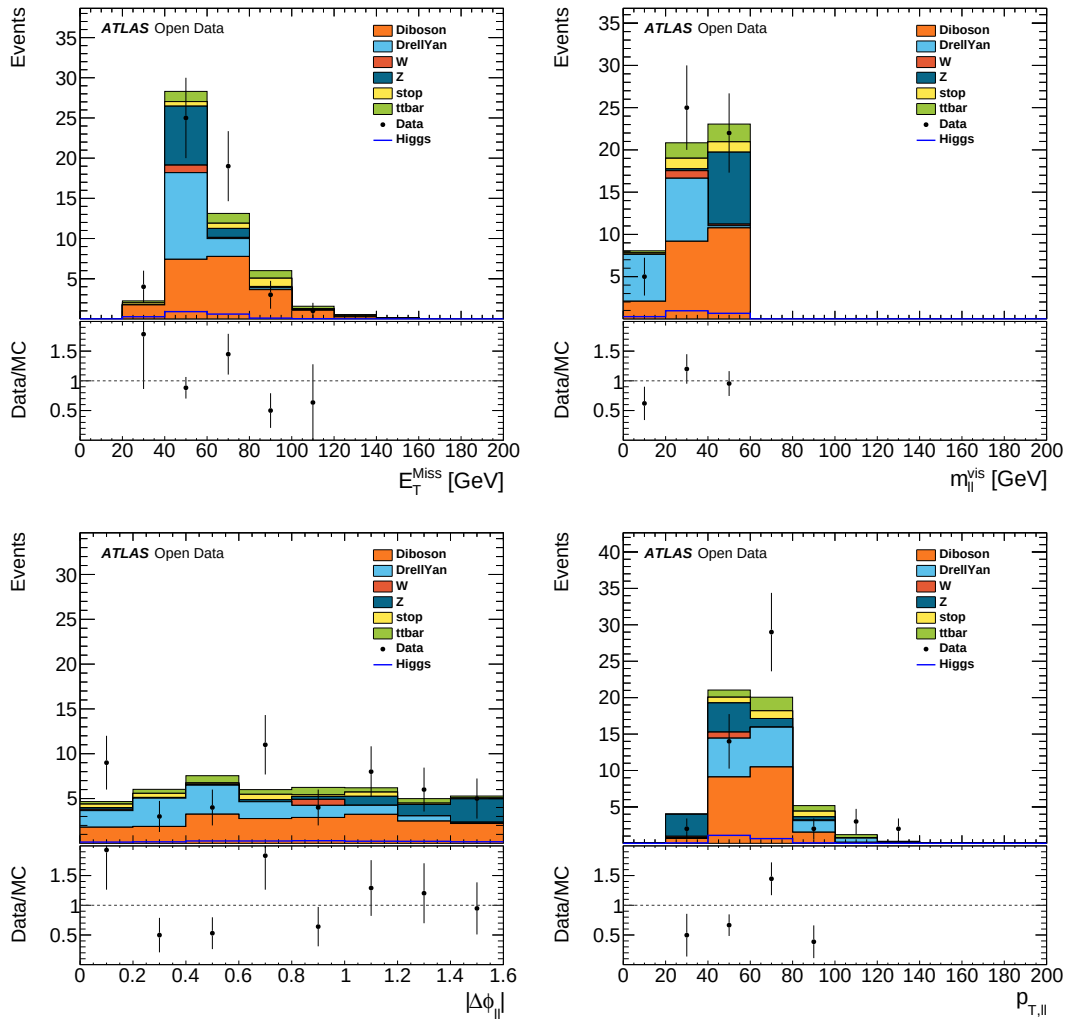

Figure 26: $H \to WW$ Analysis: Event variable histograms. The variables plotted from upper left to lower right are the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, the visible mass of the $H \to WW$ boson candidate $m_{\ell\ell}^{\mathrm{vis}}$, the opening angle in ϕ between the two selected leptons $|\Delta\phi_{\ell\ell}|$, and the transverse momentum of the dilepton system $p_{\mathrm{T},\ell\ell}$.

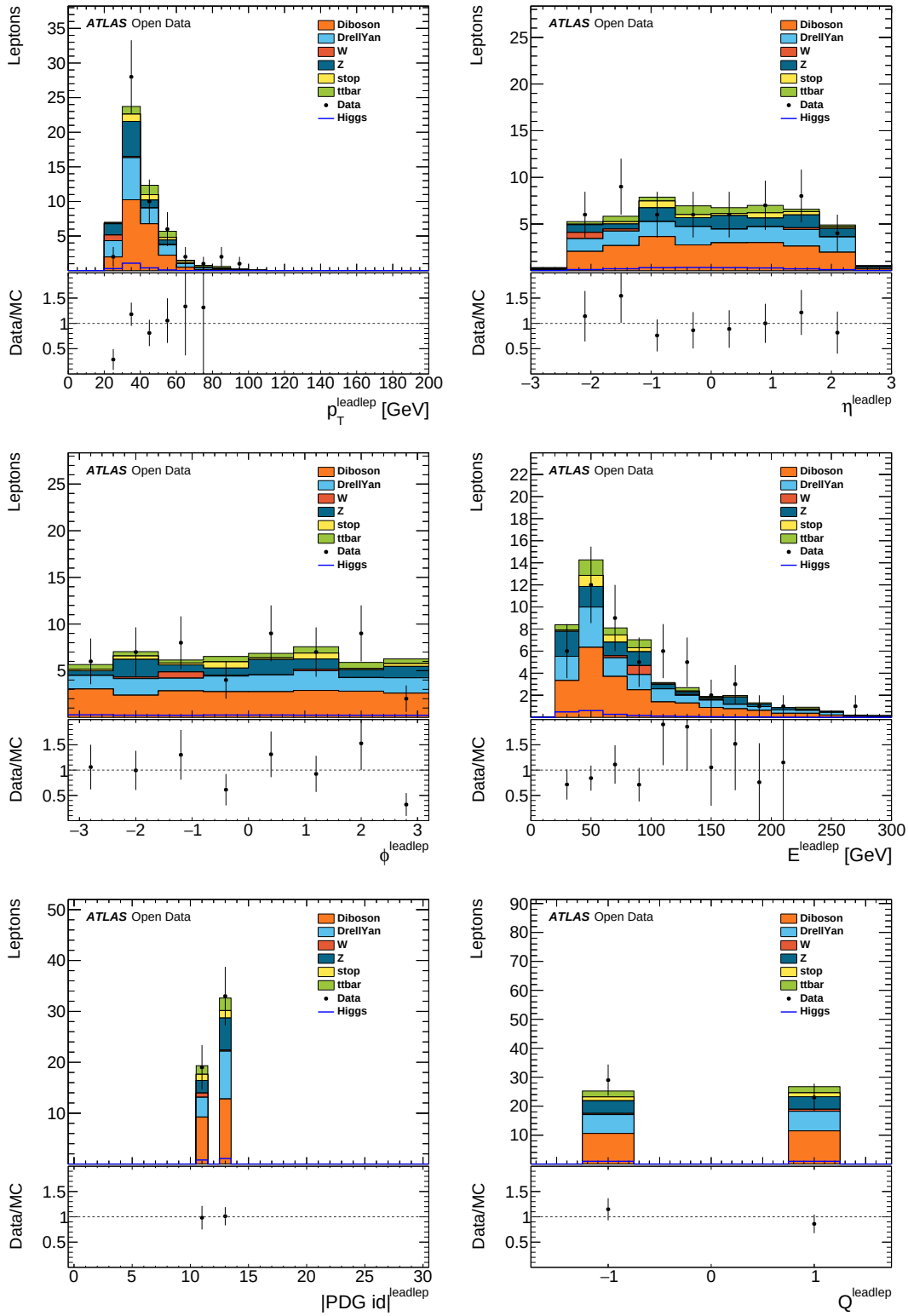

Figure 27: $H \to WW$ Analysis: Leading lepton properties. From upper left to lower right are shown: Transverse momentum p_{T}, pseudorapidity η, azimuthal angle ϕ, energy E, absolute value of the PDG id |PDG id|, and charge Q.

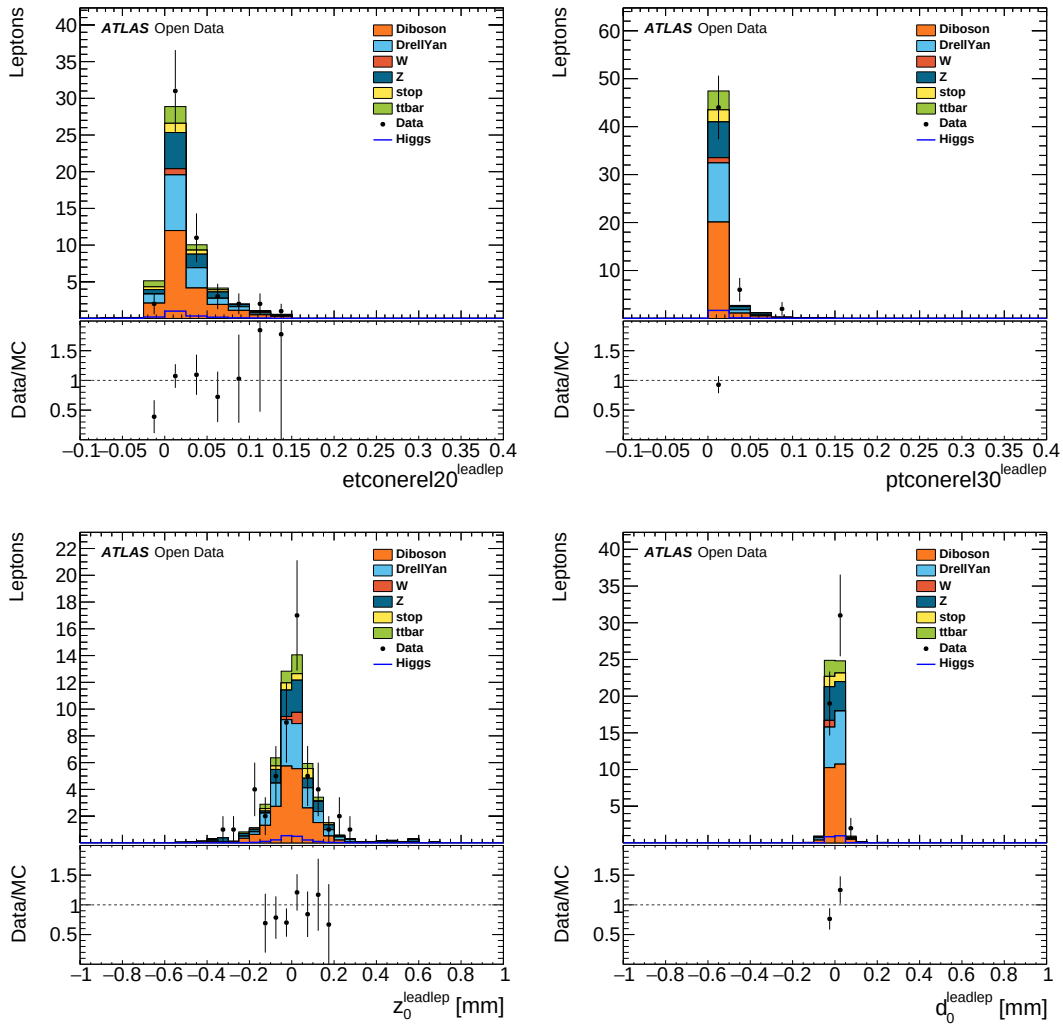

Figure 28: $H \rightarrow WW$ Analysis: Leading lepton isolation and tracking information. From upper left to lower right are shown: relative transverse energy isolation (etconerel20), relative transverse momentum isolation (ptconerel30), longitudinal impact parameter z_0, and transverse impact parameter d_0.

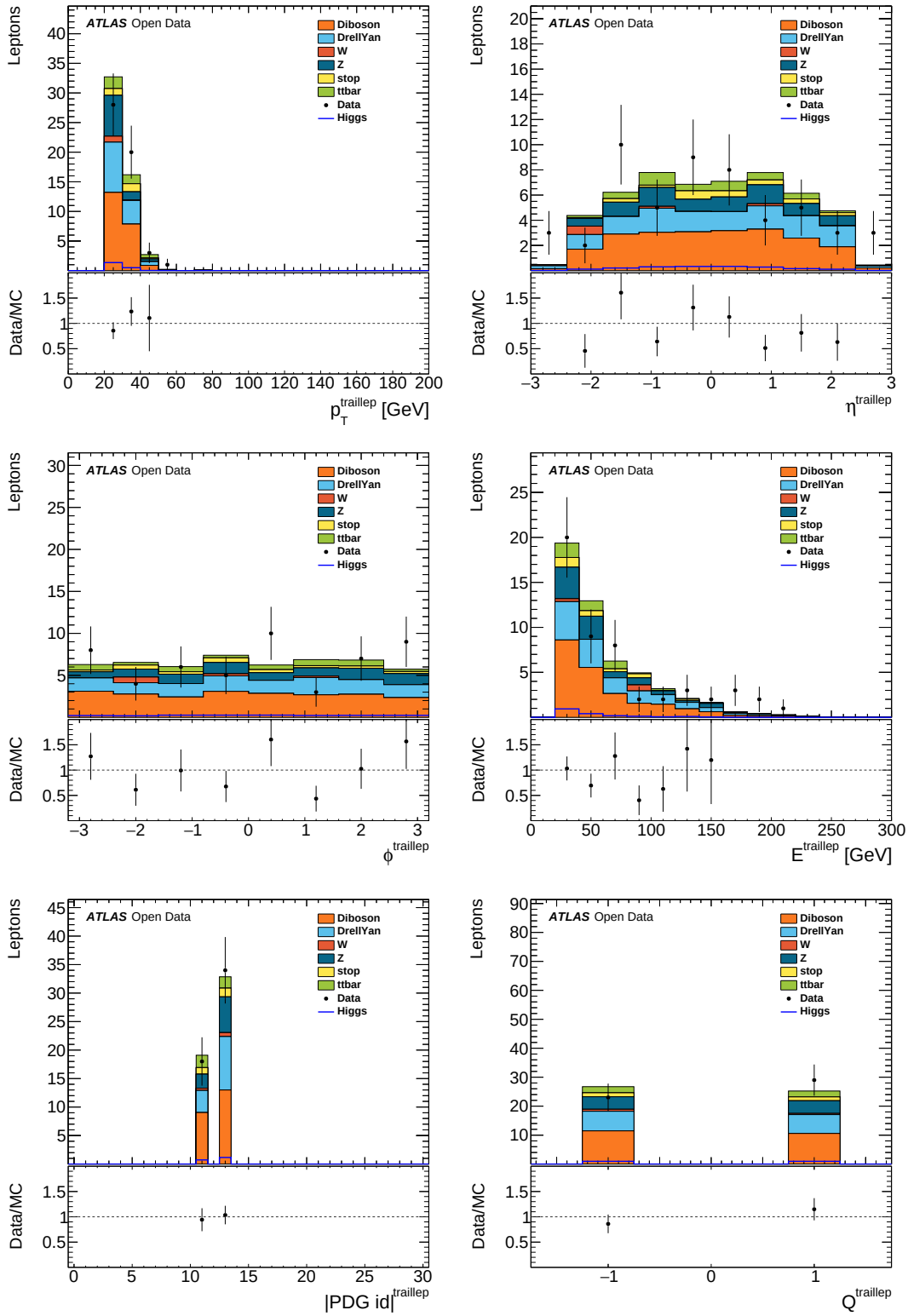

Figure 29: $H \to WW$ Analysis: Trailing lepton properties. From upper left to lower right are shown: Transverse momentum p_{T}, pseudorapidity η, azimuthal angle ϕ, energy E, absolute value of the PDG id |PDG id|, and charge Q.

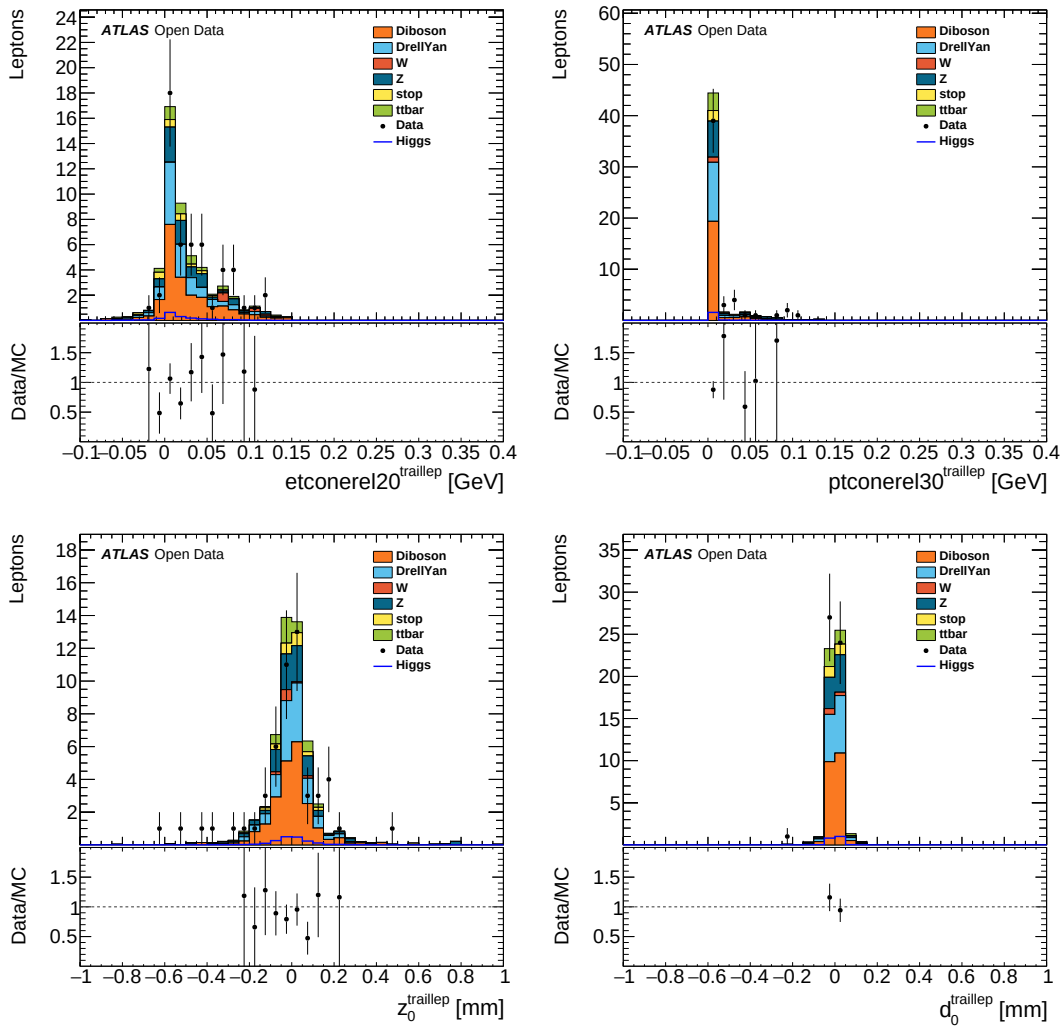

Figure 30: $H \rightarrow WW$ Analysis: Trailing lepton isolation and tracking information. From upper left to lower right are shown: relative transverse energy isolation (etconerel20), relative transverse momentum isolation (ptconerel30), longitudinal impact parameter z_0, and transverse impact parameter d_0.

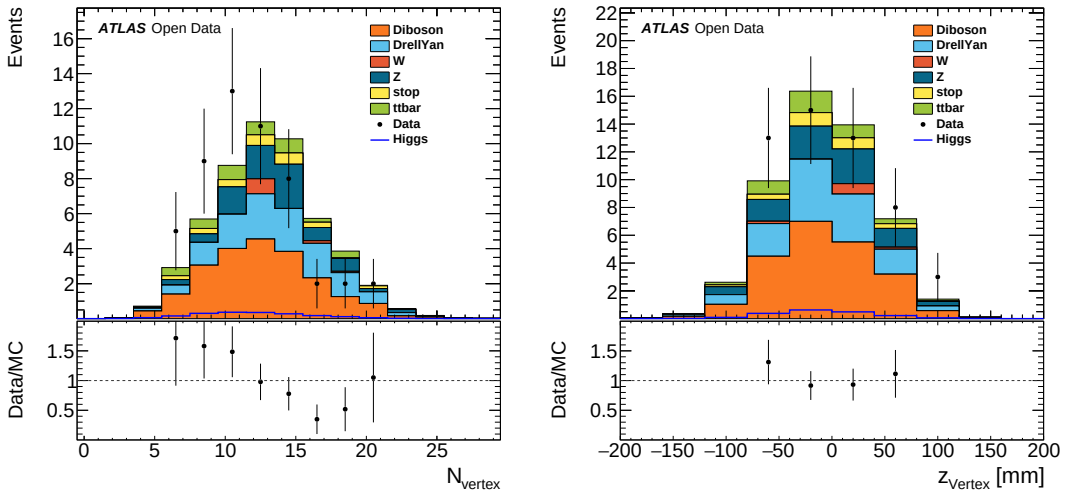

Figure 31: $H \rightarrow WW$ Analysis: Vertex histograms. The number of vertices N_{vertex} and the z coordinate of the primary vertex z_{vertex} are shown.

2.7 Z' Analysis

Searching for new physics beyond the Standard Model is a cornerstone of the ATLAS physics programme. Making such searches available in an educational context exemplifies how they are carried out and the important role of statistical analysis methods. Furthermore topics such as the sensitivity of a variable towards new physics or selection optimisation and its dependence on a free theory parameter may be discussed.

This analysis mimics a $Z' \to t\bar{t}$ analysis in the semileptonic top quark pair channel allowing electrons or muons as lepton candidates ($Z' \to t\bar{t} \to WbWb \to \ell\nu bqqb$). The standard object selection criteria (see Table 2) are applied. The event selection criteria are:

- Single electron or muon trigger is satisfied;
- Event in real data passes the Good Run List;
- Event has a good vertex ($N_{\text{tracks}} > 4$);
- Exactly one good lepton with $p_{\text{T}} > 25$ GeV;
- At least four good jets;
- At least one b-tagged jet (MV1@70%);
- $E_{\text{T}}^{\text{miss}} > 30$ GeV;
- $m_{\text{T}}^{W} + E_{\text{T}}^{\text{miss}} > 60$ GeV.

The figures in this section show the Standard Model backgrounds stacked on top of each other with the signal shapes of two Z' mass hypotheses superimposed. The signal processes have been scaled by a factor of 10 for better visibility. Data is shown as black circles.

The overall agreement between the data and simulated predictions is good. The kinematic description of the leptons is depicted in Figure 32. Isolation information and tracking information is replicated reasonably well in the regions with relevant contributions as can be seen in Figure 33. Figure 34 shows the kinematics of the jets in the selected events as well as the MV1 b-tagging weights and the jet vertex fraction. Overall, the kinematics are very well described. The jet vertex fraction and the MV1 weight are more complex variables, but are well reproduced by the simulated data.

The histograms depicting the vertex information in Figure 35 show the expected disagreement between simulated and real data. The pile-up treatment in simulated data considers the whole 2012 run period whilst the real data is taken only from period D of the 2012 data taking. A slight slope is observed in the data/simulation ratio for the missing transverse momentum (see Figure 36), which is likely to be caused by either the non inclusion of QCD contributions or the non-optimal description of $E_{\text{T}}^{\text{miss}}$ in the Z and W samples.

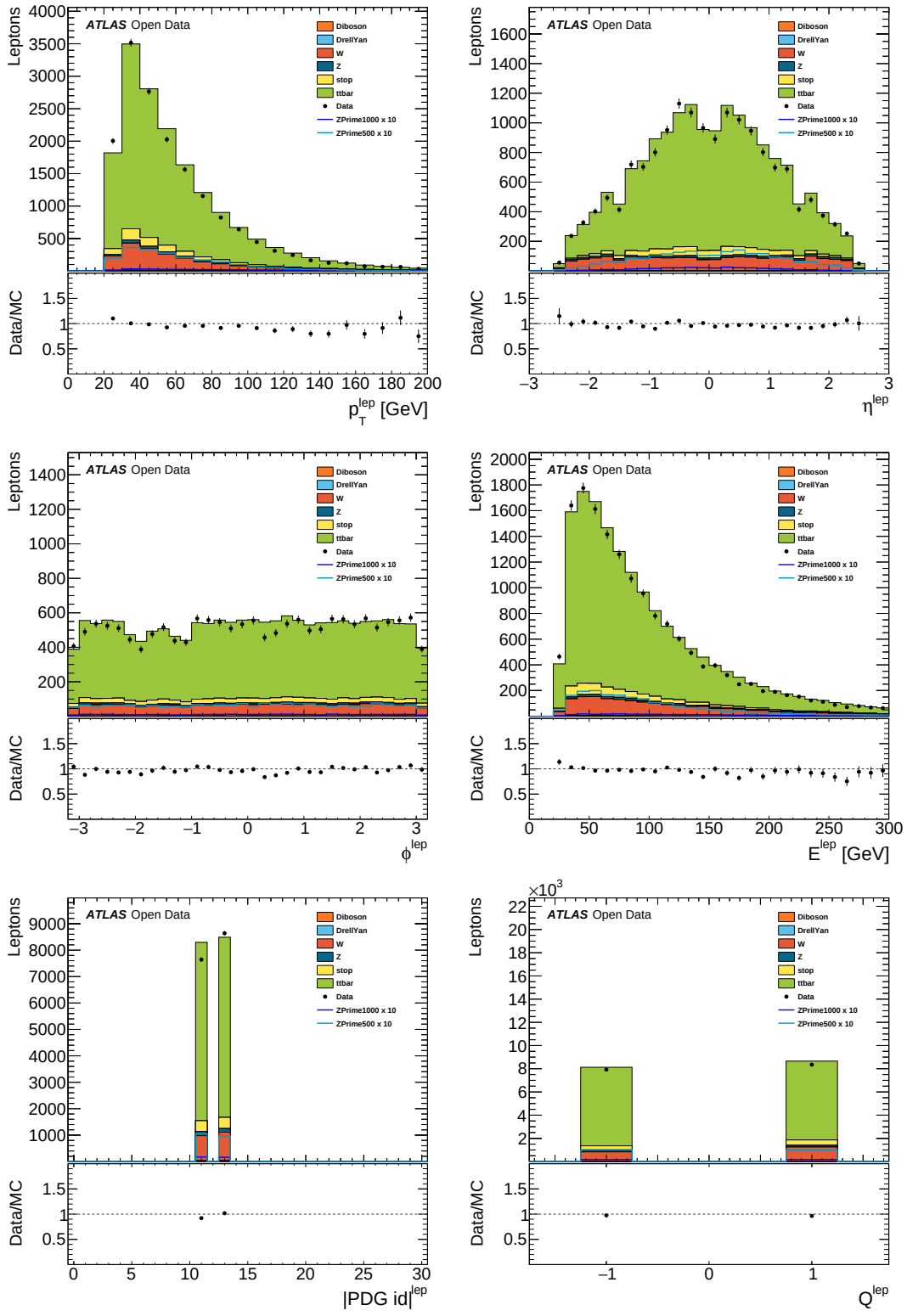

Figure 32: Z' Analysis: Leading lepton properties. From upper left to lower right are shown: Transverse momentum p_{T}, pseudorapidity η, azimuthal angle ϕ, energy E, absolute value of the PDG id |PDG id|, and charge Q.

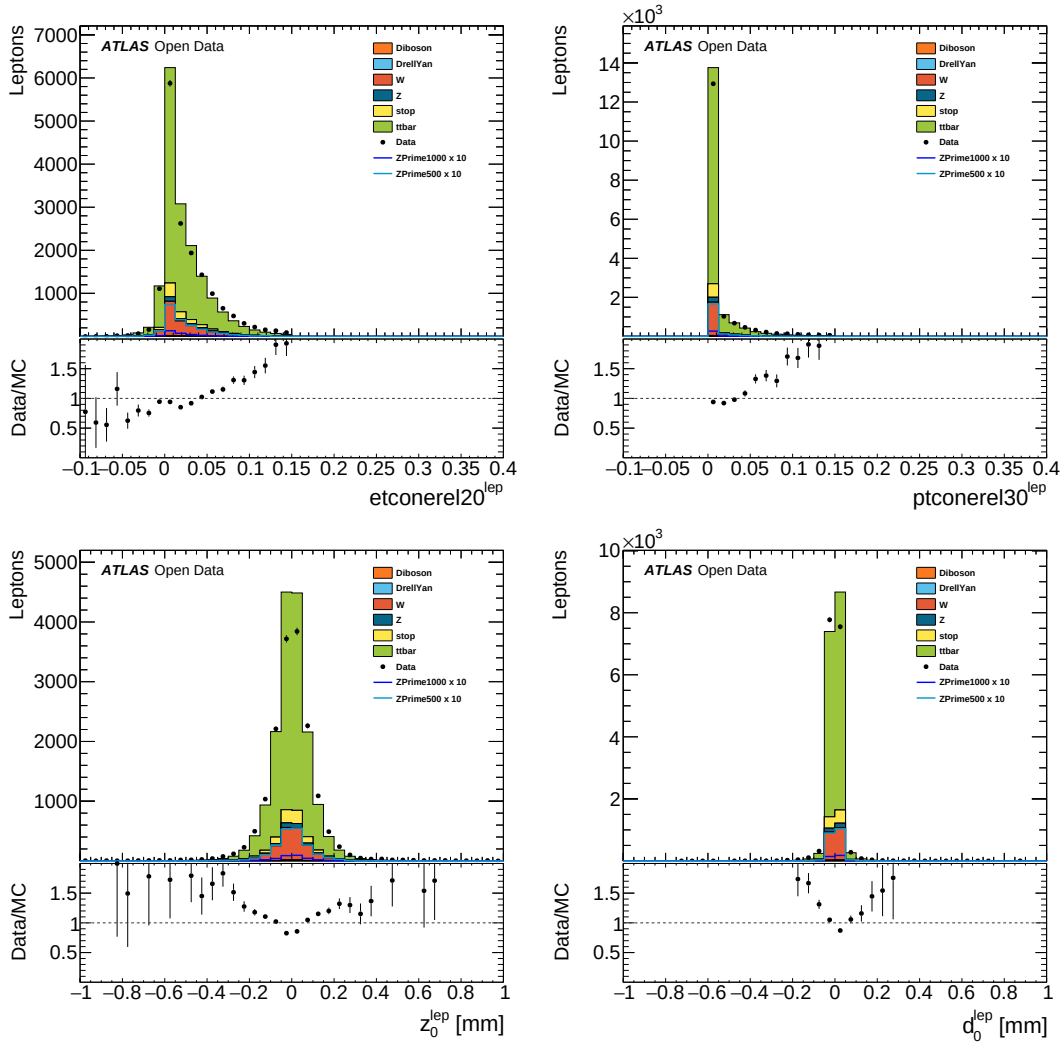

Figure 33: Z' Analysis: Leading lepton isolation and tracking information. From upper left to lower right are shown: relative transverse energy isolation (etconerel20), relative transverse momentum isolation (ptconerel30), longitudinal impact parameter z_0, and transverse impact parameter d_0.

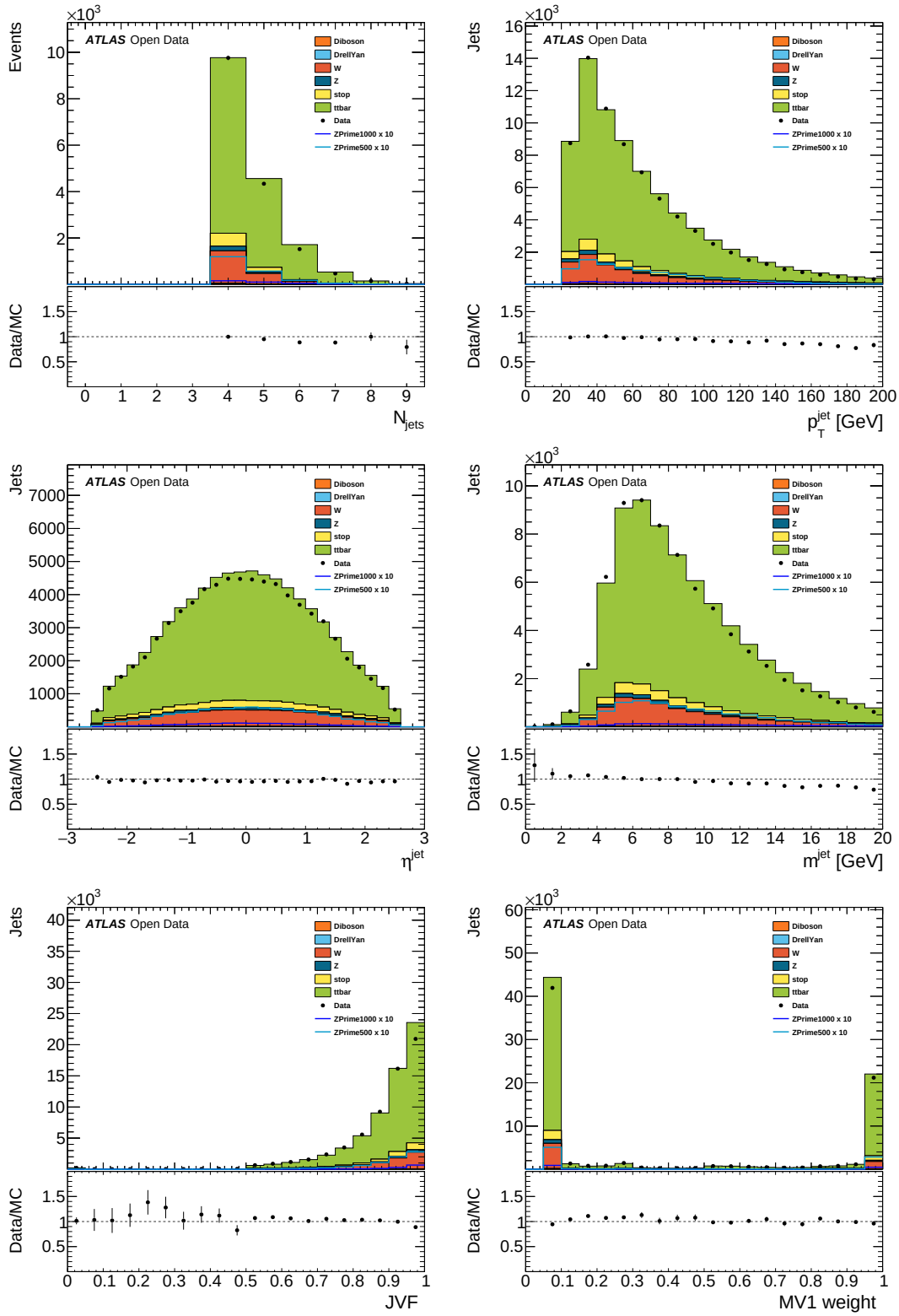

Figure 34: Z' Analysis: Jet properties. From upper left to lower right are shown: Jet multiplicity N_{jets}, transverse momentum p_{T}, pseudorapidity η, mass m, jet vertex fraction (JVF), and MV1 b-tagging weight of the selected jets.

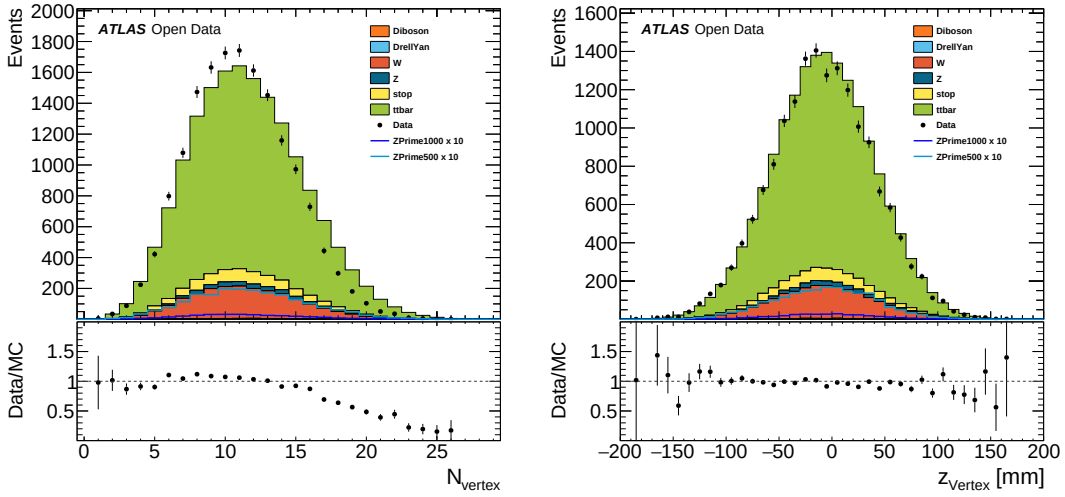

Figure 35: Z' Analysis: Vertex histograms. The number of vertices N_{vertex} and the z coordinate of the primary vertex z_{vertex} are shown.

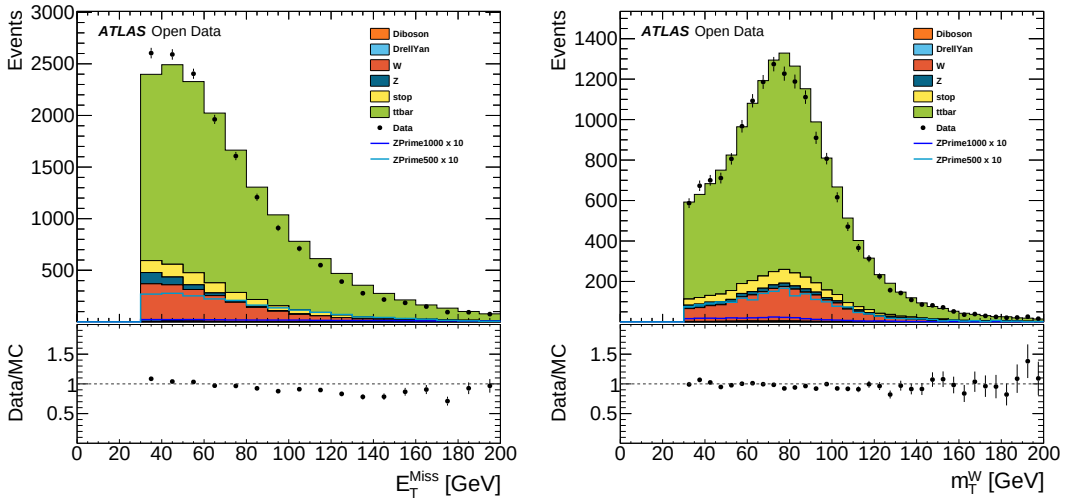

Figure 36: Z' Analysis: Event variable histograms. The variables plotted are missing transverse momentum $E_{\text{T}}^{\text{miss}}$ and the transverse mass of the W boson candidate m_{T}^{W}.

3 Summary

The production of the ATLAS open data dataset and the tools accompanying it have been discussed. The prepared bundle is released in accordance with the ATLAS Data Policy [1]. Results of a number of example analyses inspired by actual analyses by ATLAS have been presented to demonstrate possible applications of the ATLAS open data dataset. Overall, a good agreement has been observed and sources of possible deviations have been identified and discussed.

It is believed that the ATLAS open data dataset and tools together with the documentation that will be made available separately will provide an engaging learning environment for undergraduate students and other interested audiences.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DSM/IRFU, France; GNSF, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, the Canada Council, CANARIE, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d'Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BSF, GIF and Minerva, Israel; BRF, Norway; Generalitat de Catalunya, Generalitat Valenciana, Spain; the Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.

Appendix

branchname	type	description
runNumber	int	run number
eventNumber	int	event number
channelNumber	int	channel number
mcWeight	float	weight of an MC event
pvxp_n	int	number of primary vertices
vxp_z	float	z-position of the primary vertex
trigE	bool	boolean whether a standard trigger is satisfied in the egamma stream
trigM	bool	boolean whether a standard trigger is satisfied in the muon stream
passGRL	bool	signifies whether event passes the Good Run List and thus put in isGoodEvent
hasGoodVertex	bool	signifies whether the event has at least one good vertex
lep_n	int	number of preselected leptons
lep_truthMatched	vector<bool>	boolean indicating whether the lepton is matched to a truth lepton
lep_trigMatched	vector<bool>	boolean signifying whether the lepton is the one triggering the event
lep_pt	vector<float>	transverse momentum of the lepton
lep_eta	vector<float>	pseudorapidity of the lepton
lep_phi	vector<float>	azimuthal angle of the lepton
lep_E	vector<float>	energy of the lepton
lep_z0	vector<float>	z-coordinate of the track associated to the lepton wrt. the primary vertex
lep_charge	vector<float>	charge of the lepton
lep_flag	vector<int>	bitmask implementing object cuts of the top group
lep_type	vector<int>	number signifying the lepton type (e, mu, tau) of the lepton
lep_ptcone30	vector<float>	ptcone30 isolation for the lepton
lep_etcone20	vector<float>	etcone20 isolation for the lepton
lep_trackd0pvunbiased	vector<float>	d0 of the track associated to the lepton at the point of closest approach (p.c.a.)
lep_tracksigd0pvunbiased	vector<float>	d0 significance of the track associated to the lepton at the p.c.a.
met_et	float	Transverse energy of the missing momentum vector
met_phi	float	Azimuthal angle of the missing momentum vector
jet_n	int	number of selected jets
jet_pt	vector<float>	transverse momentum of the jet
jet_eta	vector<float>	pseudorapidity of the jet
jet_phi	vector<float>	azimuthal angle of the jet
jet_E	vector<float>	energy of the jet
jet_m	vector<float>	invariant mass of the jet
jet_jvf	vector<float>	JetVertexFraction of the jet
jet_trueflav	vector<int>	true flavor of the jet
jet_truthMatched	vector<int>	information whether the jet matches a jet on truth level
jet_SV0	vector<float>	SV0 weight of the jet
jet_MV1	vector<float>	MV1 weight of the jet
scaleFactor_BTAG	float	scalefactor for btagging
scaleFactor_ELE	float	scalefactor for electron efficiency
scaleFactor_JVFSF	float	scalefactor for jet vertex fraction
scaleFactor_MUON	float	scalefactor for muon efficiency
scaleFactor_PILEUP	float	scalefactor for pileup reweighting
scaleFactor_TRIGGER	float	scalefactor for trigger
scaleFactor_ZVERTEX	float	scalefactor for z-vertex reweighting

Table 3: Branches of the tuples for the ATLAS data release. The content of these tuples was defined by the ATLAS top analysis group. The technical implementation used was AnalysisTop-1.9.1. Superfluous branches not needed for the educational purposes of the data release were dropped.

period	$N_{\text{Events}}^{\text{preselectd}}$	$N_{\text{Events}}^{\text{total}}$	$\mathcal{L}$ [pb^{-1}]	size/Mb
Egamma	7917590	33575219	1000.6	723
Muons	7028084	33815203	1000.6	600

Table 4: Breakdown of the sample of real data with a total integrated luminosity of 1 fb^{-1}. $N_{\text{Events}}^{\text{preselected}}$ denotes the number of events after preselection, $N_{\text{Events}}^{\text{total}}$ the number of events prior to the preselection, $\mathcal{L}$ the luminosity of the sample and size after preselection. The sample is made by combining the runs 207490, 207532, 207582, 207589, 207749, 207772, 207845, 207865, 207934, 207982, 208126, 208184, 208189, and 208258. The real data is selected using the same preselection as applied on the simulated data.

process	DSID	Generator	σ*FE [pb]	f_k	$\mathcal{L}$ [fb^{-1}]	$N_{\text{Events}}^{\text{reduced}}$	$N_{\text{Events}}^{\text{preselected}}$	size/Mb
$t\bar{t} \to l + X$	117050	Powheg +Pythia	114.51	1.2	26.236	1500000	20775908	291
$t\bar{t} \to$ Jets	117049	Powheg +Pythia	96.35	1.2	85.027	25170	25170	5.7
single top t-chan top	110090	Powheg +Pythia	17.52	1.05	24.21	150000	1678087	21
single top t-chan antitop	110091	Powheg +Pythia	9.4	1.06	43.23	150000	1719075	15
single top s-chan	110119	Powheg +Pythia	1.64	1.107	167.73	100000	1966242	15
single top Wt-chan	110140	Powheg +Pythia	20.46	1.09	28.50	150000	235557	26
Z+Jets ee	147770	Sherpa	1207.4	1.028	10.08	7500000	49405819	938
Z+Jets mumu	147771	Sherpa	1207.4	1.028	9.63	7500000	60149707	918
Z+Jets tautau	147772	Sherpa	1207.1	1.028	11.08	750000	814528	93
Drell-Yan ee M08to15	173041	Sherpa	92.15	1.0	45.95	400000	447800	57
Drell-Yan ee M15to40	173042	Sherpa	279.19	1.0	47.22	750000	793055	100
Drell-Yan mumu M08to15	173043	Sherpa	92.08	1.0	51.93	500000	520562	74
Drell-Yan mumu M15to40	173044	Sherpa	279.2	1.0	41.01	750000	750246	103
Drell-Yan tautau M08to15	173045	Sherpa	92.12	1.0	27.13	9993	9993	1.5
Drell-Yan tautau M15to40	173046	Sherpa	279.11	1.0	49.54	32393	32393	4.5
W+Jets enu with b	167740	Sherpa	140.34	1.1	12.333	750000	5792095	86
W+Jets enu with jets, bveto	167741	Sherpa	537.84	1.1	9.563	2600000	2648506	296
W+Jets enu no jets, bveto	167742	Sherpa	10295	1.1	1.971	8000000	8448069	722
W+Jets munu with b	167743	Sherpa	140.39	1.1	11.935	750000	5630683	84
W+Jets munu with jets, bveto	167744	Sherpa	466.47	1.1	10.582	2500000	2759594	287
W+Jets munu no jets, bveto	167745	Sherpa	10368	1.1	1.719	7500000	7946599	666
W+Jets taunu with b	167746	Sherpa	140.34	1.1	18.245	100000	531981	13
W+Jets taunu with jets, bveto	167747	Sherpa	506.45	1.1	9.821	250000	273867	31
W+Jets taunu no jets, bveto	167748	Sherpa	10327	1.1	1.945	550000	593205	55
WW	105985	Herwig	12.42	1.683	46.32	500000	1288259	63
ZZ	105986	Herwig	0.992	1.55	151.19	125000	131435	20
WZ	105987	Herwig	3.667	1.9	138.44	500000	517196	68

Table 5: Samples for simulated data of the ATLAS open data dataset describing Standard Model processes. The individual processes are derived using the simulated datasets with the given dataset id (DSID). Cross sections combined with filter efficiencies are given with the appropriate scaling factors f_k for higher order QCD corrections where available. After being subjected to a preselection $N_{\text{Events}}^{\text{preselected}}$ are available in the samples. A reduction procedure is applied in order to decrease the processing time and storage requirements which further reduces the number of events found in the samples. Resulting event yields after preselection and reduction and the luminosity of these samples are denoted as $N_{\text{Events}}^{\text{reduced}}$ and $\mathcal{L}$, respectively.

process	DSID	Generator	σ*FE [pb]	f_k	$\mathcal{L}$ [fb^{-1}]	$N_{\text{Events}}^{\text{reduced}}$	$N_{\text{Events}}^{\text{preselected}}$	size/Mb
$Z' \to t\bar{t}$ $M_{Z'} = 400$ GeV	110899	Pythia	4.259	1.0	23.48	21941	21941	4.3
$Z' \to t\bar{t}$ $M_{Z'} = 500$ GeV	110901	Pythia	3.925	1.0	25.48	23231	23231	4.7
$Z' \to t\bar{t}$ $M_{Z'} = 750$ GeV	110902	Pythia	1.243	1.0	80.45	25021	25021	5.3
$Z' \to t\bar{t}$ $M_{Z'} = 1000$ GeV	110903	Pythia	0.394	1.0	253.81	25525	25525	5.5
$Z' \to t\bar{t}$ $M_{Z'} = 1250$ GeV	110904	Pythia	0.139	1.0	719.43	25030	25030	5.5
$Z' \to t\bar{t}$ $M_{Z'} = 1500$ GeV	110905	Pythia	0.0524	1.0	1908	24142	24142	5.4
$Z' \to t\bar{t}$ $M_{Z'} = 1750$ GeV	110906	Pythia	0.0211	1.0	4739	23084	23084	5.1
$Z' \to t\bar{t}$ $M_{Z'} = 2000$ GeV	110907	Pythia	0.00894	1.0	11186	21997	21997	4.9
$Z' \to t\bar{t}$ $M_{Z'} = 2250$ GeV	110908	Pythia	0.00394	1.0	25381	21127	21127	4.7
$Z' \to t\bar{t}$ $M_{Z'} = 2500$ GeV	110909	Pythia	0.00180	1.0	55556	20327	20327	4.5
$Z' \to t\bar{t}$ $M_{Z'} = 3000$ GeV	110910	Pythia	0.000434	1.0	230415	19646	19646	4.3
$gg \to H \to WW \to ll\nu\nu$ $M_{\text{H}} = 125$ GeV	161005	Powheg +Pythia	6.463	1.0	32.13	100000	278332	14
VBF$H \to WW \to ll\nu\nu$ $M_{\text{H}} = 125$ GeV	161055	Powheg +Pythia	0.819	1.0	229.93	100000	183101	18
$gg \to H \to ZZ \to 4l$ $M_{\text{H}} = 125$ GeV	160155	Powheg +Pythia	13.17	1.0	14.31	100000	117081	15
VBF$H \to ZZ \to 4l$ $M_{\text{H}} = 125$ GeV	160205	Powheg +Pythia	1.617	1.0	104.96	100000	130213	19

Table 6: Samples for simulated data of the ATLAS open data dataset describing Beyond the Standard Model signals and Higgs physics. The individual processes are derived using the simulated datasets with the given dataset id (DSID). Cross sections combined with filter efficiencies are given with the appropriate scaling factors f_k for higher order QCD corrections where available. After being subjected to a preselection $N_{\text{Events}}^{\text{preselected}}$ are available in the samples. A reduction procedure is applied in order to decrease the processing time and storage requirements which further reduces the number of events found in the samples. Resulting event yields after preselection and reduction and the luminosity of these samples are denoted as $N_{\text{Events}}^{\text{reduced}}$ and $\mathcal{L}$, respectively.

References

[1] ATLAS Collaboration, *ATLAS Data Access Policy*, (2014).

[2] 'Kaggle Higgs Boson Machine Learning Challenge', https://www.kaggle.com/c/higgs-boson, Accessed: 2016-06-16.

[3] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81.

[4] 'ATLAS Masterclasses W-path', http://atlas.physicsmasterclasses.org/en/wpath.htm, Accessed: 2016-06-14.

[5] ATLAS Collaboration, *Measurement of the W charge asymmetry in the $W \to \mu\nu$ decay mode in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Lett. **B701** (2011) 31, arXiv: 1103.2929 [hep-ex].

[6] G. Aad et al., *Measurement of the $W \to \ell\nu$ and $Z/\gamma^* \to \ell\ell$ production cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **12** (2010) 060, arXiv: 1010.2130 [hep-ex].

[7] K. A. Olive et al., *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001.

[8] 'ATLAS Masterclasses Z-path', http://atlas.physicsmasterclasses.org/en/zpath.htm, Accessed: 2016-06-14.

[9] ATLAS Collaboration, *Measurements of $W^{\pm}Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector and limits on anomalous gauge boson self-couplings*, Phys. Rev. **D93** (2016) 092004, arXiv: 1603.02151 [hep-ex].

[10] ATLAS Collaboration, *Measurement of the ZZ production cross section and limits on anomalous neutral triple gauge couplings in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Rev. Lett. **108** (2012) 041804, arXiv: 1110.5016 [hep-ex].

[11] ATLAS Collaboration, 'Search for the Standard Model Higgs boson in the $H \to WW \to \ell\ell\nu\nu$ decay mode using 1.7 fb-1 of data collected with the ATLAS detector at sqrt(s)=7 TeV', tech. rep. ATLAS-CONF-2011-134, CERN, 2011, URL: http://cds.cern.ch/record/1383837.