7 Preparing ATLAS data for education worldwide

Respect your parents. They passed school without Google.

Anon [190]

This chapter discusses the education work that forms part of this thesis - ATLAS Open Data. The ATLAS Open Data project provides open-source access to measured data, simulation, resources, and documentation for the purpose of education. ATLAS was the first LHC experiment to release real 13 TeV collision data [166, 191]. The development and testing of specific resources related to the $t\bar{t}Z$ 2ℓOS process are discussed in this chapter. It is important to point out however, that many other resources unrelated to the $t\bar{t}Z$ 2ℓOS process were also developed. All data and resources can be accessed from the ATLAS Open Data website [148]. This chapter is structured as follows:

1. discussion of the Histogram Analyser;
2. discussion of ATLAS Open Data Jupyter notebooks.

The author's specific contribution was to:

- create a data pipeline to go from 13 TeV data used for physics analysis to simplified data formats, which then allowed the creation of datasets that could be used for the $t\bar{t}Z$ 2ℓOS Histogram Analyser and Jupyter notebooks;
- create the 13 TeV datasets used as input for Open Data analyses, including those used in the $t\bar{t}Z$ 2ℓOS Histogram Analyser and Jupyter notebooks;
- write example physics analyses for use with 13 TeV ATLAS Open Data, for example the $t\bar{t}Z$ 2ℓOS Histogram Analyser and Jupyter notebooks;

- write corresponding documentation for 13 TeV datasets and example analyses, similar to the accompanying explanations given throughout this chapter;
- test 13 TeV datasets and example analyses, for example through the $t\bar{t}Z$ 2ℓOS Histogram Analyser and Jupyter notebooks.

7.1 The data

The 13 TeV ATLAS Open Data release constitutes 10 $\text{fb}^{-1}$of experimental data, which is approximately 1/14th of the data collected by ATLAS in Run 2. 10 fb^{-1}correspond to approximately 1000 trillion proton-proton collisions.The whole release is in .root file format, along with csv file formats for some specific processes. The variables present in the datasets were summarised in Table 4.4.1, and further information can be found in Ref. [166]. The data can be accessed through the ATLAS Open Data portal [148] or CERN Open Data portal [149]. Analysis of these data is possible through a number of tools, including the Histogram Analyser (Section 7.2) and Jupyter notebooks (Section 7.3)

7.2 Histogram Analyser

The Histogram Analyser is one of the main web-based resources that was developed for using ATLAS data for education. It allows students to apply selection requirements to histograms without the need to use computer code. It is possible to apply selection requirements on eight different variables, all of which are presented as individual histograms. This section introduces and covers the $t\bar{t}Z$ Histogram Analyser, the individual histograms that form it, and conclusions that can be drawn from three different signal regions. The $t\bar{t}Z$ Histogram Analyser is focused on because the author of this thesis was the main developer.

7.2.1 Introduction

The ATLAS Open Data Histogram Analyser [192, 193] is a web-based tool for fast, cut-based analysis of data, allowing to visualise data using online histograms with only a computer mouse. This tool shows how to differentiate between physics processes. By applying cuts to data, specific physics processes (signal) can be isolated from the background. The webpage [193] displays nine histograms of variables which can be used to isolate signal events. One can use their cursor to apply selections to a particular variable. Cutting on one histogram cuts the whole datasets, therefore changing the distributions of all 9 histograms - the effect on the other variables will be shown immediately. The Histogram Analyser helps in understanding the data and the relationship between the signal and background processes. It can simplify and speed-up the selection of cuts, before coding an analysis. The Histogram Analyser is used for an initial look at the $t\bar{t}Z$ 2ℓOS process.

7.2.2 The $t\bar{t}Z$ Histogram Analyser

The $t\bar{t}Z$ Histogram Analyser is used to help visualise rare top-quark measured data and simulations. This Histogram Analyser searches for rare top-quark processes. Data are shown by the black dots, with error bars. The error bars are statistical. The three main processes are $t\bar{t}Z$ signal, $t\bar{t}$ background and Z background. This Histogram Analyser also includes minor backgrounds, labelled as ‘Other’ in red. Minor backgrounds are required for data to match the total simulation. ‘Other’ includes single top production, WZ and ZZ diboson production and $t\bar{t}W$. Each process is represented by a different colour in the Histogram Analyser.

The Histogram Analyser displays nine histograms, shown in Figure 7.2.1 and described in the following.

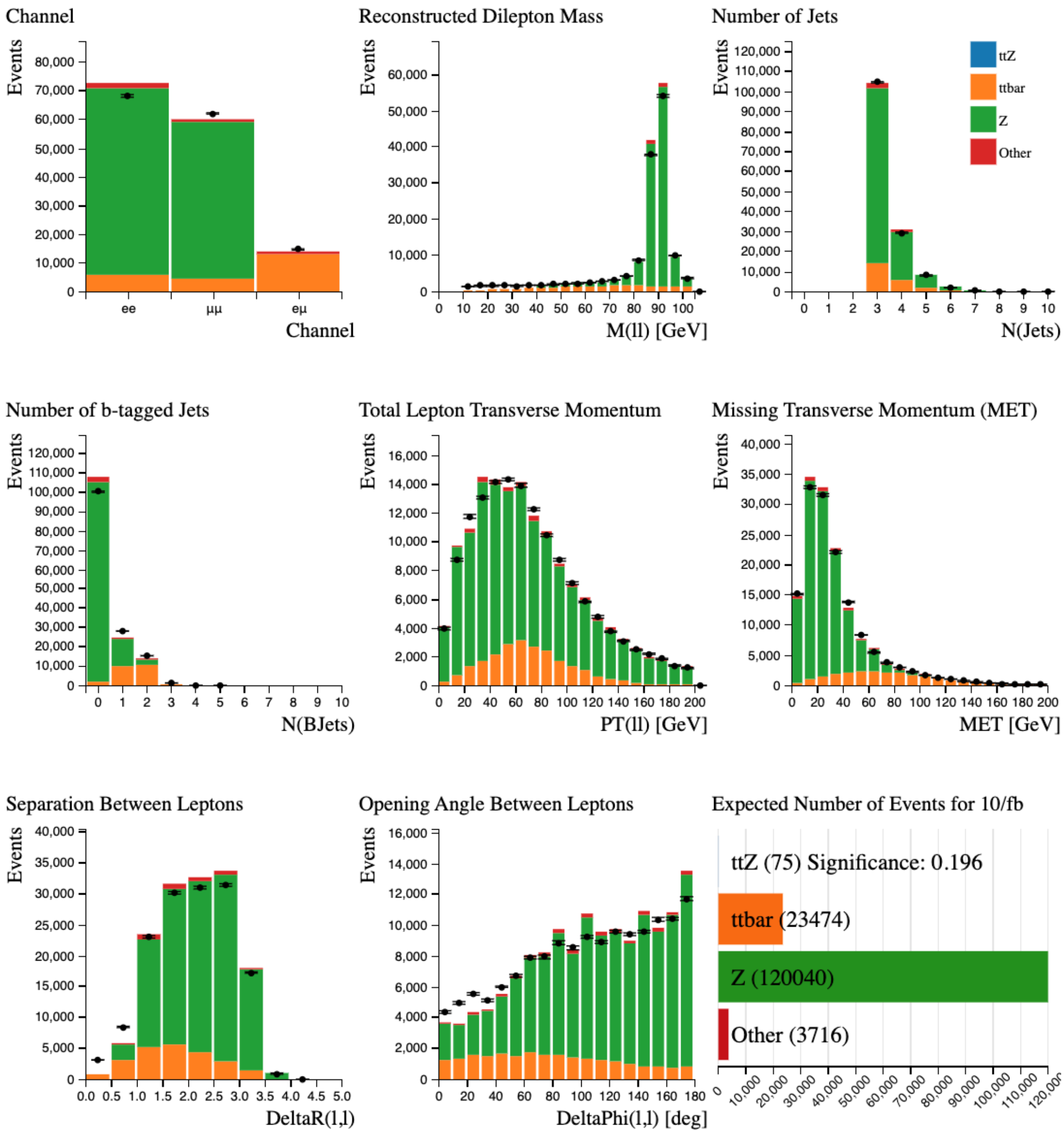


Figure 7.2.1: $t\bar{t}Z$ Histogram Analyser before any selections are applied. The 9 histograms are (top left) Channel, (top middle) Reconstructed Dilepton Mass, (top right) Number of Jets, (centre left) Number of b-tagged Jets, (centre middle) Total Lepton Transverse Momentum, (centre right) Missing Transverse Momentum, (bottom left) Separation Between Leptons, (bottom middle) Opening Angle Between Leptons, (bottom right) Expected Number of Events.

7.2.3 Expected Number of Events for 10 fb^{-1}

This histogram shows the number of events expected to be detected, reconstructed and recorded by ATLAS for 10 inverse femtobarn (10 fb^{-1}) of data, before any additional selections are made on the Histogram Analyser.

The expected number of real data events reconstructed and recorded by ATLAS is different to the number of events produced by real collisions. Some events will not be reconstructed due to the way the detector is constructed, the resolution of the sub-detectors, reconstruction efficiency and other inefficiencies.

Table 7.2.1 shows the cross-sections used by ATLAS Open Data [194], along with the expected number of events before applying additional cuts with the Histogram Analyser. With no cuts, we have 75 $t\bar{t}Z$ events, with many more background events. The majority of the background at this point is Z boson production, which can change depending on the cuts applied.

Process	Cross-section (pb)	Expected # of events
$t\bar{t}Z$	0.08258096	75
$t\bar{t}$	452.693559	23474
Z	3901.1616	120040

Table 7.2.1: Cross-sections used for the different processes of the $t\bar{t}Z$ Histogram Analyser [194], along with the expected number of events before any additional cuts are applied in the Histogram Analyser.

The **significance** of $t\bar{t}Z$ quantifies how "significant" the $t\bar{t}Z$ simulation sample is with respect to the background. It is calculated by the simplified equation:

$$\frac{\text{Number of } t\bar{t}Z \text{ events}}{\sqrt{\text{Number of background events}}}. \tag{7.2.1}$$

A larger significance value indicates better extraction of the $t\bar{t}Z$ signal amongst the backgrounds.

7.2.4 Preselections

Some pre-selections were applied to reduce the size of the datasets used as inputs to the $t\bar{t}Z$ Histogram Analyser so that the website can run quicker. These pre-selections include:

- exactly 2 leptons are required;
- decays to taus or hadrons are removed;
- events with <3 jets are removed;

7.2.5 The Histograms

Channel

The leptonic decay channels are shown in this first histogram in the top left: dielectron ee, dimuon $\mu\mu$ and electron-muon $e\mu$.

Reconstructed Dilepton Mass, M(ll)

The “Reconstructed Dilepton Mass" histogram displays the mass reconstructed from the two leptons in the final state. For $t\bar{t}Z$ 2ℓOS signal and Z background, these would originate from a Z boson. With no cuts, this peaks at 90 GeV, due to the huge Z boson contribution.

Number of Jets, NJets

The “Number of Jets" histogram displays the number of jets found in the event.

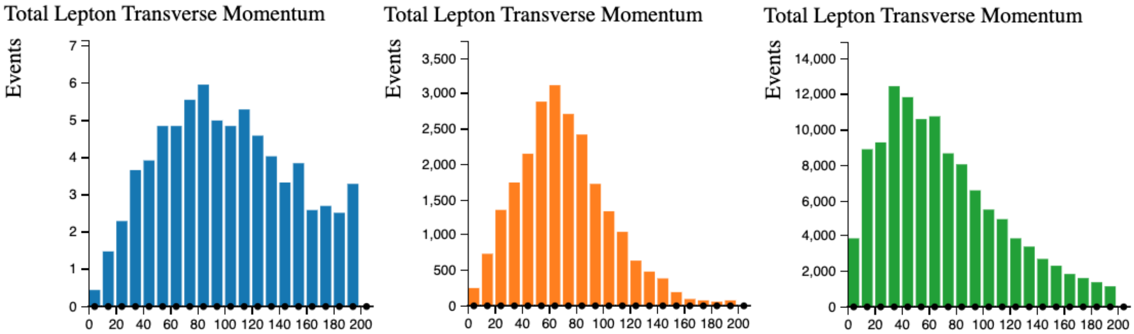
Number of b-tagged Jets, N(BJets)

Jets originating from b-quarks are identified and labelled, or **tagged**, using so-called b-tagging algorithms. b-tagged jets are expected in top quark decays, but not in leptonic W or Z boson decays.

Total Lepton Transverse Momentum, PT(l,l)

Total Lepton Transverse Momentum is the vectorial sum of the transverse momenta of the observed charged leptons.

For Z boson events, total lepton transverse momentum peaks at low values since the transverse momenta of both leptons mostly cancel each other. For the other processes this cancellation is not as pronounced, their distributions peak at between 60 and 90 GeV. This is illustrated in Figure 7.2.2.



(a) $t\bar{t}Z$ (b) $t\bar{t}$ (c) Z

Figure 7.2.2: Total Lepton Transverse Momentum (PT(ll) [GeV]) distributions for (a) $t\bar{t}Z$, (b) $t\bar{t}$, (c) Z.

Missing Transverse Momentum, MET

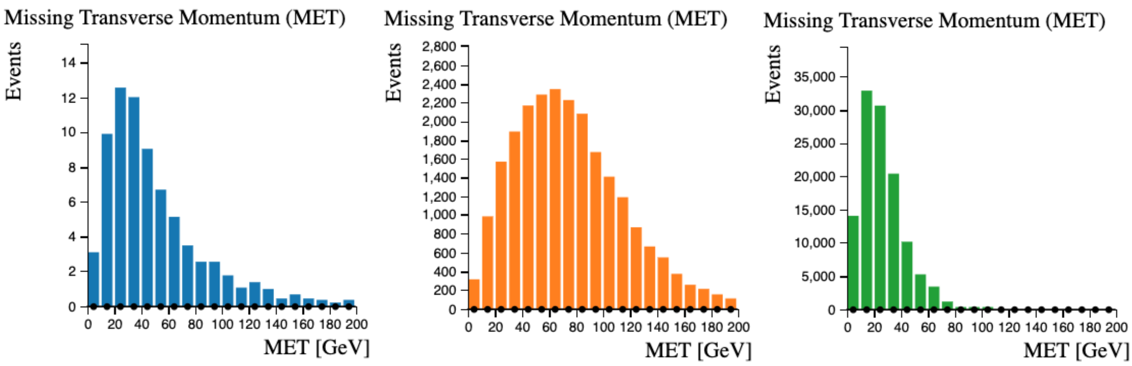
In the LHC, the initial energy of the colliding partons (quarks or gluons) along the beam axis is not known. This is due to the energy of each proton being shared and constantly exchanged between its constituents.

However, the initial momentum of particles travelling transverse to the beam axis is zero. Therefore, any net momentum in the transverse direction indicates missing transverse momentum.

Missing transverse momentum is used to infer the presence of non-detectable particles such as the neutrino. It is also expected to be a signature of many predicted physics events beyond the Standard Model, for example the lightest supersymmetric particle.

The standard abbreviation for missing transverse momentum is MET, for historical reasons.

$t\bar{t}$ decays to two leptons have two neutrinos in the final state while Z boson decays to charged leptons do not. This is illustrated in Figure 7.2.3 by the fact that the $t\bar{t}$ MET distribution peaks at higher values than the MET distributions of $t\bar{t}Z$ and Z.



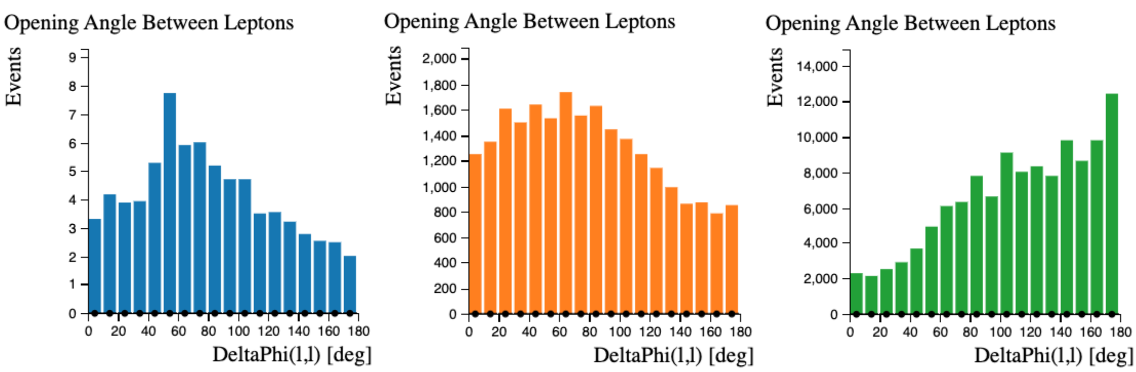
(a) $t\bar{t}Z$ (b) $t\bar{t}$ (c) Z

Figure 7.2.3: Missing Transverse Momentum (MET [GeV]) distributions for (a) $t\bar{t}Z$, (b) $t\bar{t}$, (c) Z.

Opening Angle Between Leptons, DeltaPhi(l,l)

This is the opening angle, measured in phi ϕ, between the two leptons. The azimuthal angle ϕ is measured from the x-axis, around the beam.

If the leptons are emitted back-to-back, this is displayed on the histogram as 180^0. Z events show a peak at high values in contrast to all other processes, as shown in Figure 7.2.4. The reason Z events peak at higher values than other processes is because the leptons from the Z decay are emitted close to back-to-back.



(a) $t\bar{t}Z$ (b) $t\bar{t}$ (c) Z

Figure 7.2.4: DeltaPhi(l,l) distributions for (a) $t\bar{t}Z$, (b) $t\bar{t}$, (c) Z.

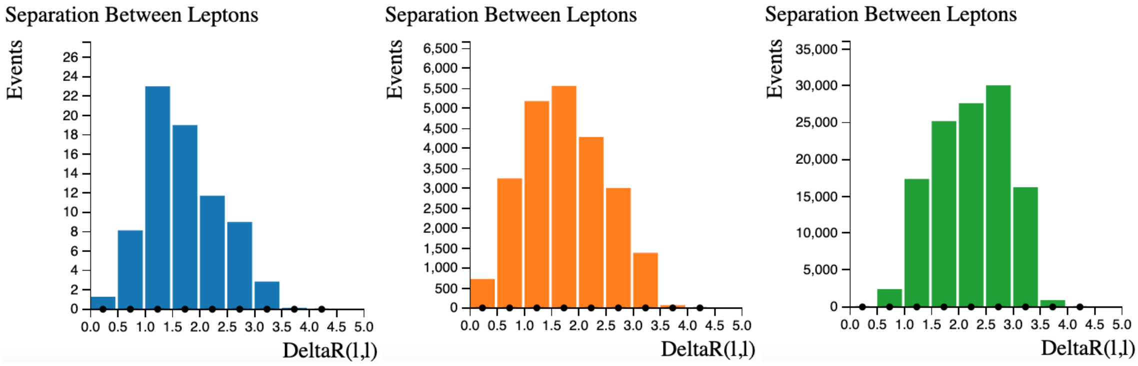
Separation Between Leptons, DeltaR(l,l)

Separation, (ΔR), is calculated using the following equation:

$$(\Delta R)^2 = (\Delta\phi)^2 + (\Delta\eta)^2, \tag{7.2.2}$$

where ϕ is the azimuthal angle between leptons and η is the pseudorapidity.

Figure 7.2.5 shows that $t\bar{t}Z$ events show a peak between 1.0 and 1.5, which is lower values than other processes, with $t\bar{t}$ peaking between 1.5 and 2.0, and Z peaking between 2.5 and 3.0.



(a) $t\bar{t}Z$ (b) $t\bar{t}$ (c) Z

Figure 7.2.5: DeltaR(l,l) distributions for (a) $t\bar{t}Z$, (b) $t\bar{t}$, (c) Z.

7.2.6 Selections for 2ℓ-Z-2b6j

Some of the variables presented in the histograms of the $t\bar{t}Z$ Histogram Analyser are shown pictorially in Figure 7.2.6.

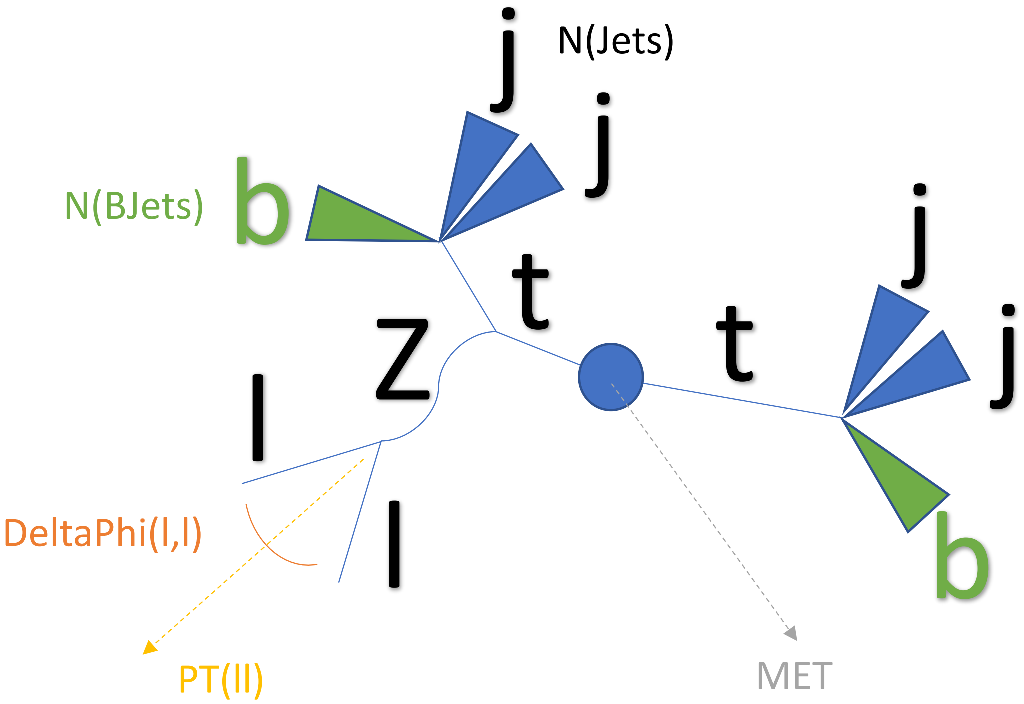


Figure 7.2.6: Schematic diagram of a $t\bar{t}Z$ decay, with some of the variables presented in the histograms of the $t\bar{t}Z$ Histogram Analyser labelled. Antiparticles are not labelled because the Z boson could be radiated from either the top or antitop.

The selections needed to define the 2ℓ-Z-2b6j region in the $t\bar{t}Z$ Histogram Analyser are:

- only the ee and $\mu\mu$ **Channels**;
- **Reconstructed Dilepton Mass** between 80 and 100 GeV;
- **Number of Jets** at least 6;
- **Number of b-tagged Jets** at least 2.

All requirements imposed so far are requirements for the 2ℓ-Z-2b6j signal region (see Table 5.2.2). The remaining variables are not used in the definitions of the final signal regions of the main analysis for this thesis (Section 6), but are used in the Multi-Variate Analysis (MVA) to described in Section 6. Therefore, exploring these variables in the Histogram Analyser can give some intuition as to what the MVA is doing to form signal-rich regions - a key learning objective of the Histogram Analyser.

These further selections are found to be optimal for increasing significance in the $t\bar{t}Z$ Histogram Analyser 2ℓ-Z-2b6j region:

- **PT(ll)** $>$ 30 GeV;
- **MET** $<$ 80 GeV;
- **DeltaPhi(l,l)** $<$ 140^0;
- **Separation** $<$ 3.

The selections for the $t\bar{t}Z$ 2ℓOS channel 2ℓ-Z-2b6j region are shown in Table 7.2.2, along with the background they most help reduce. Significance achieved after making each selection sequentially is also shown in Table 7.2.2.

Variable	Selection	To reduce	Significance afterwards
Channel	e^+e^- or $\mu^+\mu^-$	$t\bar{t}$	0.197
M(ll)	80 < M(ll) < 100 GeV	$t\bar{t}$	0.179
N(Jets)	≥6	Z	0.522
N(BJets)	≥2	Z	0.885
PT(ll)	>30 GeV	Z	0.896
MET	<80 GeV	$t\bar{t}$	0.944
DeltaPhi(l,l)	$<140^0$	Z	0.968
DeltaR(l,l)	<3	Z	0.971

Table 7.2.2: Selections for the $t\bar{t}Z$ 2ℓOS Histogram Analyser 2ℓ-Z-2b6j region, along with the background process that each selection most helps reduce, and the significance achieved after making each selection. Significance quoted is by applying these selections in order.

After each selection, both the data points and the simulated Monte Carlo distributions change. The data and simulated Monte Carlo are not exactly the same, but the general agreement is very good. This shows that these processes are well understood and well modelled.

These selections are shown in Figure 7.2.7, increasing significance to 0.971.

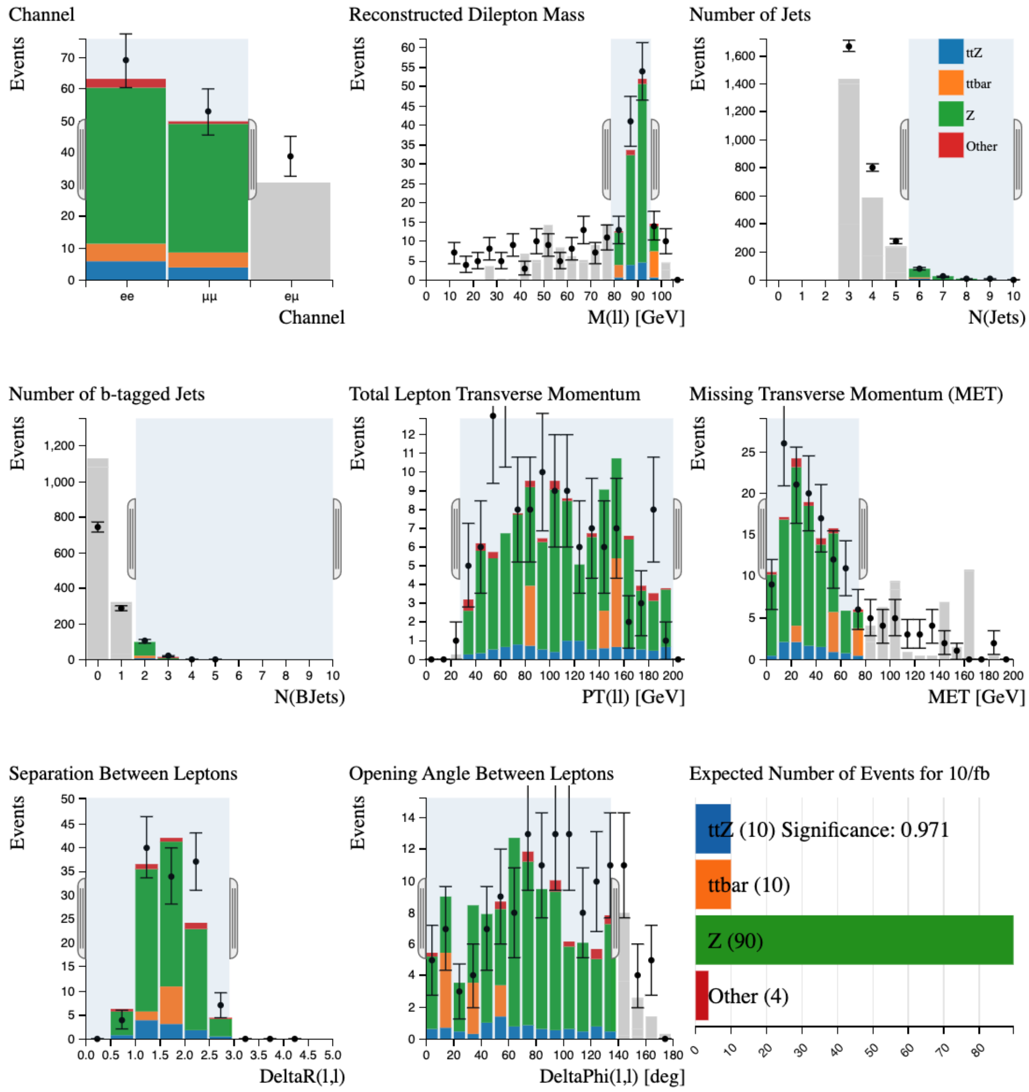


Figure 7.2.7: $t\bar{t}Z$ Histogram Analyser after applying selections for the $t\bar{t}Z$ 2ℓOS 2ℓ-Z-2b6j region. A significance of 0.971 is achieved.

No further changes in selection for any histogram increases the significance over 0.971. This indicates that the selections on Channel, M(ll), N(Jets) and N(BJets) are optimal in terms of signal region definition for 2ℓ-Z-2b6j, as is the case for $t\bar{t}Z$ 2ℓOS papers published by ATLAS [36]. The fact that the maximum significance achievable from defining a looser signal region of N(Jets)≥5 and N(BJets)≥1 indicates that the approach of defining separate signal regions can achieve higher significance than a looser signal region, e.g. with at least 5 jets rather than at least 6 jets. The significances of the separate signal regions can then be combined together to achieve a greater significance for $t\bar{t}Z$ 2ℓOS.

7.2.7 Selections for 2ℓ-Z-2b5j

To achieve a greater significance for $t\bar{t}Z$ 2ℓOS by combining signal regions, the same process can be applied to the 2ℓ-Z-2b5j signal region of Table 5.2.2 to find a significance of 0.380, shown in Figure 7.2.8. The selections for the $t\bar{t}Z$ 2ℓOS channel 2ℓ-Z-2b5j region are shown in Table 7.2.3, along with the background they most help reduce. Significance achieved after making each selection sequentially is also shown in Table 7.2.3.

Variable	Selection	To reduce	Significance afterwards
Channel	e^+e^- or $\mu^+\mu^-$	$t\bar{t}$	0.197
M(ll)	80 < M(ll) < 100 GeV	$t\bar{t}$	0.179
N(Jets)	==5	Z	0.212
N(BJets)	≥2	Z	0.329
PT(ll)	>100 GeV	Z	0.350
MET	<130 GeV	$t\bar{t}$	0.360
DeltaPhi(l,l)	$<90^0$	Z	0.380

Table 7.2.3: Selections for the $t\bar{t}Z$ 2ℓOS Histogram Analyser 2ℓ-Z-2b5j region, along with the background process that each selection most helps reduce, and the significance achieved after making each selection. Significance quoted is by applying these selections in order.

7.2.8 Selections for 2ℓ-Z-1b6j

The same process can be applied to the 2ℓ-Z-1b6j signal region of Table 5.2.2 to find a maximum significance of 0.488, shown in Figure 7.2.9. The selections for the $t\bar{t}Z$ 2ℓOS channel 2ℓ-Z-1b6j region are shown in Table 7.2.4, along with the background they most help reduce. Significance achieved after making each selection sequentially is also shown in Table 7.2.4.

Variable	Selection	To reduce	Significance afterwards
Channel	e^+e^- or $\mu^+\mu^-$	$t\bar{t}$	0.197
M(ll)	80 < M(ll) < 100 GeV	$t\bar{t}$	0.179
N(Jets)	≥6	Z	0.522
N(BJets)	==1	Z	0.472
PT(ll)	>20 GeV	Z	0.483
DeltaR(l,l)	<3	Z	0.488

Table 7.2.4: Selections for the $t\bar{t}Z$ 2ℓOS Histogram Analyser 2ℓ-Z-1b6j region, along with the background process that each selection most helps reduce, and the significance achieved after making each selection. Significance quoted is by applying these selections in order.

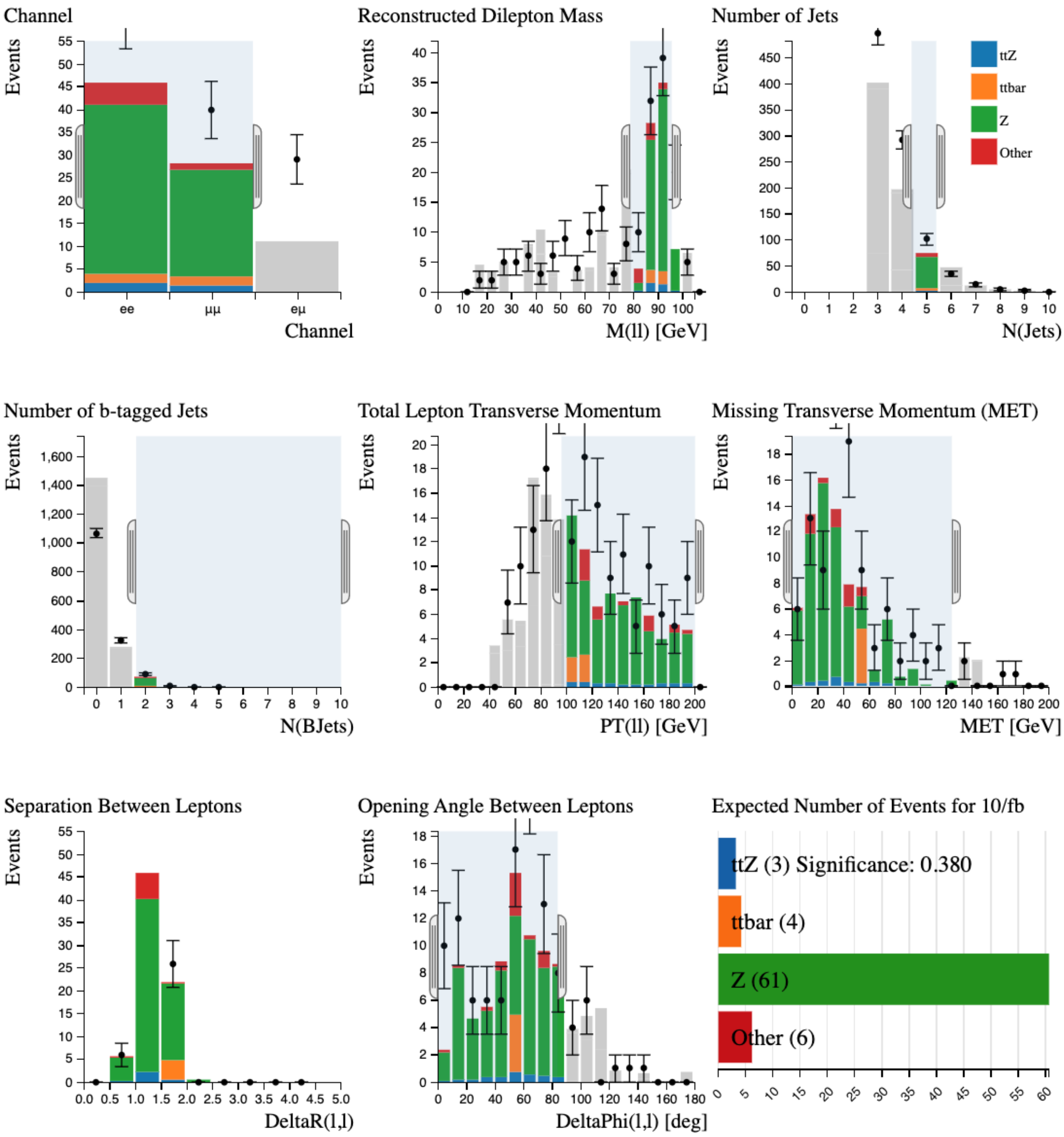


Figure 7.2.8: $t\bar{t}Z$ Histogram Analyser after applying selections for the 2ℓ-Z-2b5j signal region and optimising each variable. A significance of 0.380 is achieved.

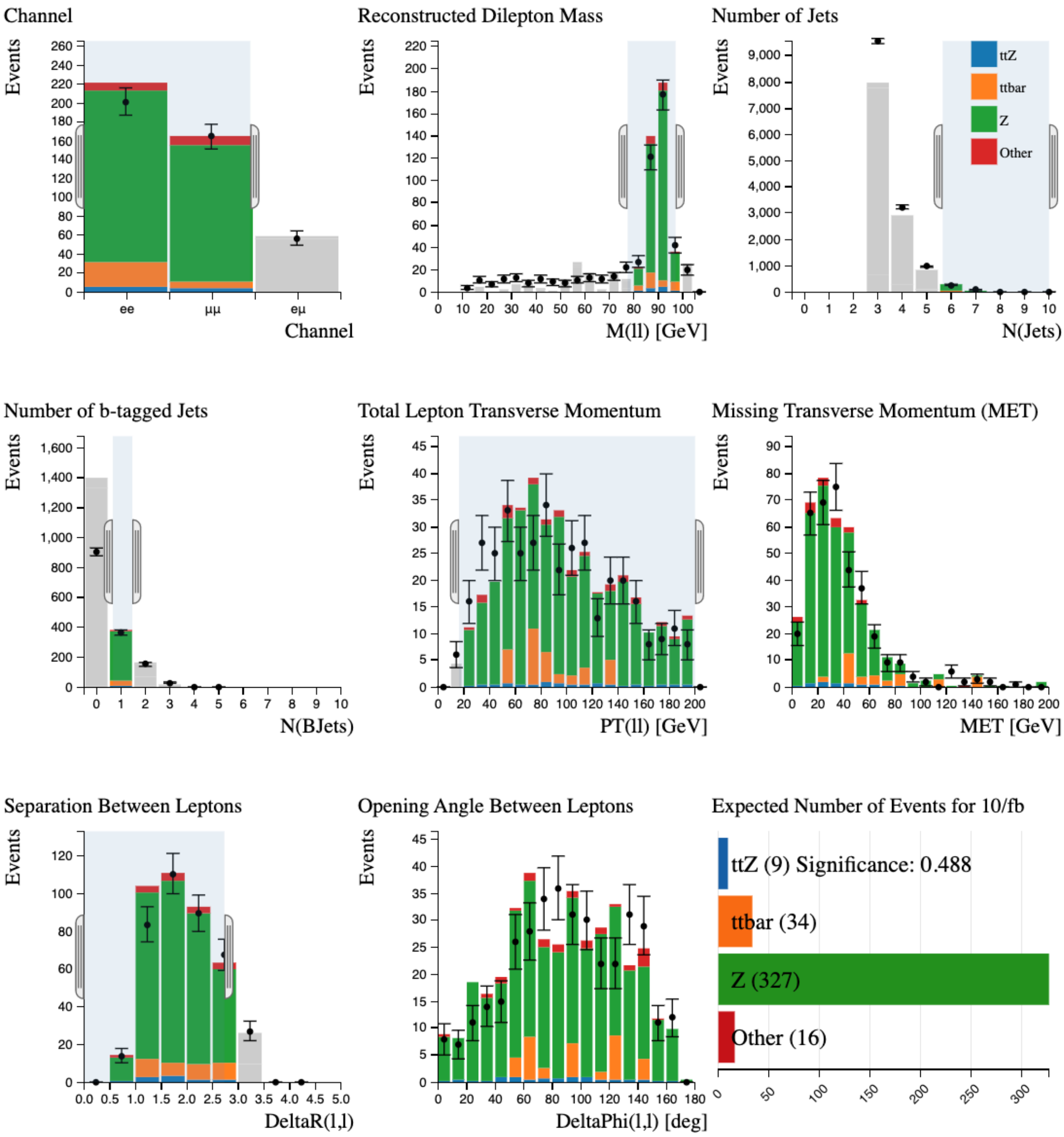


Figure 7.2.9: $t\bar{t}Z$ Histogram Analyser after applying selections for the 2ℓ-Z-1b6j signal region and optimising each variable. A significance of 0.488 is achieved.

7.2.9 Conclusion

This study indicates that an MVA will likely select:

- high PT(ll);
- low MET;
- low DeltaPhi(l,l);
- low DeltaR(l,l).

when building a signal-enriched region. No precise values can be given here because an MVA will optimise differently to the by-hand optimisation done in the Histogram Analyser. The fact that optimum selections for PT(ll), MET, DeltaR(l,l) and DeltaPhi(l,l) are different in the 3 regions illustrates why MVA training is conducted separately in different regions - because different regions will yield different optimum selections.

7.3 Jupyter notebooks

Jupyter notebooks [195] are a key online resource to introduce programming and coding, providing a very suitable arena for using ATLAS data for education. Several notebooks based on the $t\bar{t}Z$ 2ℓOS process were developed, as discussed during this section. They are presented here in sequential order of increasing difficulty.

7.3.1 Introduction

The release of the 13 TeV ATLAS Open Data was accompanied by a set of Jupyter notebooks that allow data analysis to be performed directly in a web browser [192, 196, 197]. Several notebooks with analysis examples are available, including analyses of $t\bar{t}Z$. The aim of many of these notebooks is to recreate published ATLAS results.

7.3.2 Analysis from csv

csv files are commonplace in data science outside of particle physics, therefore an analysis from csv files using ATLAS data is an opportunity to teach the transferrable skill of analysing csv files. As such, an example analysis starting from csv files and reproducing aspects of an ATLAS published result [36] is presented here.

Introduction

The csv analysis notebook [198] uses ATLAS Open Data to show the steps to implement Machine Learning in the $t\bar{t}Z$ 2ℓOS analysis, using the same input csv file as was used for the Histogram Analyser of Section 7.2. The steps taken throughout the notebook to recreate aspects of the ATLAS published result are:

1. tabulating the input data;
2. checking signal and background distributions for the variables present in the dataset;
3. checking separation between signal and background for the variables present in the dataset;

4. checking correlations between the variables present in the dataset;
5. training a MVA;
6. checking for overtraining of the MVA;
7. evaluating the performance of the MVA.

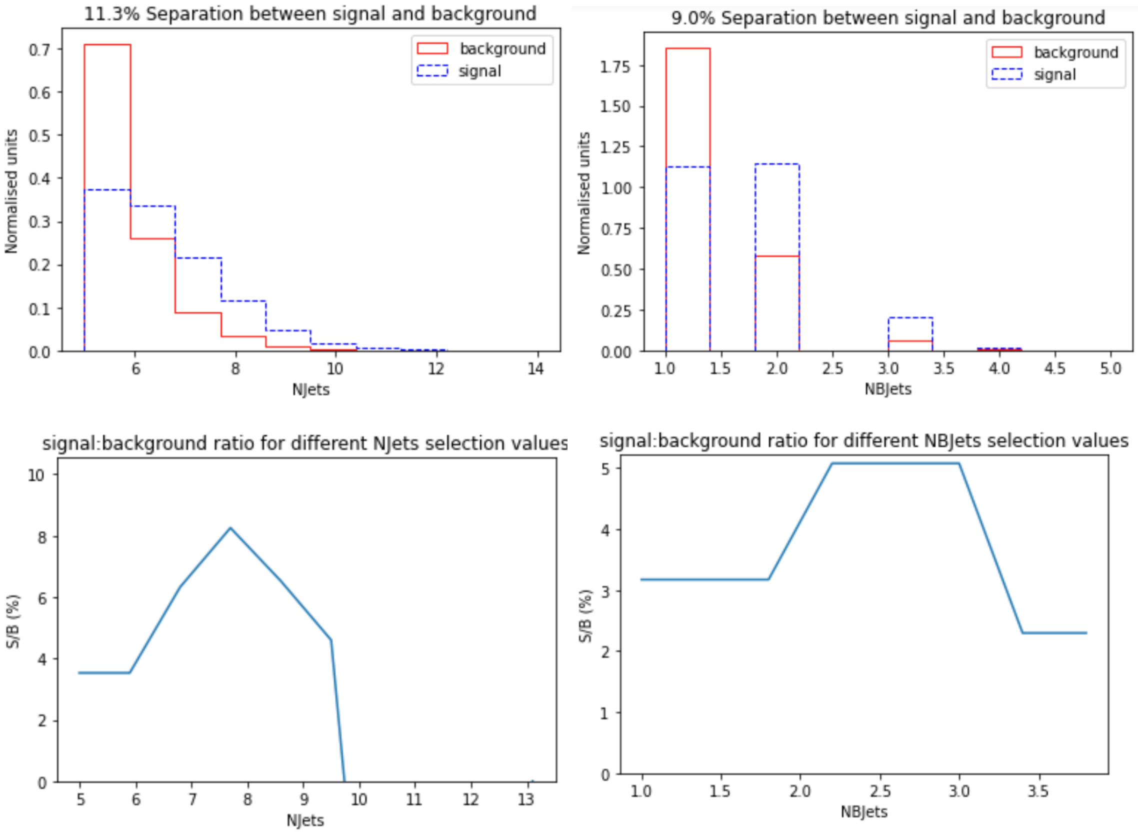
Selections

The fact that no $t\bar{t}Z$ 2ℓOS signal is visible immediately means that some selections have to be made. These selections are given in Table 7.3.1.

Reason	Code
e^+e^- or $\mu^+\mu^-$	Channel!=2
Number of jets	NJets $\geq$ 5
Number of b-jets	N(BJets) $\geq$ 1
Close to Z mass	|Mll - 91.12| < 10 GeV

Table 7.3.1: Initial selections applied to the input data in the Jupyter notebook introducing ML using $t\bar{t}Z$ 2ℓOS csv data.

After the selections of Table 7.3.1, a useful next step is to see how well signal and background are separated for each variable, and how high a signal-to-background ratio this can achieve. Such graphs are shown in Figure 7.3.1. Only 2 from 7 of the input variables are shown, for brevity.



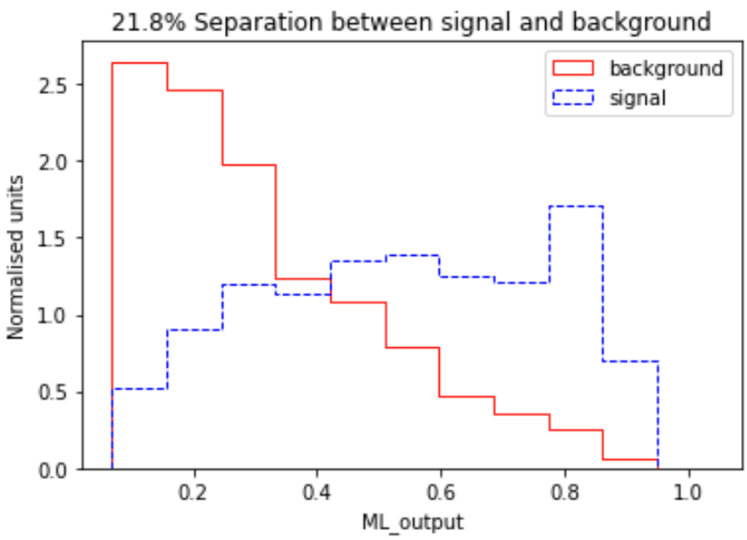
(a) Number of jets

(b) Number of b-jets

Figure 7.3.1: Separation between signal and background and signal-to-background ratio obtained by selecting above a particular value of the x-variable in question. Taking (a) NJets as an example, the starting x-value is 5. Taking the ratio of number of signal events with at least 5 jets, to the number of background events with at least 5 jets gives the S/B value at NJets=5 on the signal:background ratio plot (about 3.5%). Now imagine selecting only events with at least 7 jets. Taking the ratio of those events passing that selection gives the S/B value at NJets=7 on the signal:background ratio plot (about 6%). That is how the signal:background ratio plots are constructed.

Introducing Machine Learning

ML is introduced as a way to construct a variable that can achieve higher separation between signal and background and signal-to-background ratios. To achieve highest separation, ideally all variables would be used in the ML technique. However, for example, Mll cannot be used since values around the Z mass were selected, therefore using this sculpted distribution would lead to overtraining. To be sure all the other variables can be used, the correlations between them need to be checked. If a pair of variables is fully correlated (=1.0), using both would not add any new info. Having said this, some correlation is crucial, because this is what the ML technique exploits. No variable pair is correlated > 0.75 (absolute value), therefore each variable can be used. With a correlation check complete, the separation and signal-to-background ratio achievable using the 'ML_output' variable can be seen in Figure 7.3.2.



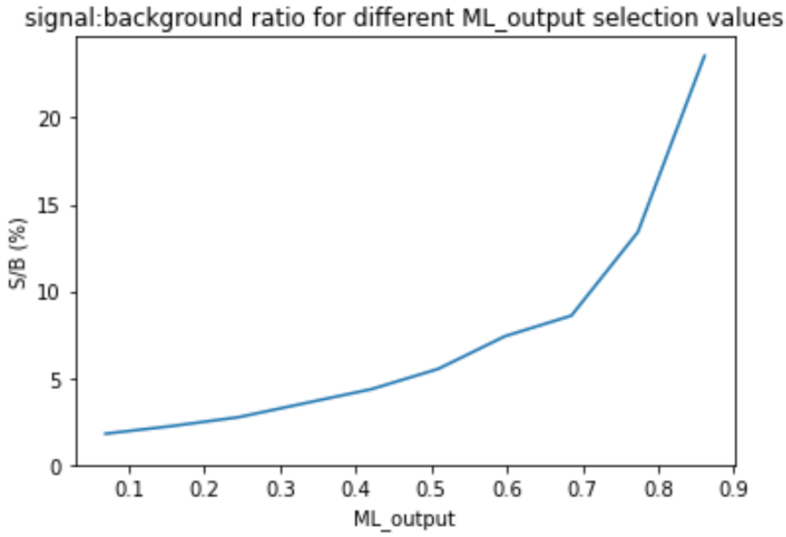


Figure 7.3.2: Separation between signal and background and signal-to-background ratio obtained by selecting above a particular value of 'ML_output'. The starting x-value is about 0.05. Taking the ratio of number of signal events with ML_output > 0.05, to the number of background events with ML_output > 0.05 gives the S/B value at ML_output = 0.05 on the signal:background ratio plot (about 2%). Now imagine selecting only events with ML_output > 0.6. Taking the ratio of those events passing that selection gives the S/B value at ML_output=0.6 on the signal:background ratio plot (about 8%). That is how the signal:background ratio plots are constructed.

ML_output compared to individual variables

The separation and S/B shown in Figure 7.3.2 is better than any of the individual variables of Figure 7.3.1 could ever have achieved. Recalling that $t\bar{t}Z$ 2ℓOS signal nominally produces at least 6 jets, including at least 2 b-jets, allows a further selection to be made, in an attempt to uncover some significant $t\bar{t}Z$ 2ℓOS signal.

Conclusion to the csv exploration notebook

After applying further selections, a significant amount of $t\bar{t}Z$ 2ℓOS signal can be seen above 0.8 in the ML_output distribution . Selecting ML_output > 0.8 would mostly eliminate background and achieve S/B 15%, as can be seen from Figure 7.3.2.

This technique of isolating signal at high ML_output allows to make precise measurements of the $t\bar{t}Z$ 2ℓOS signal process. In summary, this notebook introducing ML using $t\bar{t}Z$ shows that:

- putting data into an ML technique means only one variable has to be optimised;
- signal and background distributions are separated more when looking at ML output;
- ML achieves higher S/B than individual variables, because it finds multi-dimension correlations that give better S/B classification.

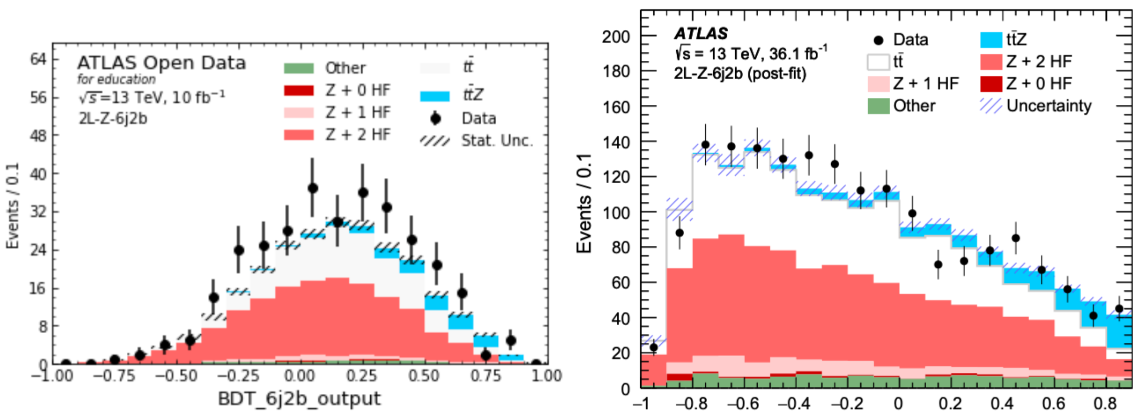
7.3.3 Full analysis

Having shown a simplified $t\bar{t}Z$ 2ℓOS analysis from csv files, similar principles can be extended to an analysis that fully reproduces a published ATLAS result [36]. The added complexity compared to the notebook of Section 7.3.2 includes:

- separating the analysis into 3 different signal regions;
- defining control regions;
- creating data-driven background estimates;
- ranking MVA input variables.

Introduction

The notebook presenting a full $t\bar{t}Z$ 2ℓOS analysis [199] uses ATLAS Open Data to show the steps to implement Machine Learning in the $t\bar{t}Z$ 2ℓOS analysis, following the ATLAS published paper “Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector" [36]. In particular, this notebook aims to recreate plots from Ref. [36] using a simplified ML workflow. The first plot that can be recreated is shown in Figure 7.3.3. Similar plots to Figure 7.3.3 are recreated for the 2ℓ-Z-2b5j and 2ℓ-Z-1b6j regions.



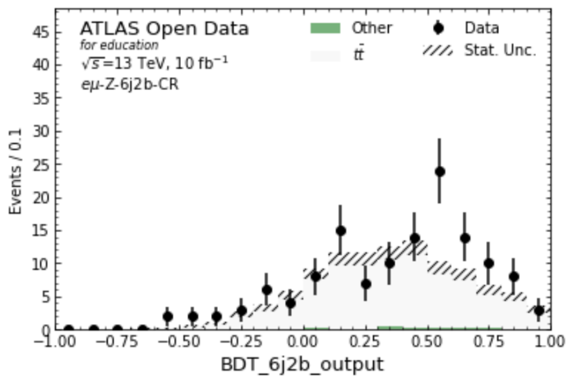
(a) Open Data

(b) Ref. [36]

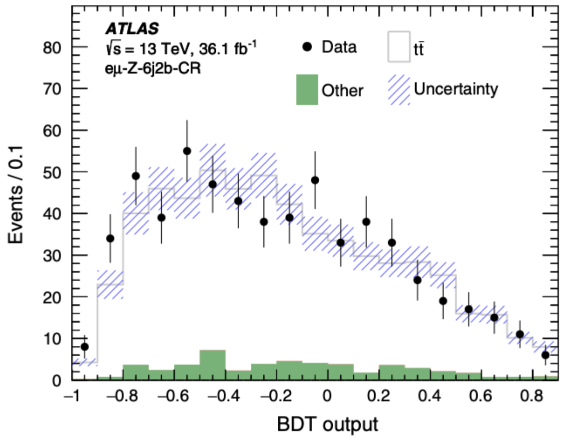
Figure 7.3.3: BDT output distributions in the signal region 2ℓ-Z-2b6j (here called 6j2b) using (a) ATLAS Open Data, (b) Ref. [36]. Considering the differences in the amount of data and the fact that not every detail from an ATLAS paper can be followed, the Open Data can reproduce this ATLAS result well. The ‘Other’ background contains SM processes with small cross sections producing two opposite-sign prompt leptons. The shaded band represents the total uncertainty. The last bin of each distribution contains the overflow.

Control regions

Plots in control regions can also be recreated, shown in Figure 7.3.4 for 2ℓ-Z-2b6j as an example.Equivalent plots for the 2ℓ-Z-2b5j and 2ℓ-Z-1b6j are also recreated.



(a) Open Data



(b) Ref. [36]

Figure 7.3.4: BDT output distributions in the $t\bar{t}$ control region of 2ℓ-Z-2b6j (here called 6j2b) using (a) ATLAS Open Data, (b) Ref. [36]. Considering the differences in the amount of data and the fact that not every detail from an ATLAS paper can be followed, the Open Data can reproduce this ATLAS result well. The 'Other' background contains SM processes with small cross sections producing two opposite-sign prompt leptons, including the $t\bar{t}Z$ process, whose contribution is negligible. The shaded band represents the total uncertainty. The last bin of each distribution contains the overflow.

Data-driven $t\bar{t}$ estimates

The $t\bar{t}$ control regions exampled in Figure 7.3.4 can then be used to build data-driven estimates of the $t\bar{t}$ contribution, rather than using the MC estimates in subfigure (a) of Figure 7.3.3.

Ranking input variables

Another result from Ref. [36] that can be recreated is Table 11, showing the definitions and ranking of input variables for the BDT. This comparison is shown in Figure 7.3.5.

Definition	6j1b	5j2b	6j2b
p_T of the lepton pair	15	14	15
p_T of the 4th jet	5	1	8
p_T of the 5th jet	-	8	-
p_T of the 6th jet	2	-	2
ΔR_η between the two leptons	6	4	7
Number of jet pairs with mass within a window of 30 GeV around 85 GeV	1	2	3
Number of top-quark candidates	-	-	1
Invariant mass of the two jets with the smallest ΔR_η	13	10	17
Invariant mass of the two untagged jets with the highest p_T	9	11	-
Invariant mass of the two jets with the highest value of the b-tagging discriminant	-	5	4
Scalar sum of p_T divided by the sum of energy of all jets	14	13	16
Average ΔR_η of all jet pairs	11	3	10
Maximum invariant mass of a lepton and the b-tagged jet with the smallest ΔR_η	10	-	13
First Fox–Wolfram moment built from jets and leptons	12	12	14
Sum of jet p_T, using up to six jets	4	6	5
η of dilepton system	3	9	9
Sum of the two closest two-jet invariant masses from jjj1 and jjj2 divided by two	7	-	11
ΔR_η between two jets with the highest value of the b-tagging discriminant in the event	-	7	6
p_T of the b-tagged jet with the highest p_T	8	-	12

(a) Open Data

Definition	6j1b	5j2b	6j2b
p_T of the lepton pair	8	11	8
p_T of the 4th jet	6	12	6
p_T of the 5th jet	-	14	-
p_T of the 6th jet	9	-	11
ΔR_η between the two leptons	7	8	12
Number of jet pairs with mass within a window of 30 GeV around 85 GeV	4	6	4
Number of top-quark candidates	-	-	17
Invariant mass of the two jets with the smallest ΔR_η	13	7	14
Invariant mass of the two untagged jets with the highest p_T	15	13	-
Invariant mass of the two jets with the highest value of the b-tagging discriminant	-	10	9
Scalar sum of p_T divided by the sum of energy of all jets	2	1	2
Average ΔR_η of all jet pairs	5	4	5
Maximum invariant mass of a lepton and the b-tagged jet with the smallest ΔR_η	14	-	13
First Fox–Wolfram moment built from jets and leptons	3	2	1
Sum of jet p_T, using up to six jets	12	5	10
η of dilepton system	1	3	3
Sum of the two closest two-jet invariant masses from jjj1 and jjj2 divided by two	10	-	15
ΔR_η between two jets with the highest value of the b-tagging discriminant in the event	-	9	7
p_T of the b-tagged jet with the highest p_T	11	-	16

(b) Ref. [36]

Figure 7.3.5: The definitions and ranking of input variables for the BDT in the $t\bar{t}Z$ 2ℓOS analysis. (a) ATLAS Open Data, (b) Ref. [36]. Some similarities can be seen between (a) and (b), for example "Number of jet pairs with mass within a window of 30 GeV around 85 GeV" ranking rather highly for both. Differences between (a) and (b) can also be seen, for example "Scalar sum of p_T divided by the sum of energy of all jets" ranking highly for (b) but not so highly for (a). Jets and leptons are ordered in descending order of p_T. Only the first eight jets are considered when calculating the input variables.

Conclusion to the full analysis notebook

Using ATLAS Open Data, a full analysis of the $t\bar{t}Z$ process can be undertaken, reproducing simplified versions of the results from an ATLAS published paper [36]. Signal and control region plots can be reproduced in the same format as the ATLAS published paper [36]. The method of obtaining data-driven $t\bar{t}$ estimates used in the ATLAS published paper [36] can also be reproduced using ATLAS Open Data. The ranking of most important variables in the MVA with ATLAS Open Data in the $t\bar{t}Z$ 2ℓOS channel show similarities to the ranking of the most important variables in the MVA from the ATLAS published paper [36].

7.4 Comparisons with full ATLAS data

This section compares results from Section 7.2 and Section 7.3.3 using 10 fb^{-1}of ATLAS Open Data in simplified analyses to Section 6 using 139.0 fb^{-1}of full Run 2 ATLAS data in a full analysis. Results will be compared in terms of:

- ranking of variables by the MVAs;
- statistical significance achievable.

7.4.1 Comparison of variable ranking between Open Data and binary BDTs

Table 6.1.2 ranking input variables using BDTs with Full Run 2 data can be compared side-by-side with the information from Figure 7.3.5 ranking input variables using BDTs with ATLAS Open Data. This comparison is shown in Table 7.4.1. A number of similarities can be seen, e.g. $N_{jj}^{V mass}$ is ranked within the top 4 in each of the six BDTs, or that $\mathrm{p_T^{ll}}$ is ranked within the bottom 3 in each of the six BDTs. However, differences can be seen also, perhaps the most stark being that $N_{bjj}^{top-mass}$

is ranked 1st in the 2b6j Open Data BDT yet 16th in the 2b6j Full Run 2 BDT. This suggests that some variables are important over a range of amount of data available, whereas other variables only become more important when more data are available.

	1b6j		2b5j		2b6j	
rank	Open Data	Full Run 2	Open Data	Full Run 2	Open Data	Full Run 2
1	N_{jj}^{Vmass}	$\mathrm{H_T^{6jets}}$	$\mathrm{p_T^{4jet}}$	$\mathrm{H_T^{6jets}}$	$N_{bjj}^{top-mass}$	$\mathrm{H_T^{6jets}}$
2	$\mathrm{p_T^{6jet}}$	η_{ll}	N_{jj}^{Vmass}	ΔR_{jj}^{ave}	$\mathrm{p_T^{6jet}}$	ΔR_{ll}
3	η_{ll}	N_{jj}^{Vmass}	ΔR_{jj}^{ave}	N_{jj}^{Vmass}	N_{jj}^{Vmass}	η_{ll}
4	$\mathrm{H_T^{6jets}}$	$\mathrm{p_T^{b1}}$	ΔR_{ll}	$\mathrm{M_{bb}^{pTord}}$	$\mathrm{M_{bb}^{pTord}}$	N_{jj}^{Vmass}
5	$\mathrm{p_T^{4jet}}$	$\mathrm{MaxM_{lepb}^{mindR}}$	$\mathrm{M_{bb}^{pTord}}$	$\mathrm{M_{jj}^{mindR}}$	$\mathrm{H_T^{6jets}}$	$\Delta R_{\mathrm{ave}}^{\mathrm{jj}}$
6	ΔR_{ll}	$\mathrm{M_{jj}^{mindR}}$	$\mathrm{H_T^{6jets}}$	ΔR_{ll}	ΔR_{bb}	ΔR_{bb}
7	$\mathrm{M_W^{avg}}$	$\mathrm{p_T^{4jet}}$	ΔR_{bb}	ΔR_{bb}	ΔR_{ll}	$\mathrm{p_T^{6jet}}$
8	$\mathrm{p_T^{b1}}$	ΔR_{ll}	$\mathrm{p_T^{5jet}}$	$\mathrm{p_T^{4jet}}$	$\mathrm{p_T^{4jet}}$	$\mathrm{MaxM_{lepb}^{mindR}}$
9	$\mathrm{M_{uu}^{pTord}}$	$\mathrm{p_T^{6jet}}$	η_{ll}	η_{ll}	η_{ll}	$\mathrm{M_W^{avg}}$
10	$\mathrm{MaxM_{lepb}^{mindR}}$	$\mathrm{M_W^{avg}}$	$\mathrm{M_{jj}^{mindR}}$	$\mathrm{p_T^{5jet}}$	ΔR_{jj}^{ave}	$\mathrm{M_{jj}^{mindR}}$
11	ΔR_{jj}^{ave}	$Centr_{jet}$	$\mathrm{M_{uu}^{pTord}}$	$\mathrm{M_{uu}^{pTord}}$	$\mathrm{M_W^{avg}}$	$\mathrm{p_T^{4jet}}$
12	H1	H1	H1	H1	$\mathrm{p_T^{b1}}$	$\mathrm{M_{bb}^{pTord}}$
13	$\mathrm{M_{jj}^{mindR}}$	$\mathrm{M_{uu}^{pTord}}$	$Centr_{jet}$	$\mathrm{p_T^{ll}}$	$\mathrm{MaxM_{lepb}^{mindR}}$	$Centr_{jet}$
14	$Centr_{jet}$	ΔR_{jj}^{ave}	$\mathrm{p_T^{ll}}$	$Centr_{jet}$	H1	H1
15	$\mathrm{p_T^{ll}}$	$\mathrm{p_T^{ll}}$			$\mathrm{p_T^{ll}}$	$\mathrm{p_T^{b1}}$
16					$Centr_{jet}$	$N_{bjj}^{top-mass}$
17					$\mathrm{M_{jj}^{mindR}}$	$\mathrm{p_T^{ll}}$

Table 7.4.1: Comparison of ranking of the variables used for BDT training, when using a single BDT per 2ℓOS region. The comparison is performed between the BDTs using ATLAS Open Data and the BDTs using Full Run 2 data.

7.4.2 Comparison of variable ranking between Open Data and binary DNNs

Figure 6.2.1 ranking variables using DNNs with Full Run 2 data can be compared side-by-side with the information from Figure 7.3.5 ranking variables using BDTs with ATLAS Open Data. This comparison is shown in Table 7.4.2. A number of similarities can be seen, e.g. N_{jj}^{Vmass} is ranked within the top 3 in each of the six MVAs, or that $\mathrm{p_T^{ll}}$ is ranked within the bottom 3 in each of the six MVAs. However, differences can be seen also, perhaps the most stark being that $\mathrm{p_T^{6jet}}$ and $\mathrm{p_T^{5jet}}$ are ranked much higher in the Open Data BDTs than they are in the Full Run 2 DNNs. This again suggests that some variables are important over a range of amount of data available, whereas other variables only become more important when more data are available.

	1b6j		2b5j		2b6j	
rank	Open Data	Full Run 2	Open Data	Full Run 2	Open Data	Full Run 2
1	N_{jj}^{Vmass}	$\mathrm{H_T^{6jets}}$	$\mathrm{p_T^{4jet}}$	$\mathrm{H_T^{6jets}}$	$N_{bjj}^{top-mass}$	$\mathrm{H_T^{6jets}}$
2	$\mathrm{p_T^{6jet}}$	η_{ll}	N_{jj}^{Vmass}	ΔR_{ll}	$\mathrm{p_T^{6jet}}$	N_{jj}^{Vmass}
3	η_{ll}	N_{jj}^{Vmass}	ΔR_{jj}^{ave}	$Centr_{jet}$	N_{jj}^{Vmass}	ΔR_{ll}
4	$\mathrm{H_T^{6jets}}$	H1	ΔR_{ll}	N_{jj}^{Vmass}	$\mathrm{M_{bb}^{pTord}}$	$N_{bjj}^{top-mass}$
5	$\mathrm{p_T^{4jet}}$	ΔR_{ll}	$\mathrm{M_{bb}^{pTord}}$	$\mathrm{M_{bb}^{pTord}}$	$\mathrm{H_T^{6jets}}$	$Centr_{jet}$
6	ΔR_{ll}	$\mathrm{p_T^{4jet}}$	$\mathrm{H_T^{6jets}}$	H1	ΔR_{bb}	ΔR_{bb}
7	$\mathrm{M_W^{avg}}$	$Centr_{jet}$	ΔR_{bb}	ΔR_{jj}^{ave}	ΔR_{ll}	$\mathrm{p_T^{4jet}}$
8	$\mathrm{p_T^{b1}}$	$\mathrm{p_T^{b1}}$	$\mathrm{p_T^{5jet}}$	$\mathrm{p_T^{4jet}}$	$\mathrm{p_T^{4jet}}$	H1
9	$\mathrm{M_{uu}^{pTord}}$	$\mathrm{p_T^{6jet}}$	η_{ll}	$\mathrm{M_{jj}^{mindR}}$	η_{ll}	$\mathrm{p_T^{6jet}}$
10	$\mathrm{MaxM_{lepb}^{mindR}}$	ΔR_{jj}^{ave}	$\mathrm{M_{jj}^{mindR}}$	ΔR_{bb}	ΔR_{jj}^{ave}	η_{ll}
11	ΔR_{jj}^{ave}	$\mathrm{M_W^{avg}}$	$\mathrm{M_{uu}^{pTord}}$	$\mathrm{p_T^{5jet}}$	$\mathrm{M_W^{avg}}$	$\mathrm{p_T^{b1}}$
12	H1	$\mathrm{MaxM_{lepb}^{mindR}}$	H1	η_{ll}	$\mathrm{p_T^{b1}}$	$\mathrm{M_{bb}^{pTord}}$
13	$\mathrm{M_{jj}^{mindR}}$	$\mathrm{M_{jj}^{mindR}}$	$Centr_{jet}$	$\mathrm{p_T^{ll}}$	$\mathrm{MaxM_{lepb}^{mindR}}$	$\mathrm{M_W^{avg}}$
14	$Centr_{jet}$	$\mathrm{p_T^{ll}}$	$\mathrm{p_T^{ll}}$	$\mathrm{M_{uu}^{pTord}}$	H1	ΔR_{jj}^{ave}
15	$\mathrm{p_T^{ll}}$	$\mathrm{M_{uu}^{pTord}}$			$\mathrm{p_T^{ll}}$	$\mathrm{p_T^{ll}}$
16					$Centr_{jet}$	$\mathrm{M_{jj}^{mindR}}$
17					$\mathrm{M_{jj}^{mindR}}$	$\mathrm{MaxM_{lepb}^{mindR}}$

Table 7.4.2: Comparison of ranking of the variables used for MVA training. The comparison is performed between the BDTs using ATLAS Open Data and the initial DNNs using Full Run 2 data.

7.4.3 Statistical significance comparison between Histogram Analyser and initial multiclass DNN

The statistical significance from Figure 6.2.5 can be compared to the significance achievable from the Histogram Analyser discussed in Section 7.2, whose final significances are shown in Figure 7.2.7, Figure 7.2.8 and Figure 7.2.9 for the 2ℓ-Z-2b6j, 2ℓ-Z-2b5j and 2ℓ-Z-1b6j channels respectively. This comparison is shown in Table 7.4.3. The Histogram Analyser only uses about 1/14th of the data used for the DNNs of Section 6.3 as this is all of the 13 TeV data currently made open by ATLAS. A more direct comparison can be made by scaling the Histogram Analyser significances by the square root of the ratio between the full Run 2 luminosity and the luminosity used in ATLAS Open Data, $\sqrt{139.0/10}$, because statistical significance scales with the square root of number of events. Even the scaled statistical significances achievable by the Histogram Analyser are about 2.5 times less than the statistical significances achievable by the DNNs. This hints at the power of DNNs in optimising for statistical significance in the $t\bar{t}Z$ 2ℓOS analysis, compared to a cut-and-count analysis.

Channel	Histogram Analyser significance	Histogram Analyser significance (scaled)	DNN significance
2b6j	0.971 (Figure 7.2.7)	3.620	10.8
2b5j	0.380 (Figure 7.2.8)	1.417	4.9
1b6j	0.488 (Figure 7.2.9)	1.819	4.5

Table 7.4.3: A comparison of the statistical significance that can be achieved using the DNNs of Section 6.3, with the Histogram Analyser of Section 7.2. It is important to remember that the Histogram Analyser uses about 1/14th of the data used for the DNNs of Section 6.3.

Bibliography

[1] *Lakers need to break from their routine*, 2002, URL: https://www.latimes.com/archives/la-xpm-2002-may-09-sp-adande09-story.html (cit. on p. 27).

[2] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579, ISSN: 0029-5582, URL: http://www.sciencedirect.com/science/article/pii/0029558261904692 (cit. on pp. 27, 29, 36).

[3] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264, URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264 (cit. on pp. 27, 29, 36).

[4] J. Goldstone, A. Salam and S. Weinberg, *Broken symmetries*, Physical Review **127** (1962) 965 (cit. on pp. 27, 29, 36).

[5] G. 't Hooft and M. Veltman, *Regularization and renormalization of gauge fields*, Nuclear Physics B **44** (1972) 189, ISSN: 0550-3213, URL: http://www.sciencedirect.com/science/article/pii/0550321372902799 (cit. on p. 27).

[6] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001, URL: https://doi.org/10.1088/1748-0221/3/08/s08001 (cit. on pp. 27, 42).

[7] T. Aaltonen et al., *Combination of the top-quark mass measurements from the Tevatron collider*, Physical Review D **86** (2012) 092003 (cit. on p. 27).

[8] S. Myers and E. Picasso, *The LEP collider*, Scientific American **263** (1990) 54 (cit. on p. 27).

[9] F. Abe et al., *Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74** (14 1995) 2626, URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2626 (cit. on pp. 27, 35).

[10] S. Abachi et al., *Observation of the Top Quark*, Phys. Rev. Lett. **74** (14 1995) 2632, URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2632 (cit. on pp. 27, 35).

[11] Particle Data Group, *Review of Particle Physics*, PTEP **8** (2020) 083C01 (cit. on pp. 27, 31, 34–37, 39, 53, 54, 63, 73).

[12] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s}$ = 13 TeV with the ATLAS experiment*, (2020), arXiv: 2006.13076 [hep-ex] (cit. on pp. 27, 35).

[13] A. Datta and M. Duraisamy, *Model independent predictions for rare top decays with weak coupling*, Physical Review D **81** (2010) 074008 (cit. on p. 27).

[14] G. Arnison et al., *Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider*, Physics Letters B **126** (1983) 398, ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/0370269383901880 (cit. on pp. 27, 36).

[15] P. Bagnaia et al., *Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ collider*, Physics Letters B **129** (1983) 130, ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/037026938390744X (cit. on pp. 27, 36).

[16] A. Kulesza, L. Motyka, D. Schwartländer, T. Stebel and V. Theeuwes, *Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy*, Eur. Phys. J. C **79** (2019) 249, arXiv: 1812.08622 [hep-ph] (cit. on pp. 27, 38).

[17] R. S. Chivukula, S. B. Selipsky and E. H. Simmons, *Nonoblique effects in the $Zb\bar{b}$ vertex from ETC dynamics*, Phys. Rev. Lett. **69** (1992) 575, arXiv: hep-ph/9204214 (cit. on p. 27).

[18] R. Chivukula, E. H. Simmons and J. Terning, *A Heavy top quark and the $Zb\bar{b}$ vertex in noncommuting extended technicolor*, Phys. Lett. **B 331** (1994) 383, arXiv: hep-ph/9404209 (cit. on p. 27).

[19] K. Hagiwara and N. Kitazawa, *Extended technicolor contribution to the Zbb vertex*, Phys. Rev. **D 52** (1995) 5374, arXiv: hep-ph/9504332 (cit. on p. 27).

[20] U. Mahanta, *Noncommuting ETC corrections to $Zt\bar{t}$ vertex*, Phys. Rev. **D 55** (1997) 5848, arXiv: hep-ph/9611289 (cit. on p. 27).

[21] U. Mahanta, *Probing noncommuting ETC effects by $e^+e^- \to t\bar{t}$ at NLC*, Phys. Rev. **D 56** (1997) 402 (cit. on p. 27).

[22] M. Perelstein, *Little Higgs models and their phenomenology*, Prog. Part. Nucl. Phys. **58** (2007) 247, arXiv: hep-ph/0512128 (cit. on p. 27).

[23] ATLAS Collaboration, *Search for direct top squark pair production in events with a Higgs or Z boson, and missing transverse momentum in $\sqrt{s}$ = 13 TeV pp collisions with the ATLAS detector*, JHEP **08** (2017) 006, arXiv: 1706.03986 [hep-ex] (cit. on p. 27).

[24] ATLAS Collaboration, *Search for new phenomena in events with same-charge leptons and b-jets in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **12** (2018) 039, arXiv: 1807.11883 [hep-ex] (cit. on p. 27).

[25] ATLAS Collaboration, *Search for squarks and gluinos in final states with same-sign leptons and jets using 139 fb^{-1} of data collected with the ATLAS detector*, JHEP **06** (2020) 046, arXiv: 1909.08457 [hep-ex] (cit. on p. 27).

[26] ATLAS Collaboration, *Analysis of events with b-jets and a pair of leptons of the same charge in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, JHEP **10** (2015) 150, arXiv: 1504.04605 [hep-ex] (cit. on p. 27).

[27] ATLAS Collaboration, *Search for pair production of up-type vector-like quarks and for four-top-quark events in final states with multiple b-jets with the ATLAS detector*, JHEP **07** (2018) 089, arXiv: 1803.09678 [hep-ex] (cit. on p. 27).

[28] ATLAS Collaboration, *Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s}$=13 TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 1085 (cit. on p. 27).

[29] CMS Collaboration, *Search for physics beyond the standard model in events with two leptons of same sign, missing transverse momentum, and jets in proton–proton collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **77** (2017) 578, arXiv: 1704.07323 [hep-ex] (cit. on p. 27).

[30] CMS Collaboration, *Search for standard model production of four top quarks with same-sign and multilepton final states in proton–proton collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **78** (2018) 140, arXiv: 1710.10614 [hep-ex] (cit. on p. 27).

[31] ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. D **97** (2018) 072003, arXiv: 1712.08891 [hep-ex] (cit. on p. 27).

[32] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. B **784** (2018) 173, arXiv: 1806.00425 [hep-ex] (cit. on p. 27).

[33] ATLAS Collaboration, *Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **07** (2020) 124, arXiv: 2002.07546 [hep-ex] (cit. on p. 27).

[34] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47, arXiv: 1909.00761 [hep-ex] (cit. on pp. 28, 59).

[35] 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. 28, 65, 66).

[36] ATLAS Collaboration, *Measurement of the* $t\bar{t}Z$ *and* $t\bar{t}W$ *cross sections in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. D **99** (2019) 072009, arXiv: 1901.03584 [hep-ex] (cit. on pp. 28, 84, 92, 101, 102, 107, 114, 121, 137, 139, 150, 155, 158–160, 165).

[37] ATLAS Collaboration, *Measurements of the inclusive and differential production cross sections of a top-quark-antiquark pair in association with a Z boson at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, (2020), arXiv: 2103.12603 [hep-ex] (cit. on pp. 28, 69, 76, 79, 81, 139).

[38] *Snow White never could read particle physics for very long.: Mean girl quotes, that one person, funny confessions*, 2012, URL: https://www.pinterest.com/pin/107101297358534412/ (cit. on p. 29).

[39] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321 (cit. on p. 29).

[40] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on p. 29).

[41] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585 (cit. on p. 29).

[42] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (4 1966) 1156, URL: https://link.aps.org/doi/10.1103/PhysRev.145.1156 (cit. on p. 29).

[43] M. Gell-Mann, *A schematic model of baryons and mesons*, Physics Letters **8** (1964) 214 (cit. on p. 29).

[44] *Standard Model of Elementary Particles*, 2022, URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (cit. on p. 30).

[45] S. Bethke, *Determination of the QCD coupling* α_S, Journal of Physics G: Nuclear and Particle Physics **26** (2000) R27, URL: https://doi.org/10.1088/0954-3899/26/7/201 (cit. on p. 30).

[46] G. Gabrielse, D. Hanneke, T. Kinoshita, M. Nio and B. Odom, *New Determination of the Fine Structure Constant from the Electron g Value and QED*, Phys. Rev. Lett. **97** (3 2006) 030802, URL: https://link.aps.org/doi/10.1103/PhysRevLett.97.030802 (cit. on pp. 30, 33).

[47] J. W. Rohlf, *Modern Physics from A to Z0 Chapter 18*, 1994 (cit. on p. 30).

[48] *How many fundamental constants does it take to explain the universe?*, URL: https://ep-news.web.cern.ch/content/how-many-fundamental-constants-does-it-take-explain-universe (cit. on p. 32).

[49] M. Peskin, *An introduction to quantum field theory*, CRC press, 2018 (cit. on p. 32).

[50] A. I. Akhiezer, *Quantum electrodynamics*, (1965) (cit. on p. 33).

[51] S. L. Glashow, *Towards a unified theory: Threads in a tapestry*, Reviews of Modern Physics **52** (1980) 539 (cit. on p. 33).

[52] S. Weinberg, *Conceptual foundations of the unified theory of weak and electromagnetic interactions*, Science **210** (1980) 1212 (cit. on p. 33).

[53] A. Salam, *Gauge unification of fundamental forces*, Reviews of Modern Physics **52** (1980) 525 (cit. on p. 33).

[54] P. Langacker, *Grand unified theories and proton decay*, Physics Reports **72** (1981) 185 (cit. on p. 33).

[55] W. Marciano and H. Pagels, *Quantum chromodynamics*, Physics Reports **36** (1978) 137 (cit. on p. 33).

[56] H. D. Politzer, *Asymptotic freedom: An approach to strong interactions*, Physics Reports **14** (1974) 129 (cit. on p. 33).

[57] A. M. Polyakov, *Quark confinement and topology of gauge theories*, Nuclear Physics B **120** (1977) 429 (cit. on p. 33).

[58] *Argonne National Laboratory*, URL: https://www.anl.gov/phy/3d-structure-of-protons-and-neutrons (cit. on p. 34).

[59] M. Kobayashi and T. Maskawa, *CP-violation in the renormalizable theory of weak interaction*, Progress of theoretical physics **49** (1973) 652 (cit. on p. 35).

[60] L. H. Orr and J. L. Rosner, *Comparison of top quark hadronization and decay rates*, Physics Letters B **246** (1990) 221 (cit. on p. 35).

[61] T. Humanic, *Extracting the hadronization timescale in TeV proton–proton collisions from pion and kaon femtoscopy*, Journal of Physics G: Nuclear and Particle Physics **41** (2014) 075105 (cit. on p. 35).

[62] M. Carena, M. Olechowski, S. Pokorski and C. Wagner, *Electroweak symmetry breaking and bottom-top Yukawa unification*, Nuclear Physics B **426** (1994) 269 (cit. on p. 35).

[63] ATLAS Collaboration, *Top working group cross-section summary plots: Autumn 2020*, ATL-PHYS-PUB-2020-029, 2020, URL: http://cds.cern.ch/record/2746598 (cit. on p. 37).

[64] A. Aivazis, *Draw feynman diagram online*, URL: https://feynman.aivazis.com/ (cit. on pp. 38, 55, 74).

[65] C. Monini, 'Single-top s channel cross-section measurement with the ATLAS detector', PhD thesis, 2014 (cit. on p. 39).

[66] G. Van Rossum and F. L. Drake Jr, *Python reference manual*, Centrum voor Wiskunde en Informatica Amsterdam, 1995 (cit. on pp. 40, 73).

[67] B. Aaronovitch, *Moon Over Soho: Book 2 in the# 1 bestselling Rivers of London series*, Hachette UK, 2011 (cit. on p. 41).

[68] L. Arnaudon et al., *Linac4 technical design report*, tech. rep., 2006 (cit. on p. 42).

[69] K. Hanke, *Past and present operation of the CERN PS Booster*, International Journal of Modern Physics A **28** (2013) 1330019 (cit. on p. 42).

[70] S. Gilardoni et al., *Fifty years of the CERN Proton Synchrotron: Volume 2*, arXiv preprint arXiv:1309.6923 (2013) (cit. on p. 42).

[71] H. van Hees and R. Rapp, *Dilepton radiation at the CERN super-proton synchrotron*, Nuclear Physics A **806** (2008) 339 (cit. on p. 42).

[72] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 42, 50, 51).

[73] E. Mobs, *The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019*, (2019), General Photo, URL: https://cds.cern.ch/record/2684277 (cit. on p. 42).

[74] J.J. Goodson, 'Search for Supersymmetry in States with Large Missing Transverse Momentum and Three Leptons including a Z-Boson', Presented 17 Apr 2012, PhD thesis: Stony Brook University, 2012 (cit. on p. 44).

[75] J. Pequenao, 'Computer generated image of the whole ATLAS detector', 2008, URL: http://cds.cern.ch/record/1095924 (cit. on p. 45).

[76] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run-2*, (2020), arXiv: 2007.07624 [hep-ex] (cit. on pp. 46, 47).

[77] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 47).

[78] ATLAS Collaboration, *ATLAS Pixel Detector: Technical Design Report*, ATLAS-TDR-11; CERN-LHCC-98-013, 1998, URL: https://cds.cern.ch/record/381263 (cit. on p. 47).

[79] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, JINST **9** (2014) P08009, arXiv: 1404.7473 [hep-ex] (cit. on p. 48).

[80] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties*, JINST **12** (2017) P05002, arXiv: 1702.06473 [hep-ex] (cit. on p. 48).

[81] ATLAS Collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, Eur. Phys. J. C **70** (2010) 1193, arXiv: 1007.5423 [hep-ex] (cit. on pp. 48, 49).

[82] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter: Technical Design Report*, ATLAS-TDR-2; CERN-LHCC-96-041, 1996, URL: https://cds.cern.ch/record/331061 (cit. on p. 48).

[83] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, ATLAS-TDR-3; CERN-LHCC-96-042, 1996, URL: https://cds.cern.ch/record/331062 (cit. on p. 48).

[84] ATLAS Collaboration, *ATLAS Muon Spectrometer: Technical Design Report*, ATLAS-TDR-10; CERN-LHCC-97-022, CERN, 1997, URL: https://cds.cern.ch/record/331068 (cit. on p. 50).

[85] T. Colombo, *Data-flow Performance Optimisation on Unreliable Networks: the ATLAS Data-Acquisition Case*, Journal of Physics: Conference Series **608** (2015) 012005 (cit. on p. 52).

[86] Proceedings of the Royal Institution of Great Britain **35** (1951) (cit. on p. 53).

[87] M. I. Jordan and T. M. Mitchell, *Machine learning: Trends, perspectives, and prospects*, Science **349** (2015) 255 (cit. on p. 53).

[88] R. Caruana and A. Niculescu-Mizil, 'An empirical comparison of supervised learning algorithms', *Proceedings of the 23rd international conference on Machine learning*, 2006 161 (cit. on p. 53).

[89] S. Höche, 'Introduction to parton-shower event generators', *Journeys Through the Precision Frontier: Amplitudes for Colliders: TASI 2014 Proceedings of the 2014 Theoretical Advanced Study Institute in Elementary Particle Physics*, World Scientific, 2016 235 (cit. on p. 53).

[90] J. C. Collins, D. E. Soper and G. Sterman, 'Factorization of hard processes in QCD', *Perturbative QCD*, World Scientific, 1989 1 (cit. on p. 53).

[91] R. D. Ball et al., *Parton distributions with LHC data*, Nuclear Physics B **867** (2013) 244 (cit. on p. 53).

[92] S. Carrazza, S. Forte and J. Rojo, *Parton distributions and event generators*, arXiv preprint arXiv:1311.5887 (2013) (cit. on p. 53).

[93] F. E. Close, *Quarks and partons*, tech. rep., 1976 (cit. on pp. 53, 54).

[94] T. Rogers and M. Strikman, *Multiple hard partonic collisions with correlations in proton-proton scattering*, Physical Review D **81** (2010) 016013 (cit. on pp. 53, 54).

[95] F. Bunkin and M. Fedorov, *Bremsstrahlung in a strong radiation field*, Sov. Phys. JETP **22** (1966) 844 (cit. on p. 53).

[96] T. Alexopoulos et al., *The role of double parton collisions in soft hadron interactions*, Physics Letters B **435** (1998) 453 (cit. on p. 53).

[97] R. Stock, *The parton to hadron phase transition observed in Pb+ Pb collisions at 158 GeV per nucleon*, Physics Letters B **456** (1999) 277 (cit. on p. 53).

[98] S. Dittmaier, *A general approach to photon radiation off fermions*, Nuclear Physics B **565** (2000) 69 (cit. on p. 53).

[99] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, Physical Review D **93** (2016) 033006 (cit. on p. 54).

[100] R. D. Ball et al., *Parton distributions for the LHC Run II*, Journal of High Energy Physics **2015** (2015) 40 (cit. on p. 54).

[101] J. D. Bjorken, *Asymptotic sum rules at infinite momentum*, Physical Review **179** (1969) 1547 (cit. on p. 54).

[102] M. Mangano, *The color structure of gluon emission*, Nuclear Physics B **309** (1988) 461 (cit. on p. 55).

[103] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross sections in deep-inelastic scattering*, arXiv preprint hep-ph/9907280 (1999) (cit. on p. 55).

[104] *File:quark confinement.svg*, URL: https://commons.wikimedia.org/wiki/File:Quark_confinement.svg (cit. on p. 56).

[105] A. H. Hoang, *The Top Mass: Interpretation and Theoretical Uncertainties*, 2014, arXiv: 1412.3649 [hep-ph] (cit. on p. 56).

[106] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani-Seymour dipole factorisation*, Journal of High Energy Physics **2008** (2008) 038 (cit. on p. 56).

[107] B. Andersson, G. Gustafson and L. Lönnblad, *Gluon splitting in the colour dipole cascades*, Nuclear Physics B **339** (1990) 393 (cit. on p. 56).

[108] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni 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 (cit. on p. 57).

[109] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on p. 57).

[110] S. Frixione, E. Laenen, P. Motylinski and B. R. Webber, *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*, JHEP **04** (2007) 081, arXiv: hep-ph/0702198 (cit. on p. 57).

[111] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015, arXiv: 1212.3460 [hep-ph] (cit. on p. 57).

[112] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on pp. 57, 63, 97).

[113] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on p. 57).

[114] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on p. 57).

[115] de Florian, D. and others, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, (2016), arXiv: 1610.07922 [hep-ph] (cit. on p. 57).

[116] ATLAS Collaboration, *Modelling of the $t\bar{t}H$ and $t\bar{t}V$ $(V = W, Z)$ processes for $\sqrt{s} = 13$ TeV ATLAS analyses*, ATL-PHYS-PUB-2016-005, 2016, URL: https://cds.cern.ch/record/2120826 (cit. on pp. 57, 96, 97).

[117] M. Bähr et al., *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639, arXiv: 0803.0883 [hep-ph] (cit. on pp. 57, 96, 97).

[118] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv: 1512.01178 [hep-ph] (cit. on pp. 57, 96, 97).

[119] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on p. 57).

[120] E. Bothmann et al., *Event Generation with Sherpa 2.2*, (2019), arXiv: 1905.09127 [hep-ph] (cit. on p. 57).

[121] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on p. 57).

[122] *vertex fitting and detector alignment*, URL: https://idpasc.lip.pt/uploads/talk/file/71/Tracking-Vertexing-Slides.pdf (cit. on p. 58).

[123] C. Kourkoumelis and S. Vourakis, *HYPATIA—an online tool for ATLAS event visualization*, Physics Education **49** (2014) 21 (cit. on p. 59).

[124] 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. 59–61, 96, 97).

[125] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton–proton collision data*, ATLAS-CONF-2014-032, 2014, URL: https://cds.cern.ch/record/1706245 (cit. on p. 59).

[126] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data*, Eur. Phys. J. C **77** (2017) 195, arXiv: 1612.01456 [hep-ex] (cit. on p. 59).

[127] 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 p. 60).

[128] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 (cit. on pp. 61–63).

[129] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on p. 61).

[130] 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 p. 61).

[131] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 61).

[132] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on p. 61).

[133] G. P. Salam, *Towards jetography*, The European Physical Journal C **67** (2010) 637 (cit. on p. 61).

[134] S. Ellis, J. Huston, K. Hatakeyama, P. Loch and M. Tönnesmann, *Jets in hadron–hadron collisions*, Progress in Particle and Nuclear Physics **60** (2008) 484 (cit. on p. 61).

[135] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, (2020), arXiv: 2007.02645 [hep-ex] (cit. on pp. 62, 96, 97).

[136] 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. 62, 63).

[137] *File:B-tagging diagram.svg*, URL: https://commons.wikimedia.org/wiki/File:B-tagging_diagram.svg (cit. on p. 64).

[138] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv: 1907.05120 [hep-ex] (cit. on p. 64).

[139] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, (2020), arXiv: 2004.13447 [hep-ex] (cit. on p. 65).

[140] M. O. Evans, *LHC open data for the world to see*, PoS **LHCP2020** (2020) 199 (cit. on p. 66).

[141] T. McCauley, *HEP data for everyone: CERN open data and the ATLAS and CMS experiments*, PoS **ICHEP2016** (2017) 335 (cit. on p. 66).

[142] F. Socher, *ATLAS and CMS Data Release & Tools*, PoS **LHCP2016** (2016) 112 (cit. on p. 66).

[143] M. O. Evans, 'Enabling Open Science with the ATLAS Open Data project at CERN', Presented 01 Jun 2018, 2018, URL: https://cds.cern.ch/record/2630961 (cit. on p. 66).

[144] M. O. Evans, *ATLAS Open Data project: HEP for everyone*, PoS **ICHEP2018** (2019) 275 (cit. on p. 66).

[145] C. Doglioni, *The ATLAS Open Data project*, tech. rep. ATL-OREACH-PROC-2018-001, CERN, 2018, URL: https://cds.cern.ch/record/2637284 (cit. on p. 66).

[146] Y.-F. Lo, *ATLAS Open Data: software for visualization and physics analysis*, (2019), URL: https://cds.cern.ch/record/2655357 (cit. on p. 66).

[147] A. Sanchez Pineda and S. Mehlhase, *ATLAS Outreach: on the dissemination of High Energy Physics and Computer Sciences*, PoS **LHCP2019** (2019) 008 (cit. on p. 66).

[148] ATLAS Collaboration, *ATLAS Open Data*, 2021, URL: http://opendata.atlas.cern/ (visited on 23/02/2021) (cit. on pp. 66, 141, 142).

[149] CERN, *CERN Open Data*, 2021, URL: http://opendata.cern.ch/ (visited on 23/02/2021) (cit. on pp. 66, 142).

[150] L. Hagesjo, 'Making ATLAS Data from CERN Accessible to the General Public: The Development and Evaluation of a Learning Resource in Experimental Particle Physics.', Presented 12 Jun 2017, 2017, URL: https://cds.cern.ch/record/2281684 (cit. on p. 66).

[151] 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 p. 66).

[152] R. C. Camacho Toro, *Outreaching particle physics to Latin America: CEVALE2VE and the use of ATLAS open data*, PoS **EPS-HEP2017** (2017) 554 (cit. on p. 66).

[153] A. Adiguzel et al., *Evaluating Analysis Description Language Concept as a First Introduction to Analysis in Particle Physics*, 2020, arXiv: 2008.12034 [hep-ph] (cit. on p. 66).

[154] A. Sanchez Pineda, *The Cevale2ve case*, PoS **ICHEP2016** (2017) 322 (cit. on p. 66).

[155] C. Kourkoumelis, S. Vourakis and S. Sotiriou, *HEP interactive activities in high schools in the framework of the CREATIONS project*, PoS **EPS-HEP2017** (2017) 559 (cit. on p. 66).

[156] D. Fassouliotis, C. Kourkoumelis and S. Vourakis, 'HEP in Greek Classes', *EPJ Web of Conferences*, vol. 137, EDP Sciences, 2017 01012 (cit. on p. 66).

[157] A. Adiguzel et al., *CutLang as an analysis description language for introducing students to analyses in particle physics*, European Journal of Physics (2021) (cit. on p. 66).

[158] G. Ünel and S. Sekmen, *CutLang: A particle physics analysis description language and runtime interpreter*, Computer Physics Communications **233** (2018) 215 (cit. on p. 66).

[159] F. Ould-Saada, 'ATLAS Open Data–Development of a simple-but-real HEP data analysis framework', *EPJ Web of Conferences*, vol. 245, EDP Sciences, 2020 08023 (cit. on p. 66).

[160] J. J. C. Mckeown, 'A Contribution to ATLAS Open Data Collaboration at CERN', Presented 15 May 2019, 2019, URL: https://cds.cern.ch/record/2704159 (cit. on p. 66).

[161] S. Biryukov, 'A Contribution to the ATLAS Open Data Collaboration at CERN', Presented 2020, 2020, URL: https://cds.cern.ch/record/2724158 (cit. on p. 66).

[162] L. Thulasidharan, *J/ψ, Y and Z studies with CERN Open Data*, DESY Summer Student Report (2018) (cit. on p. 66).

[163] M. R. Di Domenico Franco, 'Reconstruction of the invariant masses of bosons of the Standard Model using public data from ATLAS Open Data', Presented 04 Oct 2017, 2017, URL: https://cds.cern.ch/record/2293251 (cit. on p. 66).

[164] A. Prieto Tirado, 'W' and Z' bosons search at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC', Presented 11 Jun 2020, 2020, URL: https://cds.cern.ch/record/2743513 (cit. on p. 66).

[165] I. Garcia, 'Perspectivas y Evaluación de producción de Materia Oscura en asociación con un quark liviano, un quark pesado (quark b) o un boson electro débil en colisiones de particular a energías de colisión de $\sqrt(s) = 8$ TeV', Presented 14 Jul 2017, 2017, URL: https://cds.cern.ch/record/2291838 (cit. on p. 66).

[166] ATLAS Collaboration, *Review of the* 13 *TeV ATLAS Open Data release*, ATL-OREACH-PUB-2020-001, 2020, URL: https://cds.cern.ch/record/2707171 (cit. on pp. 67, 141, 142).

[167] *Top 6 quotes by Amber Naslund: A-Z quotes*, URL: https://www.azquotes.com/author/64233-Amber_Naslund (cit. on p. 69).

[168] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section and lepton differential distributions in $e\mu$ dilepton events from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 528, arXiv: 1910.08819 [hep-ex] (cit. on pp. 74, 75).

[169] ATLAS Collaboration, *Measurements of the production cross section of a Z boson in association with jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 361, arXiv: 1702.05725 [hep-ex] (cit. on pp. 74, 75).

[170] T. Kittelmann, V. Tsulaia, J. Boudreau and E. Moyse, 'The Virtual Point 1 event display for the ATLAS experiment', *Journal of Physics: Conference Series*, vol. 219, 3, IOP Publishing, 2010 032012 (cit. on p. 76).

[171] X. Wu et al., *Top 10 algorithms in data mining*, Knowledge and information systems **14** (2008) 1 (cit. on p. 84).

[172] T. Schaapherder, 'Photon Conversion Classification by Boosting Decision Trees', B.S. thesis, 2018 (cit. on p. 84).

[173] Y. Shin, *Application of Boosting Regression Trees to Preliminary Cost Estimation in Building Construction Projects*, Computational intelligence and neuroscience **2015** (2015) 149702 (cit. on p. 85).

[174] A. Dertat, *Applied deep learning - part 1: Artificial Neural Networks*, 2017, URL: https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6 (cit. on p. 86).

[175] *Tensorflow*, URL: https://www.tensorflow.org/ (cit. on pp. 87, 88, 107, 115).

[176] K. Team, *Simple. flexible. powerful.*, URL: https://keras.io/ (cit. on pp. 87, 88, 107, 115).

[177] G. C. Fox and S. Wolfram, *Event shapes in e+ e- annihilation*, Nuclear Physics B **149** (1979) 413 (cit. on p. 90).

[178] 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. 96).

[179] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 580, arXiv: 1705.02211 [hep-ex] (cit. on p. 96), Erratum: Eur. Phys. J. C **77** (2017) 712.

[180] G. Aad et al., *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s}$ = 13 TeV*, The European Physical Journal C **81** (2021) 1 (cit. on pp. 96, 97).

[181] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements+ parton showers. The NLO case*, Journal of High Energy Physics **2013** (2013) 1 (cit. on pp. 96, 97).

[182] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on pp. 96, 97).

[183] N. Lavesson and L. Lönnblad, *Extending CKKW-merging to one-loop matrix elements*, Journal of High Energy Physics **2008** (2008) 070 (cit. on pp. 96, 98).

[184] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, Journal of High Energy Physics **2009** (2009) 053 (cit. on pp. 96, 98).

[185] K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016, URL: https://cds.cern.ch/record/1456844 (cit. on p. 98).

[186] J. L. Collins, *The simple path to wealth: Your road map to financial independence and a rich, Free Life*, JL Collins, 2016 (cit. on p. 101).

[187] *Learn: Machine learning in python - scikit-learn 0.16.1 documentation*, URL: https://scikit-learn.org/ (cit. on p. 102).

[188] scikit-learn, *sklearn.ensemble.GradientBoostingClassifier*, 2020, URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html (visited on 11/01/2021) (cit. on p. 102).

[189] O. Bessidskaia Bylund et al., *Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ production cross sections in multilepton final states using 36.1 fb^{-1} of pp collisions at 13 TeV at the LHC*, tech. rep., CERN, 2016, URL: https://cds.cern.ch/record/2235582 (cit. on p. 139).

[190] *Respect your parents. they passed school without google.: Respect your parents, funny quotes, parenting quotes*, 2014, URL: https://www.pinterest.com/pin/416934877975514282/ (cit. on p. 141).

[191] L. Serkin, 'The release of the 13 TeV ATLAS Open Data: using open education resources effectively', *EPJ Web of Conferences*, vol. 245, EDP Sciences, 2020 08026 (cit. on p. 141).

[192] M. O. Evans, *ATLAS Open Data: Data visualisation and educational physics analysis to re-discover the Higgs boson*, PoS **LHCP2020** (2020) 228 (cit. on pp. 142, 155).

[193] ATLAS Collaboration, *Get started with data visualisation*, 2020, URL: http://opendata.atlas.cern/release/2020/documentation/visualization/index.html (visited on 23/02/2021) (cit. on p. 142).

[194] Atlas-Outreach-Data-Tools, *Notebooks-collection-opendata/infofile.py at master · ATLAS-outreach-data-tools/notebooks-collection-opendata*, 2021, URL: https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata/blob/master/13-TeV-examples/uproot_python/infofile.py (cit. on p. 144).

[195] T. Kluyver et al., *Jupyter Notebooks-a publishing format for reproducible computational workflows.* Vol. 2016, 2016 (cit. on p. 155).

[196] Y. Wang, *A Jupyter-based Interface Integrated with ROOT for ATLAS Open Data Analysis*, (2019), URL: https://cds.cern.ch/record/2687389 (cit. on p. 155).

[197] A. Sanchez Pineda, *Integration of ROOT NoteBook as an ATLAS analysis web-based tool in outreach and public data release projects*, PoS **ICHEP2016** (2017) 1054 (cit. on p. 155).

[198] ATLAS Collaboration, *ttZ_ML_from_csv.ipynb*, 2021, URL: https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata/blob/master/13-TeV-examples/uproot_python/ttZ_ML_from_csv.ipynb (visited on 23/02/2021) (cit. on p. 155).

[199] ATLAS Collaboration, *ttZ_ML_from_root.ipynb*, 2021, URL: https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata/blob/master/13-TeV-examples/uproot_python/ttZ%5C_ML%5C_from%5C_root.ipynb (visited on 23/02/2021) (cit. on p. 158).

[200] *A conclusion is the place where you got tired of thinking. at quotetab*, URL: https://www.quotetab.com/quote/by-source-unknown/a-conclusion-is-the-place-where-you-got-tired-of-thinking?source=thoughts-and-thinking (cit. on p. 165).