ATLAS PUB Note

PUB-OTRC-2020-01

28th April 2020

Review of the 13 TeV ATLAS Open Data release

The ATLAS Collaboration

The ATLAS Collaboration is releasing a new set of proton–proton collision data to the public for educational purposes. The data has been collected by the ATLAS detector at the Large Hadron Collider at a centre-of-mass energy $\sqrt{s}$ = 13 TeV during the year 2016 and corresponds to an integrated luminosity of 10 fb^{-1}. This dataset is accompanied by simulated events describing both several Standard Model processes, as well as hypothetical Beyond Standard Model signal production. Associated computing tools are provided to make the analysis of the dataset easily accessible. This document summarises the properties of the 13 TeV ATLAS Open Data set and the available analysis tools. Several examples intended as a starting point for further analysis work by users are shown. The general aim of the dataset and tools released is to provide user-friendly and straightforward interactive interfaces to replicate the procedures used by high-energy-physics researchers and enable users to experience the analysis of particle-physics data in educational environments.

Contents

1 Introduction

The aim of ATLAS Open Data is to provide data and tools to high school, undergraduate and graduate students, as well as teachers and lecturers, to help educate them and exercise in physics analysis techniques used in experimental particle physics. Sharing data collected by the ATLAS experiment aims to generate excitement and enthusiasm for fundamental research, inspiring physicists of the future.

The ATLAS Data Access Policy [1] sets out the guidelines regarding open access to ATLAS data by non-ATLAS members with a focus on education, training and outreach. The mandate supports the creation of multiple projects, software and educational platforms, such as the CERN and ATLAS Open Data websites [2, 3]. Since the endorsement of the policy in June 2014, the ATLAS Collaboration has released two datasets of proton–proton (pp) collision data with a centre-of-mass energy $\sqrt{s} = 8$ TeV collected by the ATLAS detector [4] at the Large Hadron Collider (LHC) during the year 2012. These datasets correspond to an integrated luminosity of 2 fb^{-1} in XML format used by the well-established IPPOG [5] "International Masterclasses" [6] activities for high school students, and 1 fb^{-1} in ROOT [7] format, accompanied by the corresponding analysis tools [8] and aimed at undergraduate physics students. All 8 TeV datasets, tools, software and documentation for studying the 8 TeV datasets, including the educational platform [3], are collectively known as "8 TeV ATLAS Open Data". A number of groups within ATLAS as well as external collaborators and users reported a large range of activities based on the latter 8 TeV datasets, described in Refs. [9, 10].

A new set of pp collision data at $\sqrt{s} = 13$ TeV is released to the public for educational purposes. The dataset corresponds to an integrated luminosity of 10 fb^{-1} of pp collisions with $\sqrt{s} = 13$ TeV recorded by the ATLAS detector at the LHC in 2016. Monte Carlo (MC) simulation samples describing several Standard Model (SM) and beyond the Standard Model (BSM) processes, which are used to model the expected distributions of different signal and background processes, are included in the release as well.

These datasets, referred to as "13 TeV ATLAS Open Data", are 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 and graduate students. The released data may also prove beneficial for the production of teaching materials, lectures and public talks. Furthermore, it may be used for further development of analysis methods and techniques, such as Higgs and particle-tracking machine learning (ML) challenges run on Kaggle's platform [11, 12]. The ATLAS Open Data website [3] provides access to the data and instructions on how to analyse it with further details given in Section 5. This document provides more details on the ATLAS specific event selection and different final-states physics analyses and is meant to be a guide on the content, properties, capabilities and limitations of the released datasets.

2 Overview of 13 TeV ATLAS Open Data

The 13 TeV ATLAS Open Data events belong to 61 runs from the first four periods of the 2016 pp data-taking and contain approximately 270 million of collision events. Only events for which all relevant subsystems were operational are considered. After applying quality criteria for the beam, data and detector, the publicly released dataset corresponds to an integrated luminosity of 10.06 ± 0.37 fb^{-1} [13]. The event reconstruction is affected by multiple inelastic pp collisions in a single bunch crossing and by collisions in neighbouring bunch crossings, referred to as "pile-up". The number of interactions per bunch crossing in this data set ranges from about 8 to 45.

The 13 TeV ATLAS Open Data set is comprised not only of pp collisions data recorded with the ATLAS detector in 2016. It is accompanied by MC simulation samples describing several SM processes which are used to model the expected distributions of different signal and background events. All simulated samples were processed through the same reconstruction algorithms and analysis chain as the data and subjected to a loose event preselection to reduce processing time. Events are selected by applying several event-quality and trigger criteria, and classified according to the type and multiplicity of reconstructed objects with high transverse momentum (p_{T}).[1]

Electron candidates are reconstructed [14] from an isolated electromagnetic calorimeter (EMCAL) energy deposit, matched to a track in the inner tracking detector (InDet), within the fiducial region of $|\eta_{\mathrm{cluster}}| < 2.47$, where η_{cluster} is the pseudorapidity of the calorimeter energy deposit associated with the electron candidate. Candidates within the transition region between the barrel and endcap electromagnetic calorimeters, $1.37 < |\eta_{\mathrm{cluster}}| < 1.52$, are excluded. Furthermore, each electron candidate is required to have $p_{\mathrm{T}} > 7$ GeV and to pass the "loose" identification criteria [14]. Muon candidates are reconstructed [15] by combining tracks reconstructed in both the inner detector and the muon spectrometer (MS). The muon candidates are selected if they have a transverse momentum above 7 GeV and pass the "loose" identification requirements defined in Ref. [15]. To reduce the contribution from non-prompt leptons (e.g. from semileptonic b- or c-hadron decays), photon conversions and hadrons, "loose" isolation requirements are applied to both the electron and muon candidates [14, 15]. Events with at least one electron or muon are selected with single-lepton triggers with either low p_{T} thresholds of 26 GeV and isolation requirements, or with higher thresholds of 50–60 GeV but with a looser identification criterion and without any isolation requirement. Corrections are applied to the MC samples to match the leptons' trigger, reconstruction and isolation efficiencies in data [14–16].

The reconstruction and calibration of photon candidates [16, 17] proceeds similarly as for electron candidates, with energy clusters in the EMCAL seeded with transverse energy greater than 2.5 GeV. Converted photon candidates are associated with the conversion of photons into electron–positron pairs in the material upstream the EMCAL. Both unconverted and converted photon candidates are required to have $E_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.37$, excluding the region $1.37 < |\eta| < 1.52$, and to satisfy a set of "tight" identification criteria [17] to reduce the contamination from the background, primarily associated with neutral hadrons in jets decaying into photon pairs. In order to reduce hadronic background, photon candidates must satisfy "loose" isolation criteria [17] based on both track and calorimeter information. Events are selected using a diphoton trigger requiring the presence in the EMCAL of two clusters of energy depositions with transverse energy above 35 GeV and 25 GeV for the leading (highest transverse energy) and subleading (second highest transverse energy) cluster.

Jet candidates are reconstructed [18] from three-dimensional topological energy clusters in the EMCAL and hadronic calorimeter (HCAL) using the anti-k_t jet algorithm [19] with a radius parameter of 0.4, and these are referred to as "small-R jets". After energy calibration [20], all jet candidates are required to have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. To reduce the effect of pile-up, an additional requirement is made on the jet vertex tagger (JVT) discriminant [21] for jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$.

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector. The positive x-axis is defined by the direction from the IP to the centre of the LHC ring, with the positive y-axis pointing upwards, while the beam direction defines the z-axis. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity η is defined in terms of the polar angle θ by $\eta = -\ln\tan(\theta/2)$. Unless stated otherwise, the angular distance is defined as $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Transverse momentum and energy are defined as $p_{\mathrm{T}} = p\sin\theta$ and $E_{\mathrm{T}} = E\sin\theta$, respectively.

Table 1: Preselection requirements for electron, muon, photon, hadronically decaying τ-lepton, small-R and large-R jet candidates within the 13 TeV ATLAS Open Data. Reconstruction (rec.), identification and isolation criteria of the objects, as well as other additional requirements are given [14, 15, 17, 18, 24].

Electron (e)	Muon (μ)	Photon (γ)
InDet & EMCAL rec.	InDet & MS rec.	InDet & EMCAL rec.
loose identification	loose identification	tight identification
loose isolation	loose isolation	loose isolation
$p_{\mathrm{T}} > 7$ GeV	$p_{\mathrm{T}} > 7$ GeV	$E_{\mathrm{T}} > 25$ GeV
$\lvert\eta\rvert < 2.47$	$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.37$

Hadronically decaying τ-leptons (τ_h)	Small-R jets	Large-R jets
InDet & EMCAL rec.	EMCAL & HCAL rec.	EMCAL & HCAL rec.
medium identification	anti-k_t, R = 0.4	anti-k_t, R = 1.0
$p_{\mathrm{T}} > 20$ GeV	$p_{\mathrm{T}} > 20$ GeV	$p_{\mathrm{T}} > 250$ GeV
$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.0$
1 or 3 associated tracks	b-tagging (MV2c10)	trimming: $R_{\mathrm{sub}} = 0.2$, $f_{\mathrm{cut}} = 0.05$

Identification of jets containing b-hadrons (b-tagging) is performed with a multivariate discriminant, MV2c10, making use of track impact parameters, the b- and c-hadron flight paths inside the jet and reconstructed secondary vertices [22]. For each jet, a value for the multivariate b-tagging discriminant is calculated, and the jet is considered b-tagged if this value is above a given threshold: the so-called algorithm "working point" (WP), which corresponds to a given efficiency [23] to tag a jet containing a b-hadron.

The reconstruction of the visible decay products of hadronic τ-lepton decays [24] starts with a reconstructed jet required to have $p_{\mathrm{T}} > 10$ GeV and $\lvert\eta\rvert < 2.5$. As in the case of electron reconstruction, the transition region between the barrel and endcap calorimeters, $1.37 < \lvert\eta_{\mathrm{cluster}}\rvert < 1.52$, is excluded. The τ_h candidates are required to have $p_{\mathrm{T}} > 20$ GeV and to pass the "medium" identification requirement [25] and have one or three associated tracks and an absolute electric charge of one. Their energy is reconstructed by multivariate regression techniques using information about the associated tracks and calorimeter clusters, as well as the average number of collisions recorded.

Large-R jets are built from three-dimensional topological clusters of energy in the calorimeters, calibrated with the local cluster weighting procedure [26], using the anti-k_t algorithm with a radius parameter of $R = 1.0$. A trimming algorithm [27] is applied to large-R jets to mitigate the impact of initial-state radiation, underlying-event activity and pile-up, with the goal of improving the mass resolution. Trimmed large-R jets are considered if they fulfil $\lvert\eta\rvert < 2.0$ and $p_{\mathrm{T}} > 250$ GeV.

The above requirements are referred to as "preselection" and are summarised in Table 1 for electron, muon, photon, hadronically decaying τ-lepton, small-R and large-R jet candidates. In addition, several data quality criteria ensure that the detector was functioning properly and events are rejected if they contain reconstructed jets associated with energy deposits that can arise from hardware problems, beam-halo events or cosmic-ray showers. Furthermore, events are required to have at least one reconstructed vertex with two or more associated tracks with $p_{\mathrm{T}} > 0.4$ GeV.

Table 2: Final-state collections as provided within the 13 TeV ATLAS Open Data release. The corresponding multiplicity of final-state objects, minimum p_{T} requirement and collection name are shown.

Final-state categories	Leading object p_{T} (min) [GeV]	Collection name
$N_\ell = 1$	25	1lep
$N_\ell \geq 2$	25	2lep
$N_\ell = 3$	25	3lep
$N_\ell \geq 4$	25	4lep
$N_{\mathrm{largeRjet}} \geq 1$ & $N_\ell = 1$	250 (large-R jet), 25 (lepton)	1largeRjet1lep
$N_{\tau-\mathrm{had}} = 1$ & $N_\ell = 1$	20 (τ_h), 25 (lepton)	1lep1tau
$N_\gamma \geq 2$	35	GamGam

If multiple vertices are reconstructed, the vertex with the largest sum of the squares of the transverse momenta of associated tracks is taken as the primary vertex [28]. The tracks associated with the signal electrons and muons must match the primary vertex of the event: the longitudinal impact parameter z_0 is required to satisfy $|z_0 \sin\theta| < 0.5$ mm, where θ is the polar angle of the track. The transverse impact parameter significance $|d_0|/\sigma(d_0)$ must be less than 5 for electrons and 3 for muons.

Furthermore, the preselected events are classified into seven different categories, labelled with the multiplicity of final-state objects in an event (e.g. number of light leptons,[2] N_ℓ, number of hadronically decaying τ-leptons, $N_{\tau-\mathrm{had}}$, number of photons, N_γ, and number of large-R jets, $N_{\mathrm{largeRjet}}$) and combined into different "final-state" collections, as summarised in Table 2. The selected events classified into these separate final-state collections are provided in a simplified data format reducing the information content of the original data analysis format used within the ATLAS Collaboration. The resulting format is a ROOT [7] tuple with more than 80 branches as detailed in Table 4 and Table 5. The layout is optimised to reduce the complexities encountered in a full-scale analysis, emphasising the educational character of the dataset.

A set of simulated SM processes includes $t\bar{t}$, single-top, W+jets, Z+jets, diboson and SM Higgs production. The basic set of SM processes is complemented by simulations of BSM processes (heavy Z' and SUSY production). A summary of the used signal and background samples is shown in Table 6. A total of about 150 GB of storage is needed for all collision and simulated 13 TeV ATLAS Open Data.

3 13 TeV ATLAS Open Data physics analysis examples

The general aim of the 13 TeV ATLAS Open Data and tools released is to provide a straightforward interface to replicate the procedures used by high-energy-physics (HEP) researchers and enable users to experience the analysis of particle physics data in educational environments. Therefore, it is of significant interest to check the correct modelling of several SM process by the 13 TeV ATLAS Open Data MC simulation. This section is devoted to illustrate the examples of physics analysis using the 13 TeV ATLAS Open Data inspired and following as closely as possible the procedures and selections taken in already published ATLAS Collaboration physics results.

[2] In the following, lepton refers to an electron or muon. Leptonic τ-lepton decays ($\tau_\ell \to \ell\nu\bar{\nu}$) are not reconstructed directly; however, isolated electron and muon by-products of the leptonically decaying τ-leptons might be considered.

Table 3: Summary of standard selection requirements applied for each of the reconstructed physics objects in the analysis examples. Additional requirements are placed on the calorimeter and tracking isolation of electrons, muons and photons. The calorimeter isolation (etcone20) is computed as the sum of the energies of calorimeter energy clusters in a cone of size $\Delta R = 0.2$ around the object, while the track isolation (ptcone30) is defined as the scalar sum of the p_{T} of tracks within a cone of size $\Delta R = 0.3$ around the object. Since large-R jets with $m < 50$ GeV or $p_{\mathrm{T}} > 1500$ GeV are outside of a well-calibrated region of phase-space, they are excluded from the selection.

Electrons & Muons	Small-R jets	Photons	Large-R jets	τ_h
$p_{\mathrm{T}} > 25$ GeV lep_ptcone30 < 0.15 lep_etcone20 < 0.15	$p_{\mathrm{T}} > 25$ GeV JVT > 0.59	 photon_ptcone30 < 0.065 photon_etcone20 < 0.065	$p_{\mathrm{T}} < 1500$ GeV mass > 50 GeV	$p_{\mathrm{T}} > 25$ GeV

Twelve analyses, grouped into different final-states, are introduced:

- *four high statistics* analyses with a selection of $W^{\pm} \to \ell^{\pm}\nu$ leptonic-decay events, single-Z-boson events, where the Z boson decays into an electron–positron or muon–antimuon pair, or into a τ-lepton pair with a hadronically decaying τ-lepton accompanied by a τ-lepton that decays leptonically, and top-quark pairs in the $t\bar{t} \to W^{+}W^{-}b\bar{b} \to \ell\nu q\bar{q}^{'}b\bar{b}$ final state. Each of these analyses have sufficiently high event yields to study the SM processes in detail, and are intended to show the general good agreement between the released 13 TeV data and MC prediction. They also enable the study of SM observables, such as the mass of the W and Z bosons, and that of the top quark;
- *three low statistics* analyses with a selection of single top-quarks produced in the t-channel in the $t + q \to W^{\pm}b + q \to \ell^{\pm}\nu b + q$ final state, diboson events with $W^{\pm}Z \to \ell^{\pm}\nu\ell^{'+}\ell^{'-}$ and $ZZ \to \ell^{+}\ell^{-}\ell^{'+}\ell^{'-}$ fully-leptonic final states. These analyses illustrate the statistical limitations of the released dataset given the low production cross-section of the rare processes, where the variations between data and MC prediction are attributed to sizeable statistical fluctuations;
- *three SM Higgs boson* analyses with a selection of events in the $H \to WW^{*} \to e\nu\mu\nu$, $H \to ZZ^{*} \to 4\ell$ and $H \to \gamma\gamma$ decay channels, which serve as examples to implement simplified analyses in different final-state scenarios and "re-discover" the production of the SM Higgs boson;
- *two BSM physics* analyses searching for new hypothetical particles: one implementing the selection criteria of a search for direct production of superpartners of SM leptons, and the second one implementing the selection criteria of a search for new heavy particles that decay into top-quark pairs, provided to implement a simplified analysis for searching for new physics using different physics objects.

An additional object selection, summarised in Table 3, is applied on top of the preselection described in Table 1 for each of the reconstructed physics objects. It provides 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.

3.1 Single-lepton final state: the case of SM W-boson production

W bosons are produced abundantly at the LHC and the measurements of the inclusive production cross section of the W bosons and of the asymmetry between the W^+ and W^- cross sections constitute important tests of the SM. In addition, W+jets processes are a significant background to studies of SM processes such as single-top production, discussed in Section 3.2, top-quark pair ($t\bar{t}$) production, described in Section 3.3, as well as searches for the SM Higgs boson and for BSM physics.

The analysis presented is focused on implementing the selection criteria for $W^{\pm} \to \ell^{\pm}\nu$ leptonic decays (with $\ell = e$ or μ), and the selection is loosely based on the W-boson production cross-section measurement carried out with early $\sqrt{s} = 13$ TeV data with the ATLAS detector [29]. Standard object-selection criteria (defined in Table 3) are applied, with a *stricter* lepton p_{T} (> 35 GeV) and lepton calorimeter and tracking isolation (< 0.1) requirements and and *tight* lepton identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly one light lepton (electron or muon) with $p_{\mathrm{T}} > 35$ GeV;
- Missing transverse momentum[3] larger than 30 GeV;
- The transverse mass of the W-boson[4] ($M_{\mathrm{T}}^{\mathrm{W}}$): $M_{\mathrm{T}}^{\mathrm{W}} > 60$ GeV.

The distributions of the pseudorapidity, azimuthal angle, transverse momentum, and energy of the selected lepton, as well as the number of jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and $M_{\mathrm{T}}^{\mathrm{W}}$ in the $W \to e\nu$ and $W \to \mu\nu$ selections are shown in Figure 1. Good agreement within statistical uncertainties is seen between data and MC prediction.

Sources of background in the single-lepton final state are events involving semileptonic decays of heavy quarks, hadrons misidentified as leptons, and, in the case of the electron channel, electrons from photon conversions (all referred to collectively as "multijet background"), usually estimated using data-driven techniques [31]. Given that no data-driven estimation of the multijet background is attempted, stricter lepton identification, isolation and p_{T} requirements, as well as tight cuts on $E_{\mathrm{T}}^{\mathrm{miss}}$ and $M_{\mathrm{T}}^{\mathrm{W}}$ distributions are applied in order to reduce its contribution. Any residual disagreement might be understood as a sign that the multijet contribution to the electron and muon channels, which predominantly populate the low $E_{\mathrm{T}}^{\mathrm{miss}}$ and transverse mass regions, are not taken into account.

3.2 Single-lepton final state: the case of t-channel single-top-quark production

At hadron colliders, top quarks are predominantly produced in pairs via the flavour conserving strong interaction, but single top-quark production can occur via charged current electroweak processes involving a Wtb vertex. At leading order (LO) in quantum chromodynamics (QCD) perturbation theory, three sub-processes contribute to single top-quark production: an exchange of a virtual W boson either in the t-channel or in the s-channel, or the associated production of a top quark with an on-shell W boson (Wt).

[3] The missing transverse momentum in the event, whose magnitude will be denoted in the following by $E_{\mathrm{T}}^{\mathrm{miss}}$, is defined as the negative vector sum of the p_{T} of reconstructed and calibrated objects in the event [30].

[4] $M_{\mathrm{T}}^{\mathrm{W}} = \sqrt{2p_{\mathrm{T}}^{\ell}E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\Delta\phi)}$, where p_{T}^{ℓ} is the transverse momentum of the lepton and $\Delta\phi$ is the azimuthal angle separation between the lepton and the direction of the missing transverse momentum.

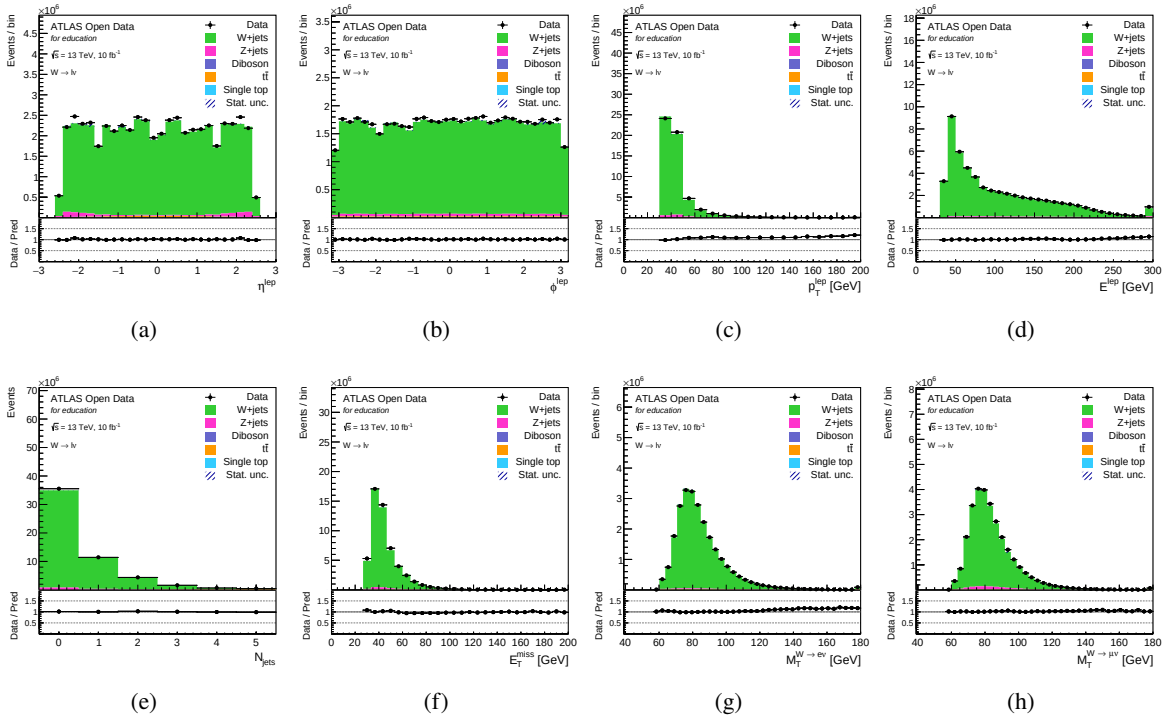


Figure 1: Comparison between data and MC prediction for several distributions in the $W \to \ell\nu$ selection. Distributions of (a) pseudorapidity η, (b) azimuthal angle ϕ, (c) transverse momentum p_T and (d) energy E of the selected lepton, as well as the (e) number of jets, (f) missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, and transverse mass in the (g) $W \to e\nu$ and (h) $W \to \mu\nu$ selections are shown. The points represent experimental data. The filled histograms show the prediction from different MC simulations. The contributions are stacked. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram.

In pp collisions, the t-channel exchange is the dominant production process of single top quarks: an exchange of a space-like W boson due to the interaction of a light quark with a b-quark produces a top quark and a forward light-quark (called the spectator quark) in the final state.

The analysis presented considers only W-boson decay modes involving an electron or a muon, and the selection is loosely based on previous measurements of the single-top-quark t-channel production at $\sqrt{s} = 8$ and 13 TeV with the ATLAS detector [32, 33]. Standard object-selection criteria (see Table 3) are applied, with a *stricter* lepton p_{T} (> 35 GeV) and lepton calorimeter and tracking isolation (< 0.1) requirements and *tight* lepton identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly one lepton (electron or muon) with $p_{\mathrm{T}} > 35$ GeV;
- Missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ larger than 30 GeV;
- Transverse mass of the W-boson $M_{\mathrm{T}}^{\mathrm{W}}$ larger than 60 GeV;
- Exactly two jets with $p_{\mathrm{T}} > 30$ GeV, with exactly one of them b-tagged (MV2c10 @ 70% WP) [22];

- The pseudorapidity of the untagged jet must satisfy $|\eta| > 1.5$;
- The separation in η between the untagged jet and the b-tagged jet must be larger than 1.5;
- The scalar sum (H_{T}) of the p_{T} of the lepton, the p_{T} of the jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ must be larger than 195 GeV;
- The mass of the lepton and the b-tagged jet, m_{lb}, must be larger than 150 GeV.

The largest background contributions to t-channel single-top-quark production arise from $t\bar{t}$ and W+jets production. The latter contributes to the background if there is a b-quark in the final state or due to mistagging of jets containing other quark flavours. Other minor backgrounds originate from Wt, s-channel single-top-quark, Z + jets and diboson production. No multijet background estimation is attempted and the contributing effects of fake or non-prompt leptons are considered to be negligible.

The distributions of the pseudorapidity and transverse momentum of the untagged and b-tagged jets, as well as the $E_{\mathrm{T}}^{\mathrm{miss}}$, $M_{\mathrm{T}}^{\mathrm{W}}$, H_{T} and m_{lb} are shown in Figure 2. Good agreement within statistical uncertainties is found between data and prediction.

3.3 Single-lepton final state: the case of top-quark pair production

The top quark is the heaviest elementary particle in the SM, with a mass m_t of around 172.5 GeV [34], which is close to the electroweak symmetry breaking scale. At the LHC, top quarks are primarily produced in quark–antiquark pairs ($t\bar{t}$), and due to its large production cross section ($\approx$ 830 pb at $\sqrt{s}$ = 13 TeV), the LHC can be viewed as "a top-quark factory". Top quarks have a rich phenomenology which includes high-p_{T} jets, b-jets, leptons and missing transverse momentum. Its understanding is crucial for studying rarer processes, given that $t\bar{t}$ production is a background to virtually all processes having leptons and multiple jets in their final states. This is illustrated in the case of the search for new heavy particles that decay into top-quarks pairs, described in Section 3.11.

The analysis presented targets the lepton+jets final state, where one W boson originating from the top-quark decays leptonically and the other W boson decays hadronically, i.e. $t\bar{t} \to W^+W^-b\bar{b} \to \ell\nu q\bar{q}'b\bar{b}$. The selection is loosely based on the top-quark-pair differential cross-section measurement carried out with $\sqrt{s}$ = 13 TeV data recorded in 2015 with the ATLAS detector [35]. Standard object-selection criteria described in Table 3 are applied, with a *stricter* lepton p_{T} requirement and *tight* lepton identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly one lepton (electron or muon) with $p_{\mathrm{T}} > 30$ GeV;
- Missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ larger than 30 GeV;
- Transverse mass of the W-boson $M_{\mathrm{T}}^{\mathrm{W}}$ larger than 30 GeV;
- At least four jets with $p_{\mathrm{T}} > 30$ GeV, out of which at least two are b-tagged (MV2c10 @ 70% WP).

The dominant source of background comes from SM single-top-quark production and the production of W bosons in association with jets, and the shapes of all background distributions are modelled with MC simulation. The contribution of multijet background is considered to be negligible.

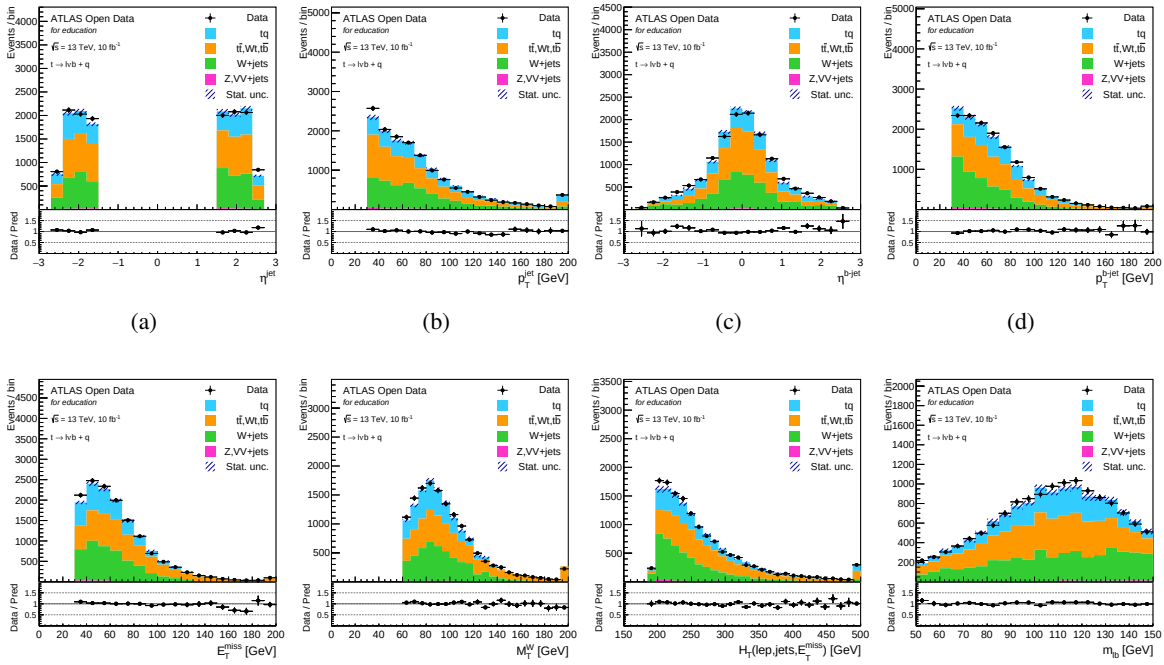


Figure 2: Comparison between data and MC prediction for several distributions in the $t + q \to Wb + q \to \ell\nu b + q$ selection. Distributions of the untagged jet (a) pseudorapidity η and (b) transverse momentum p_T, b-tagged jet (c) pseudorapidity η and (d) transverse momentum p_T, (e) missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, (f) transverse mass of the W-boson $M_{\mathrm{T}}^{\mathrm{W}}$, (g) scalar sum H_{T} and (h) mass of the lepton and the b-tagged jet m_{lb} are shown. The points represent experimental data. The filled histograms show the prediction from different MC simulations. The t-channel signal is shown independently from the other single-top quark production channels labelled tq. The label $t\bar{b}$ refers to the s-channel single-top-quark processes, and VV to diboson production, respectively. The contributions are stacked. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram.

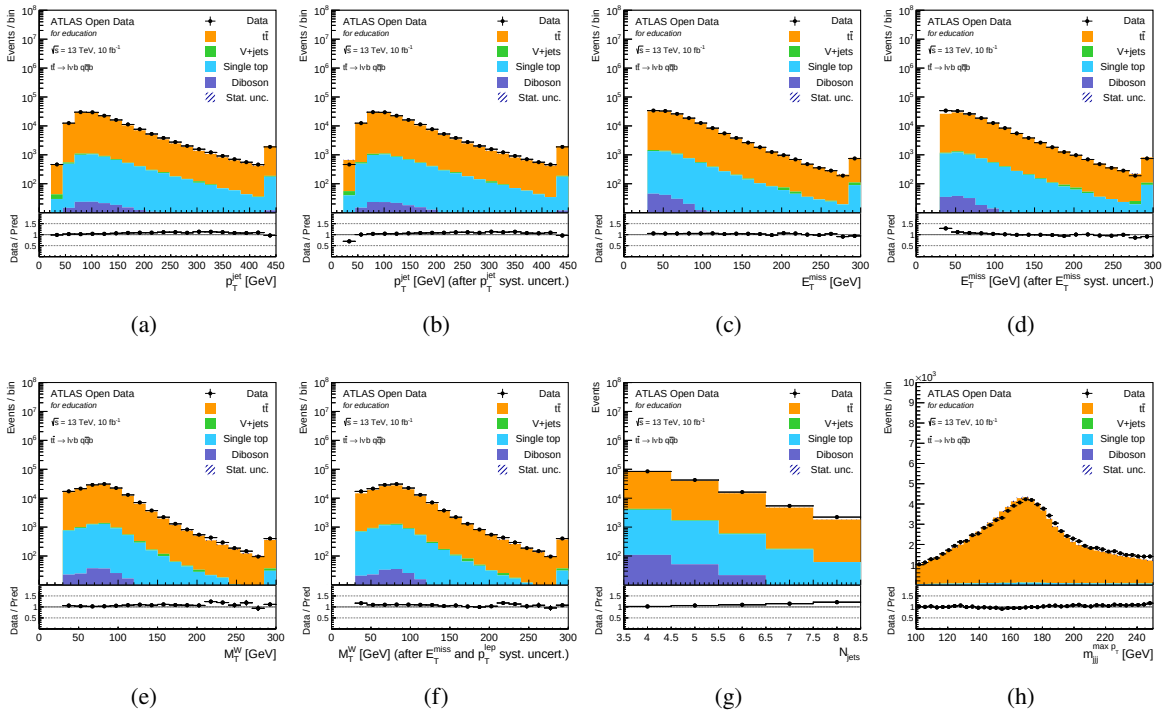


Figure 3: Comparison between data and MC prediction for several distributions in the $t\bar{t} \to W^+W^-b\bar{b} \to \ell\nu q\bar{q}'b\bar{b}$ selection. Distributions of the leading selected jet transverse momentum (a) before and (b) after a jet energy scale systematic uncertainty variation, missing transverse momentum (c) before and (d) after $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution systematic uncertainty variation, transverse mass of the W-boson (e) before and (f) after the $E_{\mathrm{T}}^{\mathrm{miss}}$ and lepton scale and resolution systematic uncertainty variations, (g) number of selected jets and (h) invariant mass of the three-jets combination with the highest vector p_{T} sum are shown. The points represent experimental data. The filled histograms show the prediction from different MC simulations. The contributions are stacked. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram.

Several sources of systematic uncertainties related to the reconstruction and the physics objects affect the $t\bar{t}$ single-lepton final-state measurements. Hence, a simplified single-component systematic uncertainty reflecting: lepton momentum scale and resolution performance for electrons and muons [15, 36], jet energy scale uncertainty [37] and missing-transverse-energy scale and resolution [30] is added to the 13 TeV ATLAS Open Data set.

The distributions of: the leading-jet transverse momentum before and after a systematic uncertainty variation; $E_{\mathrm{T}}^{\mathrm{miss}}$ and W-boson $M_{\mathrm{T}}^{\mathrm{W}}$ before and after a systematic uncertainty variation on the objects; number of selected jets and the invariant mass of the three-jets combination with the highest vector p_{T} sum, an observable that roughly approximates the mass of the top quark [38], are shown in Figure 3. Data agree well with the SM expectation within the statistical uncertainties.

3.4 Two-lepton final state: the case of SM Z-boson production

The study of Z-boson production in pp collisions provides a stringent test of perturbative QCD. In addition, the SM Z-boson process, often produced in association with one or more jets, is a significant background to searches for the SM Higgs boson and for physics beyond the SM (as in the case described in Section 3.6). It is therefore of vital to check the modelling of this process by the ATLAS Open Data MC simulation.

The analysis presented implements the selection criteria for single Z-boson events, where the Z boson decays into an electron–positron or muon–antimuon pair ($Z \to e^+e^-$ and $Z \to \mu^+\mu^-$). The selection is loosely based on the Z-boson production cross-section measurement carried out with early $\sqrt{s} = 13$ TeV data with the ATLAS detector [29]. Standard object-selection criteria (see Table 3) are applied, and leptons are required to pass *tight* identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly two same-flavour opposite-charge (SFOS) leptons (electrons or muons) with $p_{\mathrm{T}} > 25$ GeV;
- Dilepton invariant mass is required to be within: $66 < m_{\ell\ell} < 116$ GeV.

The background contribution to the electron and muon channels from SM processes is estimated using MC simulated samples. No data-driven estimation of the multijet background is attempted.

The distributions of the pseudorapidity, azimuthal angle, transverse momentum, energy, electric charge, identification code of the leading selected lepton, as well as the dilepton invariant mass for the $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ selections are shown in Figure 4. The data and simulation agree within statistical uncertainties.

3.5 Two-lepton final state: the case of SM Higgs boson production in the $H \to WW^*$ decay channel

The SM of particle physics postulates the existence of a complex scalar doublet with a vacuum expectation value, which spontaneously breaks the electroweak symmetry, gives masses to all the massive elementary particles in the theory, and gives rise to a physical scalar known as the Higgs boson. Since the observation of a new particle by the ATLAS [40] and CMS [41] Collaborations in the search for the SM Higgs boson, the properties of the new particle have been measured and the results are consistent with SM predictions. The SM Higgs boson mass has been measured to be $m_H = 125.09 \pm 0.24$ GeV [42] by combining ATLAS and CMS measurements.

The $H \to WW^*$ decay branching ratio for $m_H = 125$ GeV is predicted to be 0.214 [43] in the SM, and corresponds to the second-largest branching fraction after the dominant $H \to b\bar{b}$ decay mode. The predicted Higgs-boson production cross sections via the dominant gluon–gluon fusion (ggF) and vector-boson fusion (VBF) times $H \to WW^*$ branching fraction are 10.4 pb and 0.81 pb for ggF and VBF [44], respectively. Reducing the numerous backgrounds contributing to this channel and accurately estimating the remainder is a major challenge in this analysis. The dominant background stems from non-resonant WW diboson production, while $t\bar{t}$, single-top-quark and W+jets (with the jet misidentified as a lepton) events, as well as non-resonant WZ and ZZ processes contribute to the overall background.

The analysis presented implements the criteria for the selection of the $H \to WW^*$ decay channel, with one of the W-bosons decaying to an electron and a neutrino, and the other W-boson decaying to a muon and a neutrino.

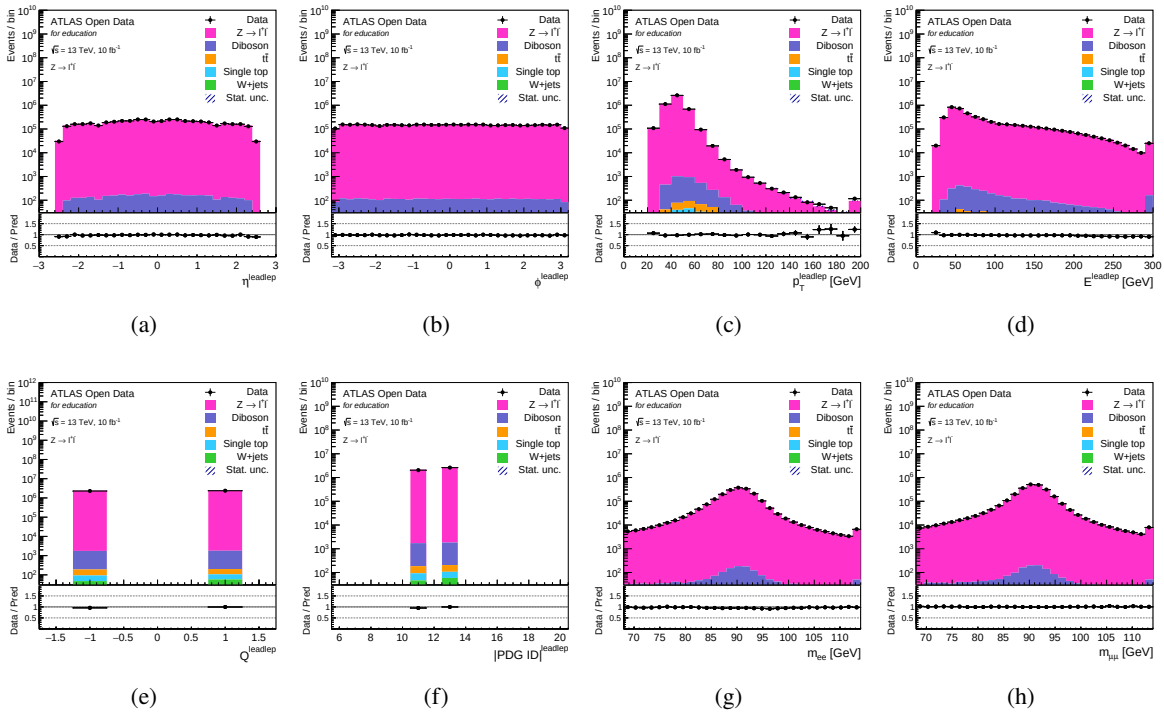


Figure 4: Comparison between data and MC prediction for several distributions in the $Z \to \ell\ell$ selection. Distributions of (a) pseudorapidity η, (b) azimuthal angle ϕ, c) transverse momentum p_{T}, (d) energy E, (e) electric charge Q and (f) PDG identification code [39] of the leading selected lepton, as well as the dilepton invariant mass $m_{\ell\ell}$ from the (g) $Z \to e^+e^-$ and (h) $Z \to \mu^+\mu^-$ selections are shown. The points represent experimental data. The filled histograms show the prediction from different MC simulations. The contributions are stacked. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram.

The selection is loosely based on the Higgs-boson production cross-section measurement in the $H \to WW^* \to e\nu\mu\nu$ decay channel at $\sqrt{s} = 13$ TeV and recorded by the ATLAS detector in 2015 and 2016 [45]. Standard object-selection criteria defined in Table 3 are applied, with a *stricter* lepton calorimeter and tracking isolation (< 0.1) requirements and *tight* lepton identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly two isolated, different-flavour opposite-sign leptons (electrons or muons) with $p_{\mathrm{T}} > 22$ and 15 GeV, respectively;
- Missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ larger than 30 GeV;
- Exactly zero or at most one jet with $p_{\mathrm{T}} > 30$ GeV, and exactly zero b-tagged jets (MV2c10 @ 85% WP) with $p_{\mathrm{T}} > 20$ GeV;
- Azimuthal angle between $E_{\mathrm{T}}^{\mathrm{miss}}$ and the dilepton system $\Delta\phi(\ell\ell, E_{\mathrm{T}}^{\mathrm{miss}}) > \pi/2$;
- Transverse momentum of the dilepton system $p_T^{\ell\ell} > 30$ GeV;

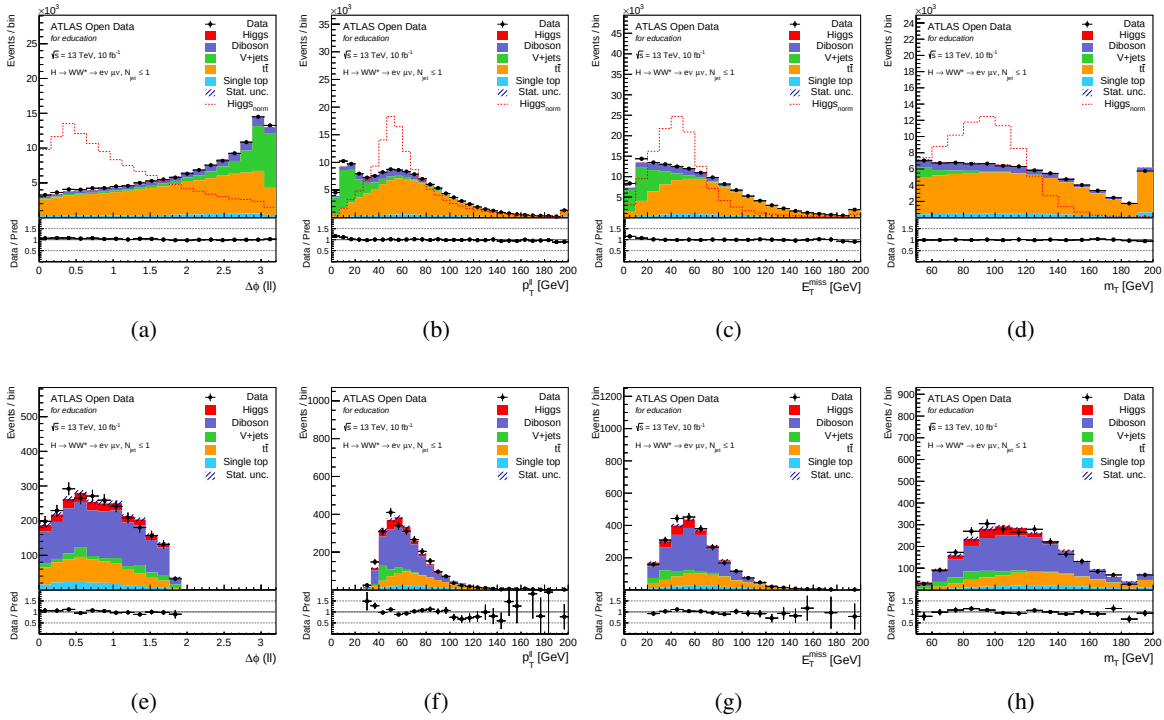


Figure 5: Comparison between data and MC prediction for several distributions in the $H \to WW^* \to e\nu\mu\nu$ selection. Distributions of (a, e) azimuthal angle between the two leptons, (b, f) transverse momentum of the dilepton system, (c, g) missing transverse momentum and (d, h) dilepton transverse mass are shown (a, b, c, d) before and (e, f, g, h) after the final event selection. The points represent experimental data. The filled histograms show the prediction from different MC simulations. Background processes such as $Z+$jets and $W+$jets are denoted V+jets. The dashed red line shows $H \to WW^*$ signal distribution normalised to background yield. The signal yield (solid) expectation for a SM Higgs with $m_H = 125$ GeV is also shown. The contributions are stacked. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram, including the SM Higgs signal.

- The invariant mass of the two leptons $m_{\ell\ell}$ must satisfy: 10 GeV < $m_{\ell\ell}$ < 55 GeV;
- Azimuthal angle between the two leptons $\Delta\phi(\ell,\ell) < 1.8$.

The distributions of the azimuthal angle between the two leptons, transverse momentum of the dilepton system, missing transverse momentum and dilepton transverse mass[5] are shown before and after the final event selection in Figure 5. Good agreement within statistical uncertainties is found between data and MC prediction before selection. After the selection criteria, a small excess in data is observed, and which corresponds to the production of the SM Higgs boson.

[5] The dilepton transverse mass is defined as $m_T = \sqrt{\left(E_{\mathrm{T}}^{\ell\ell} + E_{\mathrm{T}}^{\mathrm{miss}}\right)^2 - \left|\mathbf{p}_{\mathrm{T}}^{\ell\ell} + \mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}\right|^2}$ where $E_{\mathrm{T}}^{\ell\ell} = \sqrt{|\mathbf{p}_{\mathrm{T}}^{\ell\ell}|^2 + m_{\ell\ell}^2}$ and $\mathbf{p}_{\mathrm{T}}^{\ell\ell}$ is the vector sum of the lepton transverse momenta.

3.6 Two-lepton final state: the case of a search for supersymmetric particles

Supersymmetry (SUSY) is one of the most studied extensions of the SM. In its minimal form it predicts new fermionic (bosonic) partners of the fundamental SM bosons (fermions) and an additional Higgs doublet. These new SUSY particles, or *sparticles*, can provide an elegant solution to the gauge hierarchy problem and provide a natural candidate for dark matter in cases where the lightest supersymmetric particle is a so called weekly interacting massive particle.

The analysis presented is focused on implementing the selection criteria of a search for direct production of pairs of sleptons ($\tilde{\ell}$), the superpartners of the SM leptons, where each slepton decays directly into the lightest neutralino ($\tilde{\chi}_1^0$) and the corresponding lepton. A simplified benchmark model is chosen for this search, in which the mass of the slepton and the neutralino are the only free parameters [46]. The predicted cross section for $\tilde{\ell}\tilde{\ell}$ production, each with a mass of 600 GeV, and $\tilde{\chi}_1^0$ mass of 300 GeV is 0.7 fb [47].

The final state consists of two leptons of same flavour and opposite electric charge. The two neutralinos will escape the detector and so will contribute to the missing energy in the event. The dominant backgrounds are processes from Z + jets and diboson production, and dileptonic $t\bar{t}$ and Wt events. The selection is loosely based on the search for electroweak production of supersymmetric particles in final states with two leptons at $\sqrt{s}$ = 13 TeV with the ATLAS detector [47]. Standard object-selection criteria (given in Table 3) are applied, with a *loose* lepton p_{T} and *loose* lepton identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly two SFOS leptons (electrons or muons) with $p_{\mathrm{T}} > 25$ and 20 GeV, respectively;
- Dilepton invariant mass, $m_{\ell\ell}$, is required to be larger than 40 GeV;
- Two signal regions (SR) are constructed: loose and tight; both SR are required to have exactly zero b-tagged jets (MV2c10 @ 77% WP) with $p_{\mathrm{T}} > 20$ GeV and exactly zero non-b-tagged jets with $p_{\mathrm{T}} > 60$ GeV;
- Additional cuts on the stransverse mass variable, m_{T2} [48], and dilepton invariant mass, $m_{\ell\ell}$, are applied in the loose SR ($m_{\ell\ell} > 111$ GeV and $m_{T2} > 100$ GeV) and tight SR ($m_{\ell\ell} > 300$ GeV and $m_{T2} > 130$ GeV);

The distributions of the transverse momentum of the leading lepton, dilepton invariant mass, missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ and stransverse mass m_{T2} before and after the two signal-region selections are shown in Figure 6. Good agreement is found between data and MC prediction, even in signal regions where large statistical fluctuations are observed. Interesting to note are the events observed in data at very high values of $E_{\mathrm{T}}^{\mathrm{miss}}$ (> 900 GeV). These have been investigated further and found to contain very high-p_{T} calorimeter-tagged muons which passes the *loose* muon selection working point [15]. By applying the *tight* identification criteria on the muons, these events are rejected.

3.7 Three-lepton final state: the case of SM $W^{\pm}Z$ diboson production

The study of the diboson production is an important part of the physics programme in hadron collisions as it represents an important test of the electroweak sector of the SM. In particular, the $W^{\pm}Z$ diboson production arises from two vector bosons radiated by quarks or from the decay of a virtual W boson into a $W^{\pm}Z$ pair, the latter of which involves a triple gauge coupling.

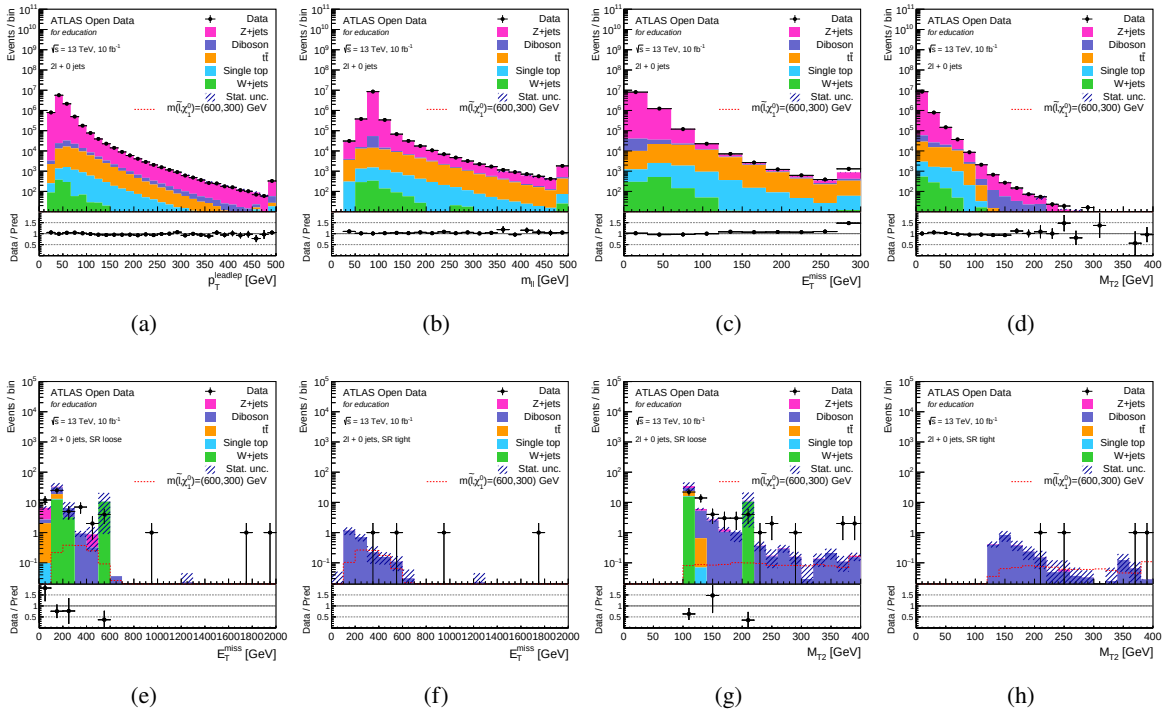


Figure 6: Comparison between data and MC prediction for several distributions in the $\tilde{\ell}\tilde{\ell}' \rightarrow \ell\tilde{\chi}_1^0\ell'\tilde{\chi}_1^{0\prime}$ selection. Distributions of (a) transverse momentum p_T of the leading lepton, (b) dilepton invariant mass $m_{\ell\ell}$, (c) missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, (d) stransverse mass m_{T2} before the signal region selections are shown. Distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the (e) loose SR and (f) tight SR, and of m_{T2} in the (g) loose SR and (h) tight SR are shown. The points represent experimental data. The filled histograms show the prediction from different MC simulations. The contributions are stacked. Simulated signal model for slepton pair production with $m_{\tilde{\ell}} = 600$ GeV and $m_{\tilde{\chi}_1^0} = 300$ GeV is overlaid as dotted red curve for comparison. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram.

The analysis presented is focused on implementing the selection criteria for $W^{\pm}Z \rightarrow \ell^{\pm}\nu\ell'^{+}\ell'^{-}$ leptonic decays, where the Z boson decays into leptons and the W boson decays into a charged lepton and a neutrino, where ℓ and ℓ' can be an electron or a muon. The selection is loosely based on the $W^{\pm}Z$ production cross-section measurement using $\sqrt{s} = 13$ TeV data collected in 2015 by the ATLAS experiment [49]. Standard object-selection criteria (see Table 3) are applied, with a *loose* lepton p_{T} requirement and *tight* lepton identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly three leptons (electrons or muons) with $p_{\mathrm{T}} > 20$ GeV, at least one of them should have $p_{\mathrm{T}} > 25$ GeV;
- Candidate events are required to have at least one SFOS-pair of leptons with an invariant mass that is consistent with the nominal Z-boson mass ($m_Z = 91.18$ GeV) to within 10 GeV, considered to be the Z-boson candidate;

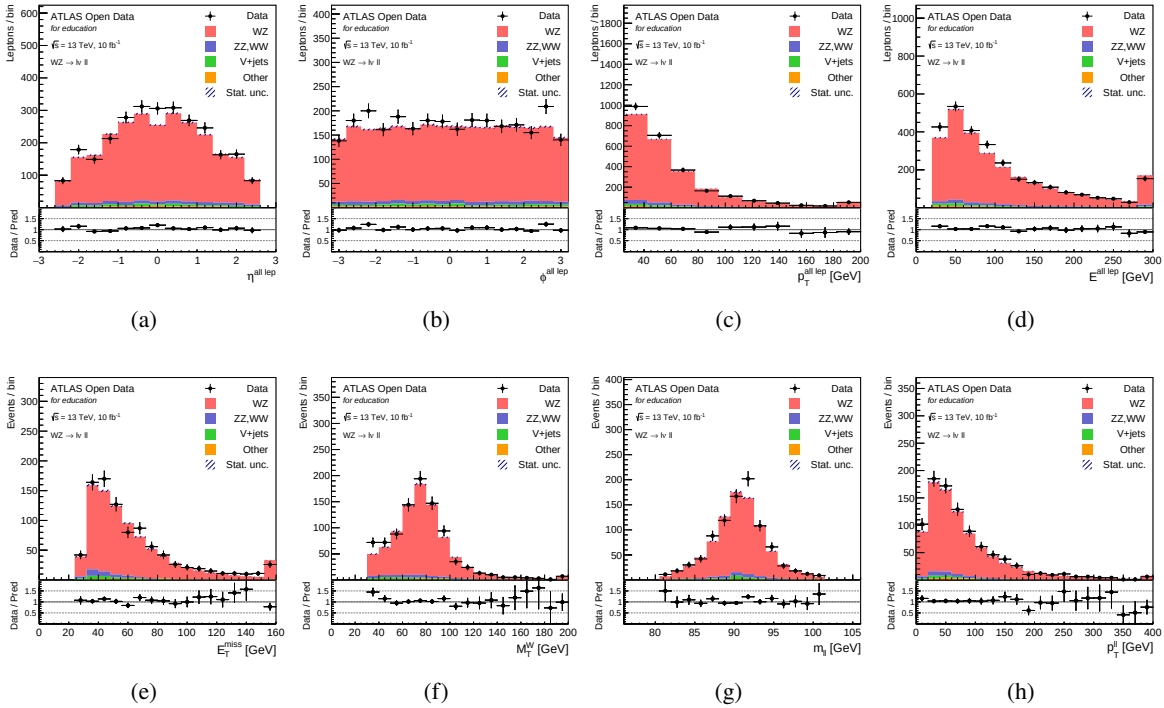


Figure 7: Comparison between data and MC prediction for several distributions in the $W^{\pm}Z \to \ell\nu\ell\ell'$ selection. Distributions of (a) pseudorapidity η, (b) azimuthal angle ϕ, (c) transverse momentum p_T and (d) energy E of the leading selected lepton are shown. Distributions of (e) missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, (f) transverse mass of the W candidate $M_{\mathrm{T}}^{\mathrm{W}}$, (g) invariant mass of the reconstructed Z-boson candidate $m_{\ell\ell}$ and (h) transverse momentum of the reconstructed Z-boson candidate $p_T^{\ell\ell}$ are shown. The points represent experimental data. The filled histograms show the prediction from different MC simulations. The $W^{\pm}Z$ signal is shown independently from the other diboson processes (WW and ZZ). Background processes such as $Z+$jets and $W+$jets are denoted $V+$jets, while $t\bar{t}$ and single-top quark backgrounds are denoted by "Other". The contributions are stacked. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram.

- Both the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ and the transverse mass of the W candidate $M_{\mathrm{T}}^{\mathrm{W}}$ are required to be larger than 30 GeV.

The distributions of the pseudorapidity, azimuthal angle, transverse momentum and energy of the leading selected lepton, as well as $E_{\mathrm{T}}^{\mathrm{miss}}$, $M_{\mathrm{T}}^{\mathrm{W}}$, invariant mass and transverse momentum of the reconstructed Z-boson candidate are shown in Figure 7. Good agreement, within statistical uncertainties, is found between data and MC prediction.

The main background events originate from ZZ diboson and $Z+$jets events. In particular, events from ZZ production survive the current event selection either because one lepton falls outside the fiducial volume or because it falls in the fiducial acceptance of the detector but is not identified. MC simulated events are used to estimate the contribution from background processes with misidentified leptons.

3.8 Four-lepton final state: the case of SM ZZ diboson production

The study of ZZ diboson production in pp interactions at the LHC not only can be used as a test of the electroweak sector of the SM. The SM ZZ production can proceed via a SM Higgs boson propagator, although this contribution is suppressed in the region where both Z bosons are produced on-shell. Hence, non-resonant ZZ diboson production is an important background for searches of the SM Higgs boson with its subsequent decay to ZZ^*, as discussed in Section 3.9.

The analysis presented is focused on implementing the selection criteria for ZZ diboson events where both Z bosons decay to leptons. The selection is loosely based on the ZZ production cross-section measurement using $\sqrt{s}$ = 13 TeV data collected with the ATLAS detector in 2015 [50]. Candidate events are reconstructed in the $ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ decay channel, where ℓ and ℓ' can be an electron or a muon. Standard object-selection criteria (described in Table 3) are applied, with a *loose* lepton p_{T} requirement and *tight* lepton identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly four leptons (electrons or muons) with $p_{\mathrm{T}} > 20$ GeV, at least one of them should have $p_{\mathrm{T}} > 25$ GeV;
- Two pairs of SFOS leptons ($\mu^+\mu^-$ or e^+e^-) are formed, giving rise to three channels: $4e$, 4μ and $2e2\mu$;
- Each SFOS lepton pair must have an invariant mass within a range of $66 < m_{\ell\ell} < 116$ GeV;
- In the $4e$ and 4μ channels, where there are two possible ways to form SFOS lepton pairs, the combination that minimises $|m_{\ell\ell,1} - m_Z| + |m_{\ell\ell,2} - m_Z|$ is chosen, where $m_{\ell\ell,1}$ and $m_{\ell\ell,2}$ are the invariant masses of the two lepton pairs.

The low SM ZZ production cross section (~15.6 pb) results in a very low yield after selection for the current dataset, and thus highlights its statistical limitations. The distributions of the pseudorapidity, azimuthal angle, transverse momentum and energy of the four selected leptons, as well as the invariant masses of the first and second reconstructed Z-boson candidates $m_{\ell\ell}$ and transverse momentum and rapidity of the four-lepton system in the selected events are shown in Figure 8. Good agreement within statistical uncertainties is found between data and ZZ signal prediction. The contribution from other background events from processes with four leptons in the final state is found to be negligible.

3.9 Four-lepton final state: the case of SM Higgs boson production in the $H \to ZZ^*$ decay channel

The search for the SM Higgs boson through the decay $H \to ZZ^* \to 4\ell$, where $\ell = e$ or μ, represents the so called "golden channel" and leads to a narrow four-lepton invariant-mass peak on top a relatively smooth and small background, largely due to the excellent momentum resolution of the ATLAS detector. The Higgs-boson decay branching ratio to the four-lepton final state for $m_H = 125$ GeV is predicted to be 0.0124% [43] in the SM, and the expected cross section times branching ratio for the process $H \to ZZ^* \to 4\ell$ is 2.9 fb for $\sqrt{s}$ = 13 TeV [44]. Hence, based on an integrated luminosity of the current ATLAS Open Data set of 10 fb^{-1}, one expects a total of 29 events to have been produced in the four-lepton final state (before reconstruction and event selection).

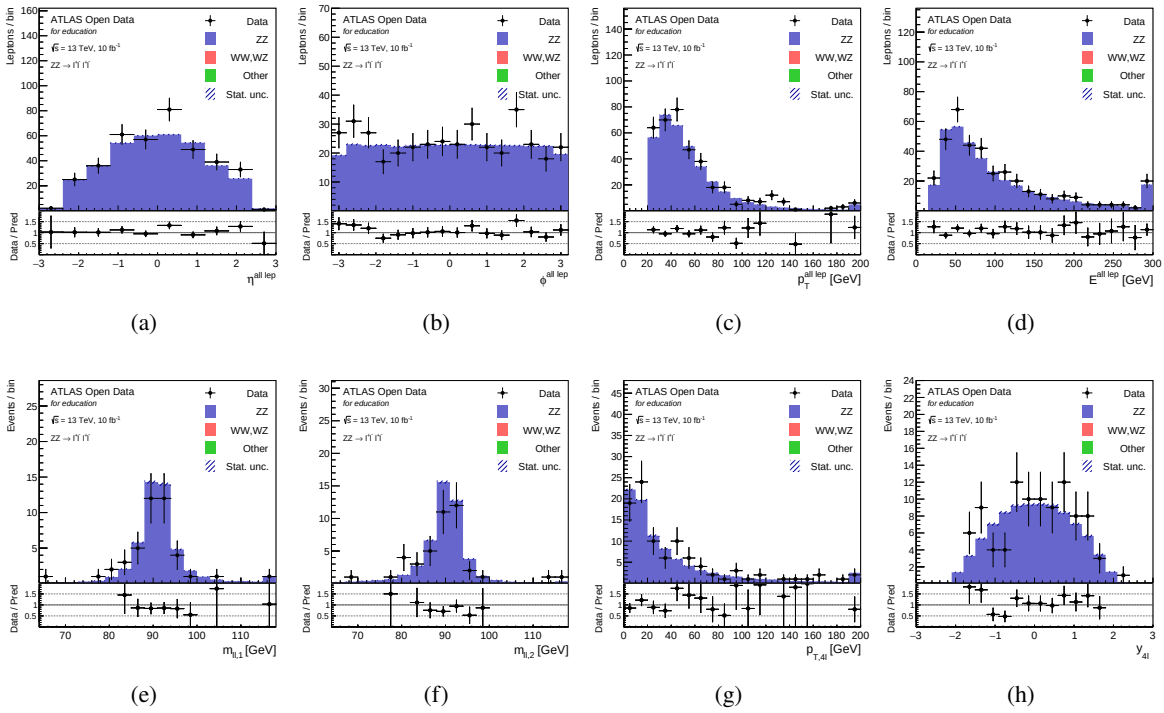


Figure 8: Comparison between data and MC prediction for several distributions in the $ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ selection. Distributions of (a) pseudorapidity η, (b) azimuthal angle ϕ, (c) transverse momentum p_T and (d) energy E of the four selected leptons are shown. Distributions of the invariant masses of the (e) first and (f) second reconstructed Z-boson candidates $m_{\ell\ell}$ and (g) transverse momentum and (h) rapidity of the four-lepton system in selected events. The points represent experimental data. The filled histograms show the prediction from different MC simulations. The contributions are stacked. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram.

Non-resonant SM ZZ production, studied in Section 3.8, can result in four prompt leptons in the final state and constitutes the largest background in this analysis. Other background contributions come from $t\bar{t}$ and Z + jets production (studied in Section 3.3 and Section 3.4, respectively), where charged lepton candidates arise either from decays of hadrons with b- or c-quark content or from misidentification of jets.

The analysis presented is focused on implementing the selection criteria of a search for Higgs-boson candidates by selecting two pairs of isolated leptons, each comprised of two leptons with the same flavour and opposite charge. The selection is loosely based on the $H \to ZZ^* \to 4\ell$ inclusive cross-section measurement carried out by the ATLAS Collaboration using $\sqrt{s}$ = 13 TeV data collected in 2015 and 2016 [51]. Standard object-selection criteria described in Table 3 are applied, with a *loose* lepton p_{T} and *loose* lepton calorimeter- and track-based isolation requirements. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly four leptons (electrons or muons) with $p_{\mathrm{T}} > 25, 15, 10, 7$ GeV, respectively;
- Higgs-boson candidates are formed by selecting two SFOS lepton pairs;

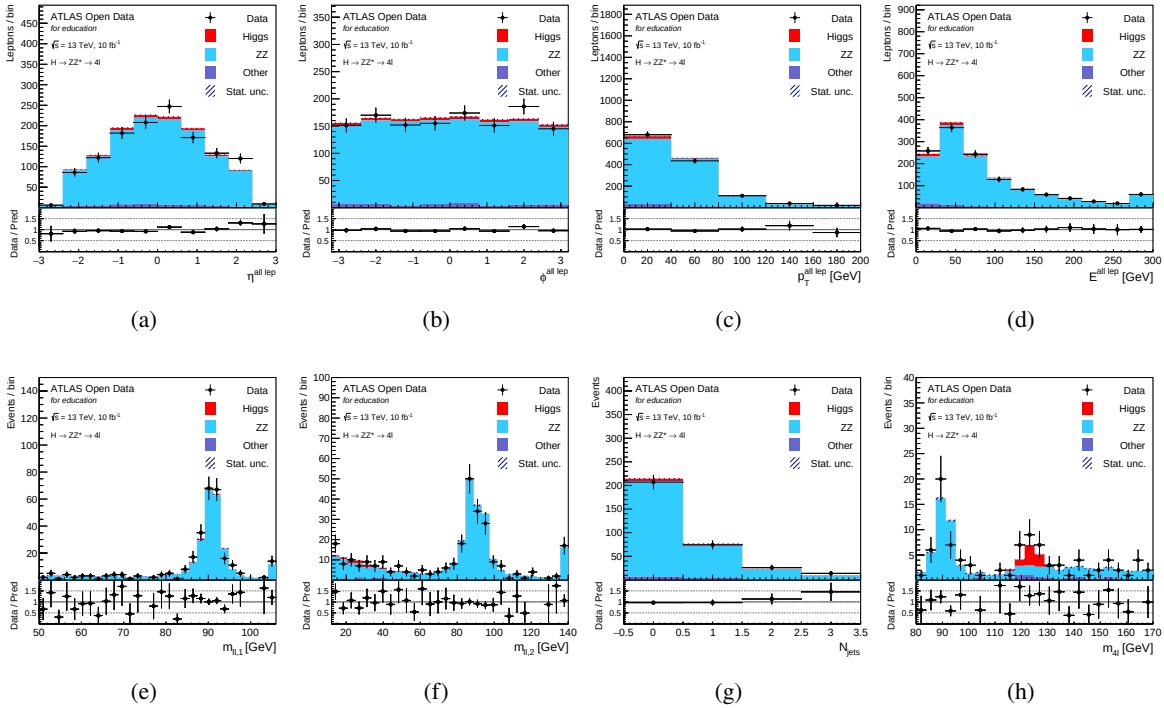


Figure 9: Comparison between data and MC prediction for several distributions in the $H \to ZZ^* \to 4\ell$ selection. Distributions of (a) pseudorapidity η, (b) azimuthal angle ϕ, (c) transverse momentum p_T and (d) energy E of the four selected leptons are shown. Distributions of the invariant masses of the (e) first and (f) second reconstructed Z-boson candidates $m_{\ell\ell}$, (g) jet multiplicity and (h) the four-lepton invariant mass distribution $m_{4\ell}$ in selected events are shown. The points represent experimental data. The filled histograms show the prediction from different MC simulations. The signal expectation for a SM Higgs with $m_H = 125$ GeV is also shown. Minor background processes such as $Z+$jets, $W+$jets, $t\bar{t}$ and single-top quark production are denoted by "Other". The contributions are stacked. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures but (h) contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram, including the SM Higgs signal.

- The leading pair is defined as the SFOS pair with the mass $m_{\ell\ell,1}$ closest to the Z boson mass m_Z, and the subleading pair is defined as the SFOS pair with the mass $m_{\ell\ell,1}$ second closest to m_Z.

The distributions of the pseudorapidity, azimuthal angle, transverse momentum and energy of the four selected leptons, as well as the invariant masses of the first and second reconstructed Z-boson candidates, jet multiplicity and the four-lepton invariant-mass distribution $m_{4\ell}$ are shown in Figure 9. Good agreement within statistical uncertainties is found between measured data and MC background prediction, including the SM Higgs signal expectation. An excess of events in the distribution of the four-lepton invariant mass near $m_{4\ell} = 125$ GeV is observed, which corresponds to the expected in the SM production of the Higgs boson. Further statistical interpretations of the excess and statements regarding the properties of the SM Higgs boson are not attempted and left to the interested reader.

3.10 Two-τ-lepton final state: the case of SM Z-boson production

The τ-leptons play an important role in the physics programme of the LHC. They serve not only as a foundation to identify and precisely measure several SM production processes such as $W \to \tau\nu$ and $Z \to \tau\tau$, which represent the main irreducible backgrounds in measurements of the Higgs boson in $H \to \tau\tau$ final states, but also are widely used in searches for new physics beyond the SM.

The analysis presented is focused on implementing a selection criteria in order to reconstruct the $Z \to \tau\tau$ decays from a hadronically decaying τ-lepton, accompanied by a τ-lepton that decays leptonically. The leptonic τ-lepton decays ($\tau_\ell \to \ell\nu\bar{\nu}$) are reconstructed as electrons and muons in the final state. The hadronic τ-lepton decays ($\tau_h \to$ hadrons ν) produce a highly collimated jet in the detector consisting of an odd number of charged hadrons and possibly additional calorimetric energy deposits from neutral decay products. To discriminate τ_h from jets initiated by light-quarks or gluons, an identification algorithm using multivariate techniques is applied to τ_h candidates [25].

The selection is loosely based on the $Z \to \tau\tau$ production cross-section measurement at $\sqrt{s} = 8$ TeV with the ATLAS detector [52], and follows the object reconstruction from the measurement of the SM Higgs boson in the $H \to \tau\tau$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector [53]. Standard object-selection criteria (defined in Table 3) are applied, with a *stricter* lepton p_{T} and lepton calorimeter and tracking isolation (< 0.1) requirements and *tight* lepton identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly one light lepton (electron or muon) with $p_{\mathrm{T}} > 30$ GeV;
- Exactly one hadronically decaying τ-lepton candidate with $p_{\mathrm{T}} > 25$ GeV;
- The two leptons are required to be of opposite charge;
- The transverse mass calculated from the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the momentum of the selected light lepton is required to have $M_{\mathrm{T}}^{\mathrm{W}} < 30$ GeV;
- The sum of the azimuthal angular separation between the τ_h candidate and the $E_{\mathrm{T}}^{\mathrm{miss}}$ directions, and the lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$ directions is required to be less than 3.5;
- The visible mass of the τ_h candidate and lepton, $m_{\mathrm{vis}} = m(\tau_h, \mathrm{lepton})$, is required to satisfy $35 < m_{vis} < 75$ GeV.

The Z-boson candidate is reconstructed from the visible decay products of the τ-leptons and from the $E_{\mathrm{T}}^{\mathrm{miss}}$, which is assumed to originate from the final-state neutrinos. The di-τ invariant mass $m_{\tau\tau}^{\mathrm{MMC}}$ is determined using the missing-mass calculator (MMC) [54]. The dominant experimental uncertainties affecting the searches with hadronic τ decays are related to the τ_h reconstruction and energy scale, and a simplified τ_h energy-scale uncertainty [24] is added to the 13 TeV ATLAS Open Data set.

The distributions of the pseudorapidity, azimuthal angle, energy, number of associated tracks, transverse momentum before and after a systematic uncertainty variation of the selected hadronic τ-lepton, as well as the visible mass and the di-τ invariant mass are shown in Figure 10. Good agreement within statistical uncertainties is found between data and prediction.

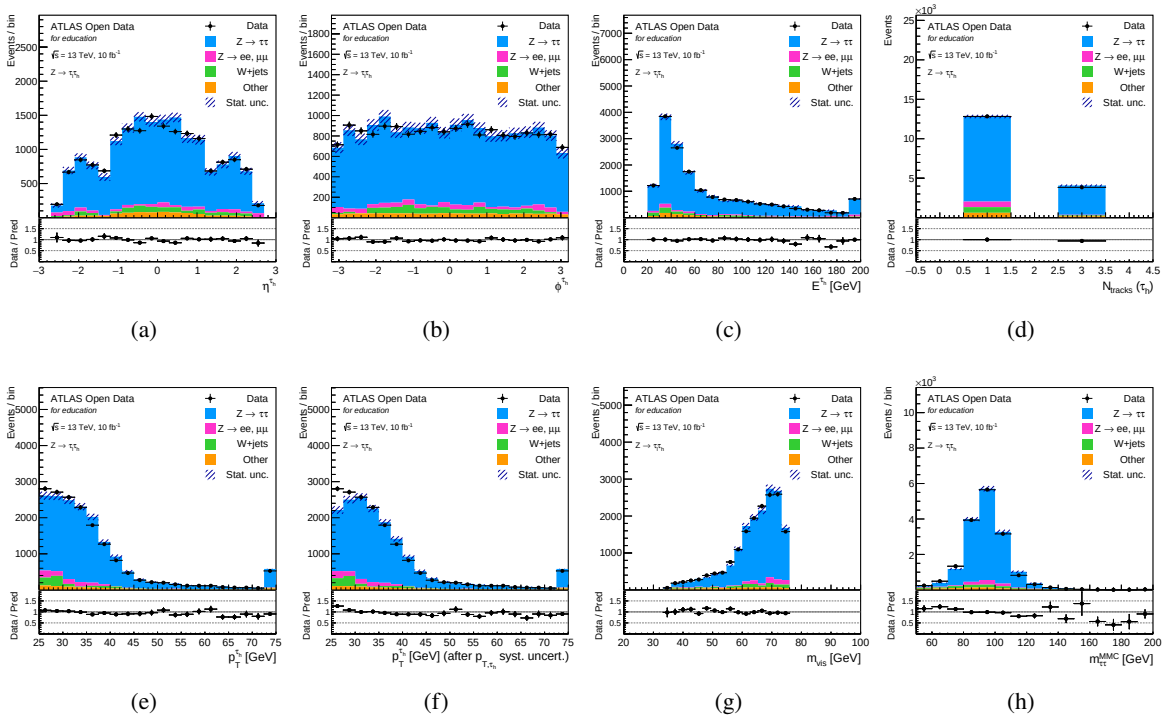


Figure 10: Comparison between data and MC prediction for several distributions in the $Z \to \tau_l\tau_h$ selection. Distributions of (a) pseudorapidity, (b) azimuthal angle, c) energy, (d) number of associated tracks, (e) transverse momentum, (f) transverse momentum after a systematic uncertainty variation of the selected hadronic τ-lepton, (g) visible mass and (h) reconstructed di-τ invariant mass $m_{\tau\tau}^{\text{MMC}}$ are shown. The points represent experimental data. The filled histograms show the prediction from different MC simulations. The $Z \to \tau\tau$ signal is shown independently from the other Z-boson decay channels. Background processes such as $t\bar{t}$, single-top quark and diboson production are denoted by "Other". The contributions are stacked. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram.

3.11 Single-lepton boosted final state: the case of a search for BSM $Z' \to t\bar{t}$

Despite the spectacular phenomenological and experimental success of the SM, searches for new physics phenomena at the LHC are constantly ongoing. As an example, with a mass close to the scale of electroweak symmetry breaking, the top quark, besides having a large coupling to the SM Higgs boson, is predicted to have large couplings to new particles hypothesised in many BSM models.

The analysis presented is focused on implementing the selection criteria of a search for new heavy particles that decay into top-quark pairs in events containing a single charged lepton, large-R jets and missing transverse momentum. A particular benchmark model chosen for this search produces a new gauge boson Z' with a mass of 1 TeV and width of 10 GeV that decays into a $t\bar{t}$-pair [55]. The analysis selection is loosely based on previous searches for heavy particles decaying into top-quark pairs using lepton+jets events in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector [56] and applies a selection referred to as a single-lepton *boosted topology* of the $t\bar{t}$ final state [35].

Large-R jets with radius parameter $R = 1.0$ associated with hadronically decaying top quarks are selected over jets originating from the fragmentation of other quarks or gluons by requiring that they contain several high-p_{T} objects and have a mass compatible with the top-quark mass. Standard object-selection criteria (defined in Table 3) are applied, with a *stricter* lepton p_{T} requirement and *tight* lepton identification criteria. The final event-selection criteria are:

- Single-electron or single-muon trigger satisfied;
- Exactly one lepton (electron or muon) with $p_{\mathrm{T}} > 30$ GeV;
- $E_{\mathrm{T}}^{\mathrm{miss}}$ larger than 20 GeV and $E_{\mathrm{T}}^{\mathrm{miss}} + M_{\mathrm{T}}^{\mathrm{W}} > 60$ GeV;
- At least one small-R jet close to the lepton, i.e. with ΔR(small-R jet, lepton) < 2.0;
- Exactly one large-R jet, passing the simplified top-tagging requirements [57]: a mass larger than 100 GeV and N-subjettiness ratio $\tau_{32} < 0.75$;
- This *top-tagged* large-R jet must be well separated from the lepton, $\Delta\phi$(large-R jet, lepton) > 1.0, and from the small-R jet associated with the lepton, ΔR(large-R jet, small-R jet) > 1.5;
- At least one b-tagged (MV2c10 @ 70% WP) small-R jet that fulfils the following requirements: it is either inside the top-tagged large-R jet, ΔR(large-R jet, b-tagged jet)< 1.0, or it is the small-R jet associated with the lepton.

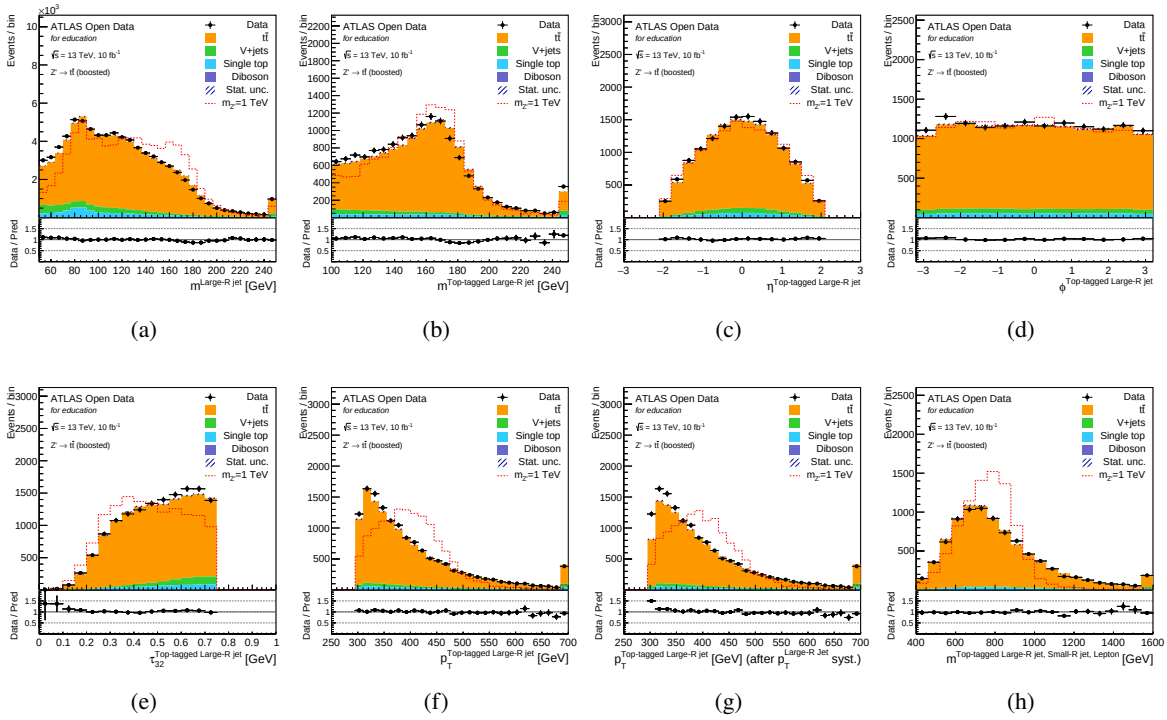
The dominant source of background comes from $t\bar{t}$ production, studied in Section 3.3, and the shapes of all background distributions are modelled with MC simulation. No multijet background estimation is attempted and the contributing effects of fake or non-prompt leptons are considered to be negligible.

An observable that roughly approximates the mass of the $t\bar{t}$ system is constructed [56] by summing the four-momentum of the top-tagged large-R jet, the four-momenta of the charged lepton and the b-tagged small-R jet associated with the lepton. For simplicity, the neutrino momentum is not added. Reconstructing the full $t\bar{t}$ system and studying its properties is possible, but introduces a slight challenge due to the reconstruction of the z component of the neutrino's momentum [58], which may be seen as an interesting educational challenge.

The distributions of the large-R jet mass before and after the top-tagging requirements, top-tagged large-R jet pseudorapidity, azimuthal angle, N-subjettiness, transverse momentum before and after a systematic uncertainty variation and the approximate mass of the $t\bar{t}$ system are shown in Figure 11. Good agreement within statistical uncertainties is found between data and MC prediction. The benchmark 1 TeV Z' model is overlaid, clearly showing the different kinematic properties of this particular BSM prediction.

3.12 Two-photon final state: the case of SM Higgs boson production in the $H \to \gamma\gamma$ decay channel

The $H \to \gamma\gamma$ decay mode provides a very clear and distinctive signature of two isolated and highly energetic photons, and is one of the main channels studied at the LHC. Despite the small branching ratio, $(2.27 \pm 0.07) \times 10^{-3}$ for $m_H = 125$ GeV [43], a reasonably large signal yield can be obtained thanks to the high photon reconstruction and identification efficiency at the ATLAS experiment. Furthermore, due to the excellent photon energy resolution of the ATLAS calorimeter, the signal manifests itself as a narrow peak in the diphoton invariant mass ($m_{\gamma\gamma}$) spectrum on top of a smoothly falling irreducible background from QCD production of two photons.



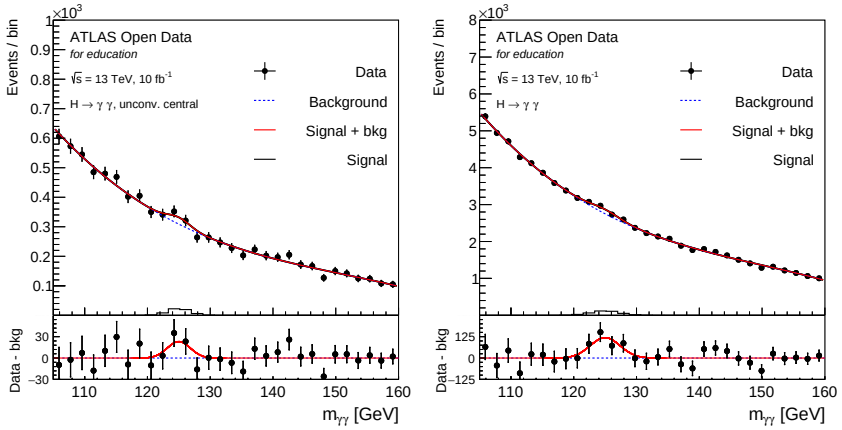
(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 11: Comparison between data and MC prediction for several distributions in the $Z' \to t\bar{t}$ single-lepton boosted selection. Distributions of the large-R jet mass (a) before and (b) after the top-tagging requirements, top-tagged large-R jet (c) pseudorapidity, (d) azimuthal angle, (e) N-subjettiness, (f) transverse momentum, (g) transverse momentum after a systematic uncertainty variation and (h) approximate mass of the $t\bar{t}$ system (neutrino momentum not added). The points represent experimental data. The filled histograms show the prediction from different MC simulations. The contributions are stacked. The benchmark 1 TeV Z' signal prediction is overlaid as dotted red curve. The statistical uncertainty is represented by the error bars on the data points and the hashed area on the MC prediction. The last bin in all figures contains the overflow. The lower panels in each figure show the ratio of the data points to the stacked histogram.

The analysis presented is focused on implementing the selection criteria for $H \to \gamma\gamma$ event candidates and is loosely based on the measurement of Higgs boson in the diphoton decay channel with $\sqrt{s} = 13$ TeV data collected with the ATLAS detector [59]. Standard object-selection criteria (see Table 3) are applied. The final event-selection criteria are:

- Diphoton trigger is satisfied;
- Exactly two photons with $E_{\mathrm{T}} > 35$ and 25 GeV, respectively;
- Leading and subleading photon candidates are respectively required to have $E_{\mathrm{T}}/m_{\gamma\gamma} > 0.35$ and 0.25;
- Diphoton invariant mass $m_{\gamma\gamma}$ between 105 GeV and 160 GeV.



(a) (b)

Figure 12: Diphoton invariant-mass spectrum as observed in the selected diphoton events, with (a) two unconverted central ($|\eta| < 0.75$) photons and (b) inclusively (for both unconverted and converted photons). The solid red curve shows the fitted signal-plus-background model when the Higgs boson mass is constrained to be 125 GeV. The background component of the fit is shown with the dotted blue curve. The signal component of the fit is shown with a solid black curve. The bottom plot shows the residuals between the data and the background component of the fitted model.

The background is estimated from data, without the use of MC simulation, by fitting the diphoton invariant-mass distribution in a range ($105 < m_{\gamma\gamma} < 160$ GeV) with a third-order polynomial function with free shape and normalisation parameters. Signal MC simulations of the five main Higgs-boson production mechanisms (ggF, VBF, WH, ZH, $t\bar{t}H$) are used to model the shape of the invariant mass of the signal, modelled as a Gaussian function.

The distributions of the invariant mass, $m_{\gamma\gamma}$, of the selected diphoton events are shown in Figure 12 in the case of events with two unconverted central ($|\eta| < 0.75$) photons and inclusively (for both unconverted and converted photons). In the inclusive case, a total of 76394 events in the data are selected and approximately 400 signal events for $m_H = 125$ GeV are expected in the diphoton invariant mass range between 105 GeV and 160 GeV. The result of a fit including a signal component fixed to $m_H = 125$ GeV and a background component described by a third-order polynomial is superimposed. An excess, corresponding to the production of the SM Higgs boson, is visible in data near an invariant mass value of 125 GeV.

4 General capabilities and limitations of the released 13 TeV dataset

The publicly released 13 TeV dataset can be used for educational purposes with different levels of task difficulty. At a beginner level, one could visualise the content of the datasets and produce simple distributions. An intermediate-level task would consist of making histograms with collision data after some basic selection, while advanced-level tasks would allow for a deeper look into the ATLAS data, with possibilities of measuring real event properties and physical quantities.

A non-exhaustive list of possible tasks with the proposed datasets include: comparisons of several distributions of event variables for simulated signal and background events; finding variables that are able to separate signal from background (jet multiplicity, transverse momenta of jets and leptons, lepton isolation, b-tagging, missing transverse energy, angular distributions); development and modification of cuts on these variables in order to enrich the signal-over-background separation; optimisation of the signal-over-background ratio and estimation of the purity based on simulation only; comparisons of the selection efficiency between data and simulation.

Advanced-level tasks might include: derivation of production cross sections and masses of objects, reconstruction of the objects (quarks or bosons) by assigning the detector physics objects (jets, leptons, missing energy) to the hypothetical decay trees; estimation of the impact of other sources of systematic uncertainties (luminosity uncertainty, b-tagging efficiency, background modelling) by adding approximate and conservative values.

In addition it may be used as a test-bed for new data-analysis techniques, e.g. kinematic fitting procedures, multivariate discrimination of signal from background and other ML-related tasks. The current content does not support the creation of unfolded distributions or searches for signals that are not supported by signal datasets made available by the ATLAS Collaboration.

An important aspect of the 13 TeV ATLAS Open Data is that it is prepared specifically for educational purposes. To this end, precision has been traded for simplicity of use. The simplifications are:

- Scale factors implementing corrections for different object efficiencies are calculated using the preselection cuts described in Table 1. This 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 per-jet b-tagging scale factor (scaleFactor_BTAG) is computed for a specific WP given a specific b-tagging algorithm (MV2c10 with a 70% b-jet efficiency [22, 23]). In case a different operating point for the MV2c10 b-tagging algorithm is specified, this introduces a potential mismatch between data and MC simulation.
- No data-driven estimation of the multijet background is provided, and the contributing effects of fake or non-prompt leptons may be countered using strict object definitions such as lepton identification, isolation and p_{T} requirements, as well as tighter cuts on $E_{\mathrm{T}}^{\mathrm{miss}}$ and $M_{\mathrm{T}}^{\mathrm{W}}$ distributions. Yet, any residual disagreement might be understood as a sign that the multijet contribution to the electron and muon channels are not taken into account. Designing and implementing approximate data-driven methods to estimate multijet backgrounds is left as a challenging exercise to advanced students.
- In order to provide ground for systematic-uncertainty estimation studies, but reduce large complexities, only a simplified single-component systematic-uncertainty estimate related to object transverse-momentum reconstruction is included in the datasets, detailed in Table 5.

5 ATLAS Open Data educational tools

The release of the 13 TeV ATLAS Open Data, hosted on the CERN and ATLAS Open Data online portals [2, 3] as well as preloaded into the ATLAS Open Data Key flash drives [9], is accompanied by a set of associated educational tools:

- The 13 TeV ATLAS Open Data *analysis framework*, written in C++ and interfaced with ROOT and publicly available on the GitHub [60] platform, implements the protocols needed for reading the datasets, making an analysis selection, writing out histograms and plotting the results. The analysis examples described in Section 3 are created with this framework, and the ease of use and a clear and simple structure of the tools is emphasised.
- The 13 TeV ATLAS Open Data *ROOTbooks* allow data analysis to be performed directly in a web browser by integrating the ROOT framework with the Jupyter [61] notebook technology, a combination called "ROOTbook". Several notebooks with analysis examples and an interface to launch the framework mentioned above, are available on the ATLAS Open Data educational portal using SWAN [62] and Binder [63] platforms.
- The 13 TeV ATLAS Open Data *virtual machine* provides an unmodified operating system with all of its installed software to run on top of another operating system. The virtual machine contains the latest versions of the Linux and ROOT distributions, the 13 TeV ATLAS Open Data analysis framework and the complete 13 TeV ATLAS Open Data set needed to carry out educational analysis of the released datasets.
- The 13 TeV ATLAS Open Data *online documentation* platform [3] provides introductory material and detailed information for wide-audience about the ATLAS experiment, the released 13 TeV ATLAS Open Data sets and their visualisation, the analysis frameworks and ROOTbooks, as well as the virtual machine usage instructions.

These educational tools are aimed to foster educational efforts at the high-school and university levels by providing the material for hands-on physics exercises at different levels of complexity.

6 Summary

A review of a new set of pp collision data released by the ATLAS Collaboration to the public for educational purposes is presented. The data has been collected by the ATLAS detector at the LHC at $\sqrt{s} = 13$ TeV during the year 2016 and corresponds to an integrated luminosity of 10 fb^{-1}. The pp collision data is accompanied by a set of MC simulated samples describing several processes which are used to model the expected distributions of different signal and background events. Associated educational tools are released to make the analysis of the dataset easily accessible. These educational tools provide a straightforward interface to replicate the procedures used by HEP researchers while enabling users to experience the analysis of particle-physics data in educational environments. A number of physics-analysis examples inspired by published ATLAS Collaboration results are presented to demonstrate the wide range of final-state scenarios provided within the 13 TeV ATLAS Open Data.

References

[1] ATLAS Collaboration, *ATLAS Data Access Policy*, ATL-CB-PUB-2015-001, 2015, URL: http://cds.cern.ch/record/2002139 (cit. on p. 3).

[2] *CERN Open Data Portal*, URL: http://opendata.cern.ch (cit. on pp. 3, 28).

[3] *ATLAS Open Data and Tools for Education*, URL: http://opendata.atlas.cern (cit. on pp. 3, 28).

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

[5] *International Particle Physics Outreach Group*, URL: http://ippog.org/ (cit. on p. 3).

[6] *International Masterclasses*, URL: http://physicsmasterclasses.org (cit. on p. 3).

[7] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997) 81 (cit. on pp. 3, 6).

[8] ATLAS Collaboration, *Review of the ATLAS Open Data Dataset*, ATL-OREACH-PUB-2016-001, 2016, URL: https://cds.cern.ch/record/2203649 (cit. on p. 3).

[9] ATLAS Collaboration, *Review of ATLAS Open Data 8 TeV datasets, tools and activities*, ATL-OREACH-PUB-2018-001, 2018, URL: https://cds.cern.ch/record/2624572/files/ATL-OREACH-PUB-2018-001.pdf (cit. on pp. 3, 28).

[10] M. O. Evans, 'Enabling Open Science with the ATLAS Open Data project at CERN', 2018, URL: http://cds.cern.ch/record/2630961 (cit. on p. 3).

[11] *ATLAS Higgs Boson Machine Learning Challenge*, URL: http://www.kaggle.com/c/higgs-boson (cit. on p. 3).

[12] *ATLAS TrackML Particle Tracking Challenge*, URL: http://www.kaggle.com/c/trackml-particle-identification (cit. on p. 3).

[13] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653, arXiv: 1608.03953 [hep-ex] (cit. on p. 3).

[14] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex] (cit. on pp. 4, 5).

[15] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on pp. 4, 5, 12, 16, 35).

[16] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, JINST **14** (2019) P03017, arXiv: 1812.03848 [hep-ex] (cit. on pp. 4, 35).

[17] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, Eur. Phys. J. C **79** (2019) 205, arXiv: 1810.05087 [hep-ex] (cit. on pp. 4, 5).

[18] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on pp. 4, 5).

[19] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 4).

[20] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, arXiv: 1703.09665 [hep-ex] (cit. on p. 4).

[21] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, URL: https://cds.cern.ch/record/1700870 (cit. on pp. 4, 34).

[22] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: https://cds.cern.ch/record/2160731 (cit. on pp. 5, 9, 27, 34).

[23] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s}$ = 13 TeV*, JHEP **08** (2018) 089, arXiv: 1805.01845 [hep-ex] (cit. on pp. 5, 27).

[24] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s}$ = 13 TeV*, ATLAS-CONF-2017-029, 2017, URL: https://cds.cern.ch/record/2261772 (cit. on pp. 5, 22, 35).

[25] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, 2015, URL: https://cds.cern.ch/record/2064383 (cit. on pp. 5, 22).

[26] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 17, arXiv: 1406.0076 [hep-ex] (cit. on p. 5).

[27] D. Krohn, J. Thaler and L.-T. Wang, *Jet trimming*, JHEP **02** (2010) 084, arXiv: 0912.1342 [hep-ph] (cit. on p. 5).

[28] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. C **77** (2017) 332, arXiv: 1611.10235 [hep-ex] (cit. on p. 6).

[29] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Lett. B **759** (2016) 601, arXiv: 1603.09222 [hep-ex] (cit. on pp. 8, 13).

[30] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv: 1802.08168 [hep-ex] (cit. on pp. 8, 12, 35).

[31] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton–proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS Detector*, ATLAS-CONF-2014-058, 2014, URL: https://cds.cern.ch/record/1951336 (cit. on p. 8).

[32] ATLAS Collaboration, *Probing the Wtb vertex structure in t-channel single-top-quark production and decay in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, JHEP **04** (2017) 124, arXiv: 1702.08309 [hep-ex] (cit. on p. 9).

[33] ATLAS Collaboration, *Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **04** (2017) 086, arXiv: 1609.03920 [hep-ex] (cit. on p. 9).

[34] ATLAS Collaboration, *Measurement of the top quark mass in the $t\bar{t} \to$ lepton+jets channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results*, Eur. Phys. J. C **79** (2019) 290, arXiv: 1810.01772 [hep-ex] (cit. on p. 10).

[35] ATLAS Collaboration, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, JHEP **11** (2017) 191, arXiv: 1708.00727 [hep-ex] (cit. on pp. 10, 23).

[36] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, (2019), arXiv: 1908.00005 [hep-ex] (cit. on pp. 12, 35).

[37] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-015, 2015, URL: https://cds.cern.ch/record/2037613 (cit. on pp. 12, 35).

[38] A. Quadt, *Top quark physics at hadron colliders*, Eur. Phys. J. C **48** (2006) 835 (cit. on p. 12).

[39] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001 (cit. on p. 14).

[40] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on p. 13).

[41] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on p. 13).

[42] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv: 1503.07589 [hep-ex] (cit. on p. 13).

[43] A. Djouadi, J. Kalinowski and M. Spira, *HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56, arXiv: hep-ph/9704448 [hep-ph] (cit. on pp. 13, 19, 24).

[44] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, (2016), arXiv: 1610.07922 [hep-ph] (cit. on pp. 13, 19).

[45] ATLAS Collaboration, *Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **789** (2019) 508, arXiv: 1808.09054 [hep-ex] (cit. on p. 14).

[46] J. Alwall, P. Schuster and N. Toro, *Simplified Models for a First Characterization of New Physics at the LHC*, Phys. Rev. D **79** (2009) 075020, arXiv: 0810.3921 [hep-ph] (cit. on p. 16).

[47] ATLAS Collaboration, *Search for electroweak production of supersymmetric particles in final states with two or three leptons at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 995, arXiv: 1803.02762 [hep-ex] (cit. on p. 16).

[48] C. G. Lester and D. J. Summers, *Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders*, Phys. Lett. B **463** (1999) 99, arXiv: hep-ph/9906349 [hep-ph] (cit. on p. 16).

[49] ATLAS Collaboration, *Measurement of the $W^{\pm}Z$ boson pair-production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, Phys. Lett. B **762** (2016) 1, arXiv: 1606.04017 [hep-ex] (cit. on p. 17).

[50] ATLAS Collaboration, *Measurement of the ZZ Production Cross Section in pp Collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Detector*, Phys. Rev. Lett. **116** (2016) 101801, arXiv: 1512.05314 [hep-ex] (cit. on p. 19).

[51] ATLAS Collaboration, *Measurement of inclusive and differential cross sections in the $H \to ZZ^* \to 4\ell$ decay channel in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **10** (2017) 132, arXiv: 1708.02810 [hep-ex] (cit. on p. 20).

[52] ATLAS Collaboration, *Measurement of τ polarisation in $Z/\gamma^* \to \tau\tau$ decays in proton–proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 163, arXiv: 1709.03490 [hep-ex] (cit. on p. 22).

[53] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **99** (2019) 072001, arXiv: 1811.08856 [hep-ex] (cit. on p. 22).

[54] A. Elagin, P. Murat, A. Pranko and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. A **654** (2011) 481, arXiv: 1012.4686 [hep-ex] (cit. on pp. 22, 35).

[55] R. M. Harris and S. Jain, *Cross Sections for Leptophobic Topcolor Z' Decaying to Top-Antitop*, Eur. Phys. J. C **72** (2012) 2072, arXiv: 1112.4928 [hep-ph] (cit. on p. 23).

[56] ATLAS Collaboration, *Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 565, arXiv: 1804.10823 [hep-ex] (cit. on pp. 23, 24).

[57] ATLAS Collaboration, *Boosted hadronic top identification at ATLAS for early 13 TeV data*, ATL-PHYS-PUB-2015-053, 2015, URL: https://cds.cern.ch/record/2116351 (cit. on pp. 24, 35).

[58] ATLAS Collaboration, *A search for $t\bar{t}$ resonances using lepton-plus-jets events in proton–proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, JHEP **08** (2015) 148, arXiv: 1505.07018 [hep-ex] (cit. on p. 24).

[59] ATLAS Collaboration, *Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **98** (2018) 052005, arXiv: 1802.04146 [hep-ex] (cit. on p. 25).

[60] *ATLAS Outreach data and tools GitHub account*, URL: http://github.com/atlas-outreach-data-tools (cit. on p. 28).

[61] *Project Jupyter*, URL: http://jupyter.org/ (cit. on p. 28).

[62] *Service for Web based ANalysis*, URL: http://swan.web.cern.ch/ (cit. on p. 28).

[63] *Binder platform*, URL: http://mybinder.org/ (cit. on p. 28).

[64] ATLAS Collaboration, *Validation of Monte Carlo event generators in the ATLAS Collaboration for LHC Run 2*, ATL-PHYS-PUB-2016-001, 2016, URL: https://cds.cern.ch/record/2119984 (cit. on p. 36).

[65] ATLAS Collaboration, *Multi-boson simulation for 13 TeV ATLAS analyses*, ATL-PHYS-PUB-2016-002, 2016, URL: https://cds.cern.ch/record/2119986 (cit. on p. 36).

[66] ATLAS Collaboration, *Monte Carlo Generators for the Production of a W or Z/γ^* Boson in Association with Jets at ATLAS in Run 2*, ATL-PHYS-PUB-2016-003, 2016, URL: https://cds.cern.ch/record/2120133 (cit. on p. 36).

[67] ATLAS Collaboration, *Simulation of top-quark production for the ATLAS experiment at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2016-004, 2016, URL: https://cds.cern.ch/record/2120417 (cit. on p. 36).

[68] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on p. 36).

[69] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on p. 36).

[70] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv: hep-ph/0603175 (cit. on p. 36).

[71] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on p. 36).

[72] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on p. 36).

Appendices

A Content of the 13 TeV ATLAS Open Data tuple

The detailed content of the 13 TeV ATLAS Open Data ROOT tuple is given in Table 4 and Table 5.

Table 4: Description of the 13 TeV ATLAS Open Data branches and variables.

Tuple branch name	C++ type	Variable description
runNumber	int	number uniquely identifying ATLAS data-taking run
eventNumber	int	event number and run number combined uniquely identifies event
channelNumber	int	number uniquely identifying ATLAS simulated dataset
mcWeight	float	weight of a simulated event
XSection	float	total cross-section, including filter efficiency and higher-order correction factor
SumWeights	float	generated sum of weights for MC process
scaleFactor_PILEUP	float	scale-factor for pileup reweighting
scaleFactor_ELE	float	scale-factor for electron efficiency
scaleFactor_MUON	float	scale-factor for muon efficiency
scaleFactor_PHOTON	float	scale-factor for photon efficiency
scaleFactor_TAU	float	scale-factor for tau efficiency
scaleFactor_BTAG	float	scale-factor for b-tagging algorithm @70% efficiency
scaleFactor_LepTRIGGER	float	scale-factor for lepton triggers
scaleFactor_PhotonTRIGGER	float	scale-factor for photon triggers
trigE	bool	boolean whether event passes a single-electron trigger
trigM	bool	boolean whether event passes a single-muon trigger
trigP	bool	boolean whether event passes a diphoton trigger
lep_n	int	number of pre-selected leptons
lep_truthMatched	vector<bool>	boolean indicating whether the lepton is matched to a simulated lepton
lep_trigMatched	vector<bool>	boolean indicating whether the lepton is the one triggering the event
lep_pt	vector<float>	transverse momentum of the lepton
lep_eta	vector<float>	pseudo-rapidity, η, of the lepton
lep_phi	vector<float>	azimuthal angle, ϕ, 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. primary vertex
lep_charge	vector<int>	charge of the lepton
lep_type	vector<int>	number signifying the lepton type (e or μ)
lep_isTightID	vector<bool>	boolean indicating whether lepton satisfies tight ID reconstruction criteria
lep_ptcone30	vector<float>	scalar sum of track p_{T} in a cone of R=0.3 around lepton, used for tracking isolation
lep_etcone20	vector<float>	scalar sum of track E_{T} in a cone of R=0.2 around lepton, used for calorimeter isolation
lep_trackd0pvunbiased	vector<float>	d_0 of track associated to lepton at point of closest approach (p.c.a.)
lep_tracksigd0pvunbiased	vector<float>	d_0 significance of the track associated to 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 pre-selected jets
jet_pt	vector<float>	transverse momentum of the jet
jet_eta	vector<float>	pseudo-rapidity, η, of the jet
jet_phi	vector<float>	azimuthal angle, ϕ, of the jet
jet_E	vector<float>	energy of the jet
jet_jvt	vector<float>	jet vertex tagger discriminant [21] of the jet
jet_trueflav	vector<int>	flavour of the simulated jet
jet_truthMatched	vector<bool>	boolean indicating whether the jet is matched to a simulated jet
jet_MV2c10	vector<float>	output from the multivariate b-tagging algorithm [22] of the jet

Table 5: Description of the 13 TeV ATLAS Open Data branches and variables.

Tuple branch name	C++ type	Variable description
photon_n	int	number of pre-selected photons
photon_truthMatched	vector<bool>	boolean indicating whether the photon is matched to a simulated photon
photon_trigMatched	vector<bool>	boolean indicating whether the photon is the one triggering the event
photon_pt	vector<float>	transverse momentum of the photon
photon_eta	vector<float>	pseudo-rapidity of the photon
photon_phi	vector<float>	azimuthal angle of the photon
photon_E	vector<float>	energy of the photon
photon_isTightID	vector<bool>	boolean indicating whether photon satisfies tight identification reconstruction criteria
photon_ptcone30	vector<float>	scalar sum of track p_{T} in a cone of R=0.3 around photon
photon_etcone20	vector<float>	scalar sum of track E_{T} in a cone of R=0.2 around photon
photon_convType	vector<int>	information whether and where the photon was converted
largeRjet_n	int	number of pre-selected large-R jets
largeRjet_pt	vector<float>	transverse momentum of the large-R jet
largeRjet_eta	vector<float>	pseudo-rapidity of the large-R jet
largeRjet_phi	vector<float>	azimuthal angle of the large-R jet
largeRjet_E	vector<float>	energy of the large-R jet
largeRjet_m	vector<float>	invariant mass of the large-R jet
largeRjet_truthMatched	vector<int>	information whether the large-R jet is matched to a simulated large-R jet
largeRjet_D2	vector<float>	weight from algorithm [57] for W/Z-boson tagging
largeRjet_tau32	vector<float>	weight from algorithm [57] for top-quark tagging
tau_n	int	number of pre-selected hadronically decaying τ-lepton
tau_pt	vector<float>	transverse momentum of the hadronically decaying τ-lepton
tau_eta	vector<float>	pseudo-rapidity of the hadronically decaying τ-lepton
tau_phi	vector<float>	azimuthal angle of the hadronically decaying τ-lepton
tau_E	vector<float>	energy of the hadronically decaying τ-lepton
tau_charge	vector<int>	charge of the hadronically decaying τ-lepton
tau_isTightID	vector<bool>	boolean indicating whether hadronically decaying τ-lepton satisfies tight ID reconstruction criteria
tau_truthMatched	vector<bool>	boolean indicating whether the hadronically decaying τ-lepton is matched to a simulated τ-lepton
tau_trigMatched	vector<bool>	boolean signifying whether the τ-lepton is the one triggering the event
tau_nTracks	vector<int>	number of tracks in the hadronically decaying τ-lepton decay
tau_BDTid	vector<float>	output of the multivariate algorithm [24] discriminating hadronically decaying τ-leptons from jets
ditau_m	float	di-τ invariant mass using the missing-mass calculator [54]
lep_pt_syst	vector<float>	single component syst. uncert. (lepton momentum scale and resolution [15, 36]) affecting lep_pt
met_et_syst	float	single component syst. uncert. ($E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution [30]) affecting met_pt
jet_pt_syst	vector<float>	single component syst. uncert. (jet energy scale [37]) affecting jet_pt
photon_pt_syst	vector<float>	single component syst. uncert. (photon energy scale and resolution [16]) affecting photon_pt
largeRjet_pt_syst	vector<float>	single component syst. uncert. (large-R jet energy resolution [37]) affecting largeRjet_pt
tau_pt_syst	vector<float>	single component syst. uncert. (τ-lepton reconstruction and energy scale [24]) affecting tau_pt

B MC samples released in the 13 TeV ATLAS Open Data

The summary of the 13 TeV ATLAS Open Data signal and background MC simulation samples used in Section 3 to illustrate the examples of physics analysis is given in Table 6. An overall summary of the validation of MC event generators and MC simulation setups of W/Z+jets, multi-boson, single and pair top-quark production, and other MC samples used in published ATLAS Collaboration results can be found in Refs. [64–67].

Table 6: Description of the MC samples released in the 13 TeV ATLAS Open Data.

Process	Unique "channelNumber"	Generator, hadronisation	Additional information
		Top-quark production	
$t\bar{t}$+jets	410000	Powheg-Box v2 [68] + Pythia 8 [69]	only 1ℓ and 2ℓ decays of $t\bar{t}$-system
single (anti)top t-channel	(410012) 410011	Powheg-Box v1 + Pythia 6 [70]	
single (anti)top Wt-channel	(410014) 410013	Powheg-Box v2 + Pythia 6	
single (anti)top s-channel	(410026) 410025	Powheg-Box v2 + Pythia 6	
		W/Z (+ jets) production	
$Z \to ee, \mu\mu, \tau\tau$	361106 – 361108	Powheg-Box v2 + Pythia 8	LO accuracy up to $N_{\text{jets}} = 1$
$W \to e\nu, \mu\nu, \tau\nu$	361100 – 361105	Powheg-Box v2 + Pythia 8	LO accuracy up to $N_{\text{jets}} = 1$
$W \to e\nu, \mu\nu, \tau\nu$ + jets	364156 – 364197	Sherpa 2.2 [71]	LO accuracy up to 3-jets final states
$Z \to ee, \mu\mu, \tau\tau$ + jets	364100 – 364141	Sherpa 2.2	LO accuracy up to 3-jets final states
		Diboson production	
WW	363359, 363360	Sherpa 2.2	$qq'\ell\nu$ final states
WW	363492	Sherpa 2.2	$\ell\nu\ell'\nu'$ final states
ZZ	363356	Sherpa 2.2	$qq'\ell^+\ell^-$ final states
ZZ	363490	Sherpa 2.2	$\ell^+\ell^-\ell'^+\ell'^-$ final states
WZ	363358	Sherpa 2.2	$qq'\ell^+\ell^-$ final states
WZ	363489	Sherpa 2.2	$\ell\nu qq'$ final states
WZ	363491	Sherpa 2.2	$\ell\nu\ell^+\ell^-$ final states
WZ	363493	Sherpa 2.2	$\ell\nu\nu\nu'$ final states
		SM Higgs production ($m_{\text{H}} = 125$ GeV)	
ggF, $H \to WW$	345324	Powheg-Box v2 + Pythia 8	$\ell\nu\ell'\nu'$ final states
VBF, $H \to WW$	345323	Powheg-Box v2 + Pythia 8	$\ell\nu\ell'\nu'$ final states
ggF, $H \to ZZ$	345060	Powheg-Box v2 + Pythia 8	$\ell^+\ell^-\ell'^+\ell'^-$ final states
VBF, $H \to ZZ$	344235	Powheg-Box v2 + Pythia 8	$\ell^+\ell^-\ell'^+\ell'^-$ final states
$ZH, H \to ZZ$	341947	Pythia 8	$\ell^+\ell^-\ell'^+\ell'^-$ final states
$WH, H \to ZZ$	341964	Pythia 8	$\ell^+\ell^-\ell'^+\ell'^-$ final states
ggF, $H \to \gamma\gamma$	343981	Powheg-Box v2 + Pythia 8	
VBF, $H \to \gamma\gamma$	345041	Powheg-Box v2 + Pythia 8	
$WH(ZH), H \to \gamma\gamma$	345318, 345319	Powheg-Box v2 + Pythia 8	
$t\bar{t}H, H \to \gamma\gamma$	341081	aMC@NLO [72] + Pythia 8	
		BSM production	
$Z' \to t\bar{t}$	301325	Pythia 8	$m_{Z'} = 1$ TeV
$\tilde{\ell}\tilde{\ell}' \to \ell\tilde{\chi}_1^0\ell'\tilde{\chi}_1^{0\prime}$	392985	aMC@NLO + Pythia 8	$m_{\tilde{\ell}} = 600$ GeV, $m_{\tilde{\chi}_1^0} = 300$ GeV

C Evolution of the ATLAS Open Data from the 8 TeV release (2016) to the 13 TeV release (2019)

Figure 13(a) shows the evolution of the ATLAS Open Data, Figure 13(b) shows the evolution of the ATLAS Open Data SM Higgs and BSM MC signals and Figure 13(c) shows the evolution of the ATLAS Open Data tuple structure from the 8 TeV release (2016) to the 13 TeV release (2019).

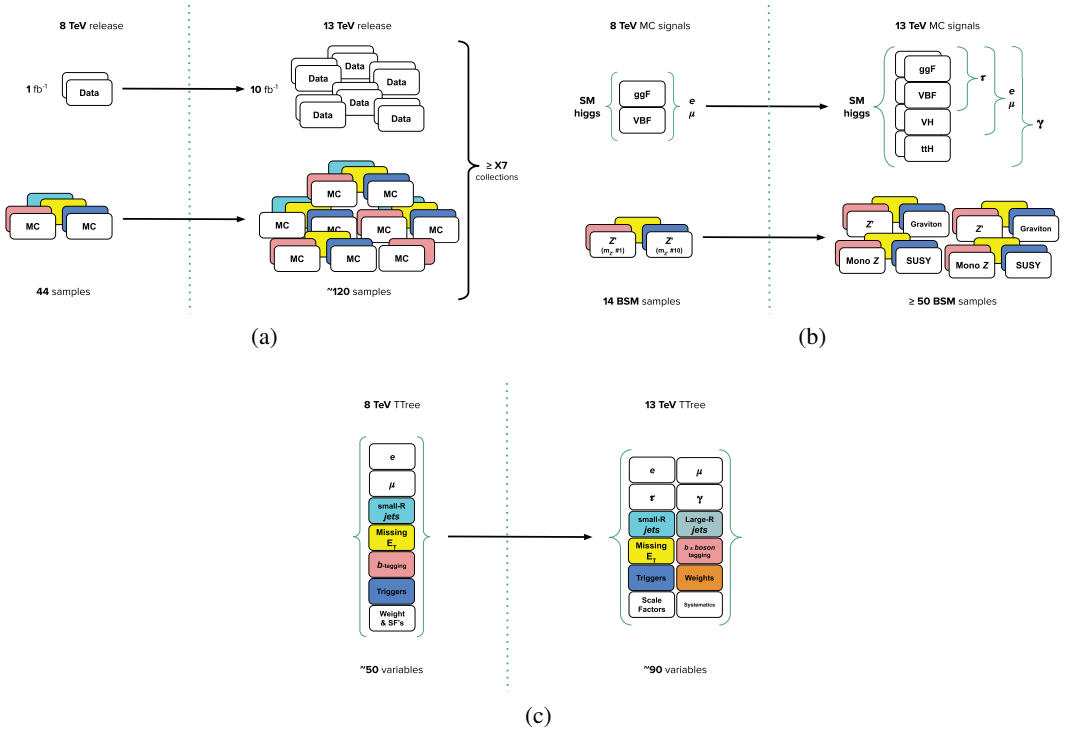


Figure 13: Evolution of ATLAS (a) Open Data, (b) Open Data SM Higgs and BSM MC signals and (c) Open Data tuple structure from the 8 TeV release (2016) to the 13 TeV release (2019).