Combined Z→2L and H→4L distributions using ATLAS and CMS (Open) Data

DESY Summer Student Programme, 2021

Aritra Bal
Indian Institute of Technology Kharagpur

Supervisors

Achim Geiser
Yewon Yang

September 8, 2021

Abstract

The aim of the project was to study the H→4L and Z→2L distributions from ATLAS and CMS (Open) Data. The analysis was already carried out for the 2011 Run A and 2012 Run B+C Open Data [both] and ATLAS 8 TeV Open Data [Z→2L only]. A common analysis code for all datasets has to be developed, along with a framework for converting ATLAS 13 TeV Open Data to the CMS nanoAODplus format usable by the aforementioned code. With the inclusion of new datasets like the ATLAS 13 TeV **Open** Data and CMS Run 2 Data, the statistics for the Higgs decay have increased by a large factor.

Contents

1 Introduction

The **CMS nanoAOD** format [1] was developed as a replacement for the AOD (Analysis Object Data) from Run 1, and the miniAOD formats due to time and resource constraints, with nanoAOD requiring only about 1 kilobyte storage per event. The nanoAOD format can be analysed simply in ROOT, and it also stores more information. Analysis using the older format was already carried out in the 2019 Summer Project. The first step was to validate the new nanoAODplus nTuples by comparison to the results [2] from the 2019 project - for both data and Monte Carlo simulations. Having successfully carried out the checks, a common framework was developed for the conversion of ATLAS 13 TeV Data to CMS nanoAODplus format, to increase the statistics. Further increase was obtained by analyzing the data from CMS 2011 Run A (Open) and Run 2 CMS Data from 2016. A preliminary investigation was also carried out for the 4μ mass spectrum from 2016-18 Run 2 Data. The generated plots were then compared against the released figures by the CMS Collaboration.
A complete documentation of the CMS nanoAODplus variables can be found at [3], and that of ATLAS 13 TeV variables at [4].

1.1 H $\rightarrow$ 4L decay channel

A total of five channels ($\gamma\gamma$, ZZ, WW, $\tau\tau$ and bb) were analyzed for the discovery of the Higgs boson. Out of these, the H$\rightarrow$ZZ$\rightarrow$4L (where L= e, μ) channel is known as the golden channel because of the large signal to background ratio, leading to a clean signal at around 125 GeV. The observation of the Higgs peak can further be validated by the Z$\rightarrow$4L around 90 GeV.

2 Event Selection Criteria

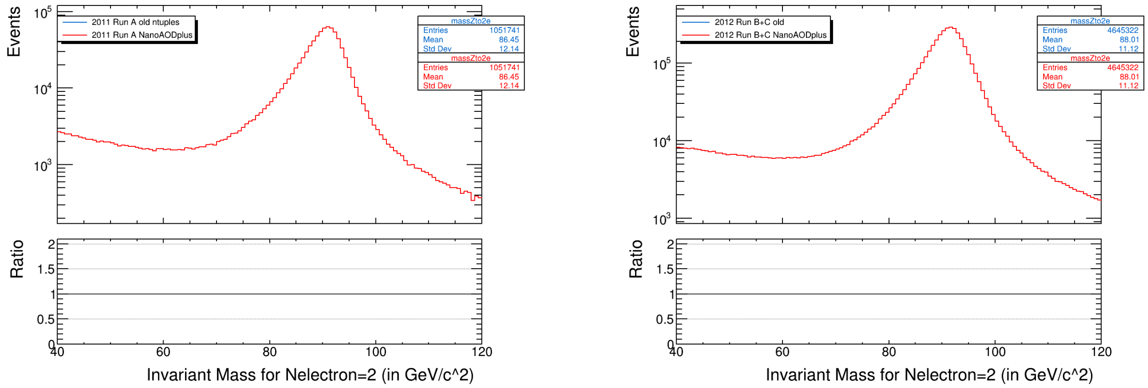
2.1 Selection and Kinematic Cuts

To discard background events as much as possible, we first apply the quality cuts, namely that the muons have to be global, and both muons and electrons must be particle flow candidates. A further set of kinematic cuts on transverse momentum (p_T), pseudorapidity (η) and impact parameters (IP) [5].

- $p_T > 7$ GeV for electrons and $p_T > 5$ GeV for muons.
- $|\eta| < 2.5$ for electrons and $|\eta| < 2.4$ for electrons.
- Transverse IP $|d_{XY}| < 0.5$ for both electrons and muons.
- Longitudinal IP $|d_Z| < 1.0$ for both electrons and muons.
- Relative Isolation: scalar sum of the transverse momenta of the particles within the distance ΔR of a muon, defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, and normalized to p_T is required to be lesser than 0.4
- Impact parameter significance, which is the ratio of the 3D Impact Parameter at point of closest approach (to lepton), to its uncertainty, must satisfy $|SIP_{3D}| < 1.0$ for both electrons and muons.

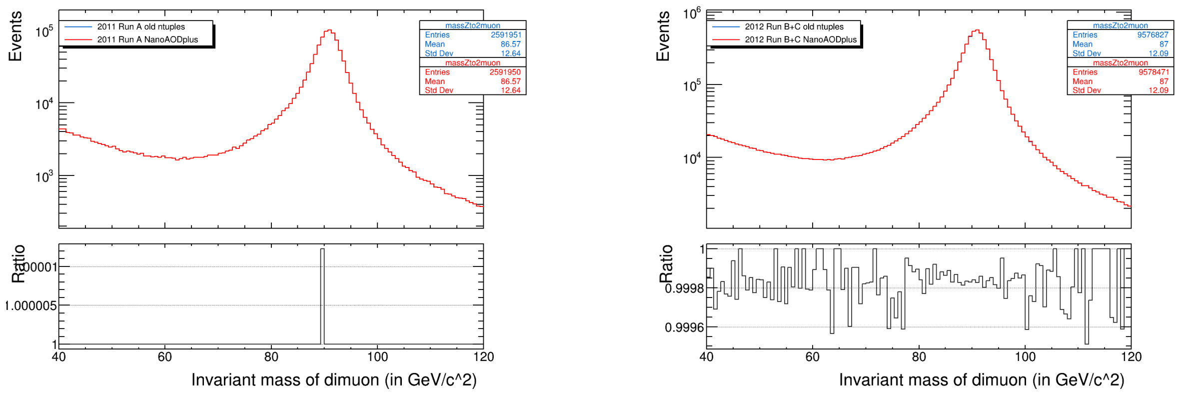
The leptons (muon/electron) that survive all the cuts above are henceforth called *good* electrons or muons.

H



(a) 2011 Run A (b) 2012 Run B/C

Figure 1: Z→ $2e$ peak validation: CMS Run 1 Open Data



(a) 2011 Run A (b) 2012 Run B/C

Figure 2: Z→ 2μ peak validation: CMS Run 1 Open Data

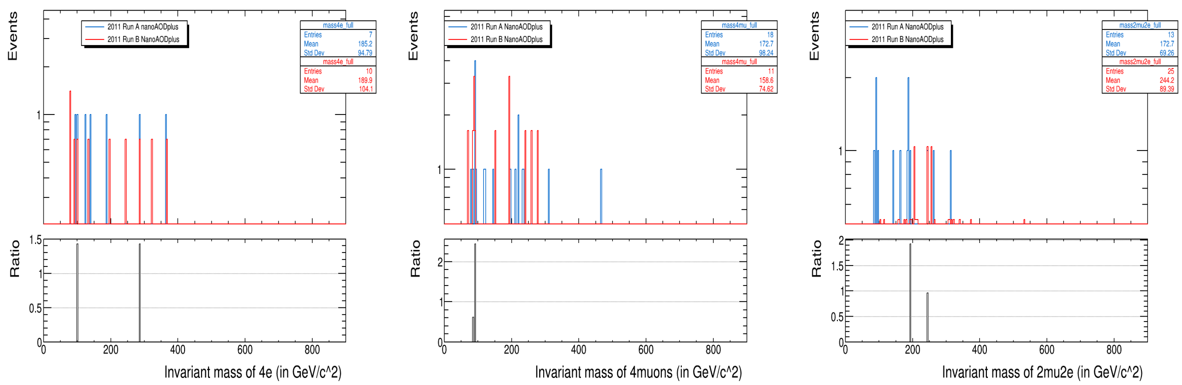
2.2 Higgs Candidate Event Selection

Z pairs are formed by selecting 2 pairs of same flavour and opposite charge leptons - 4μ or $4e$ (3 possibilities each) and $2\mu 2e$ (only 1 possibility). The pair with invariant mass closer to the Z boson mass $M_Z = 90\ GeV$ must satisfy $20 < M_A < 120$, whereas the other must satisfy $12 < M_B < 120$. The leptons that form the pairs must satisfy additional p_T cuts: $p_T > 20\ GeV$ for the higher transverse momentum lepton and $p_T > 10\ GeV$ for the lower transverse momentum lepton.

3 Validation of Datasets

3.1 new nTuples using $Z \rightarrow 2L$ peaks

Comparisons were carried out against the older plots from the 2019 project, which had been compared to the Higgs Open Data release. They were found to agree within an acceptable margin (Figures 1 and 2).



(a) 4e mass distribution (b) 4μ mass distribution (c) $2\mu 2e$ mass distribution

Figure 3: Comparison of 4L mass distributions for 2011 Run A and Run B Open Data

3.2 CMS 2011 Run B Open Data

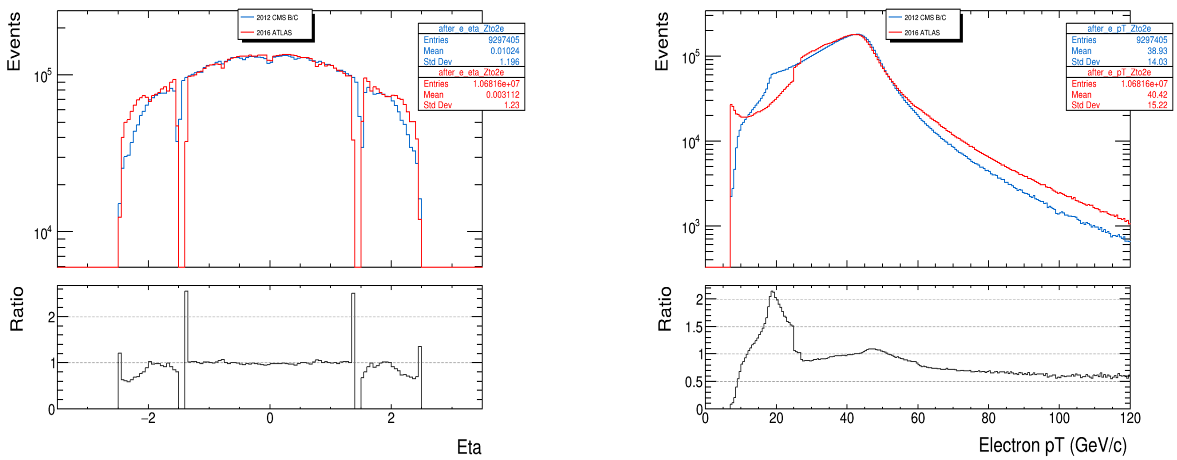
As a benchmark, we compare these plots against the 2011 Run A plots, subject to normalisation by the histogram integral. By adding the statistics from Run B, the net luminosity is increased from 2.3 fb^{-1} to 5.1 fb^{-1} for CMS 2011. The figures for the 4 lepton mass distribution are shown here in Figure 3.

3.3 ATLAS 13 TeV Open Data

With a luminosity of 10.2 fb^{-1}, the ATLAS 13 TeV data adds greatly to the statistics for the Higgs mass spectrum. A framework [similar to the one for ATLAS 8 TeV in [6]] was developed for the conversion of the data into nanoAODplus format which can be analysed by the same code used on the CMS datasets. A table for the conversion is given in the appendix.
ATLAS Data has an additional set of pre-selection cuts [7], namely that the muons are already global muons, and the leptons are all particle flow candidates. Accordingly, these branches have been initialised to *true*. The value of $|\eta|$ for the electron determines whether it is in the barrel or endcap calorimeter as, with $|\eta| < 1.37$ for barrel and $|\eta| > 1.53$ for endcap calorimeter. A complete table for the variable conversion [both 8 TeV and 13 TeV] is present in the appendix.
The following pre-selection cuts are already present in ATLAS and have been validated by control plots comparing against data from CMS 2012 Run B/C (L=11.6 fb^{-1}).

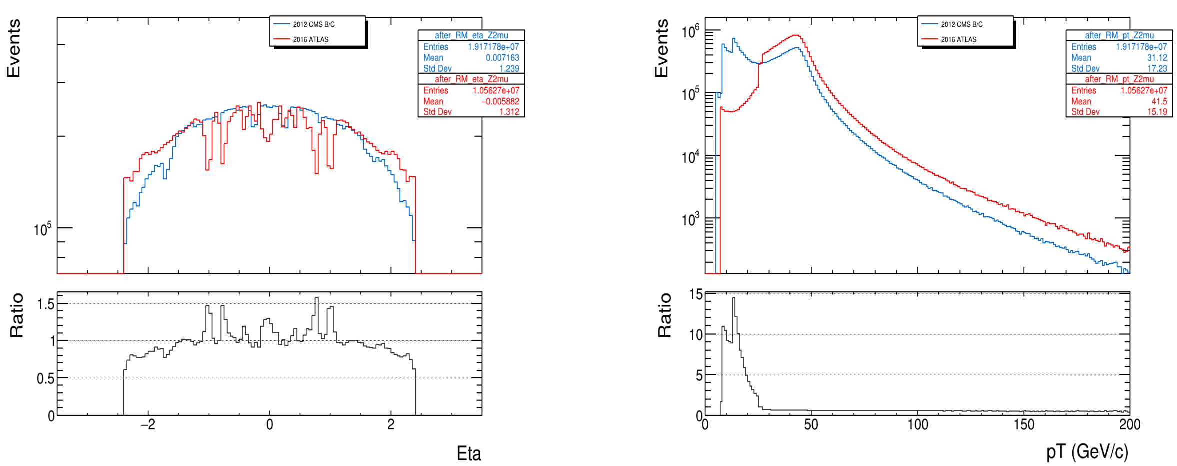
- $p_T > 7.0$ GeV for electrons and muons.
- $|\eta| < 2.47$ for electrons.
- $|\eta| < 2.5$ for muons.
- lep_ptcone30/lep_pt < 0.15 for electrons and muons.
- lep_etcone20/lep_pt < 0.15 for electrons and muons.

where:
lep_ptcone30: Scalar sum of track p_T in a cone of $R = 0.3$ around the lepton.
lep_etcone20: Scalar sum of track E_T in a cone of $R = 0.2$ around the lepton.



(a) η (b) p_T

Figure 4: electron variables after $Z \to 2e$ cuts: ATLAS 13 TeV Open Data vs CMS 2012 Run B/C Open Data



(a) η (b) p_T

Figure 5: muon variables after $Z \to 2\mu$ cuts: ATLAS 13 TeV Open Data vs CMS 2012 Run B/C Open Data

4 $H \to 4L$ mass spectrum

4.1 Open Data

The luminosities for each data set are as follows:

- 2011 Run A and B: 5.1 fb^{-1}
- 2012 Run B and C: 11.6 fb^{-1}
- ATLAS 8 TeV: 1.0 fb^{-1}
- ATLAS 13 TeV: 10.2 fb^{-1}

All these datasets are publicly available. The analysis in the $H \to 4L$ decay channel has been completed for all the above datasets, with the MC predictions and background taken from CMS Run 1 MC samples, scaled by the luminosities mentioned above. The final result is shown in 6.

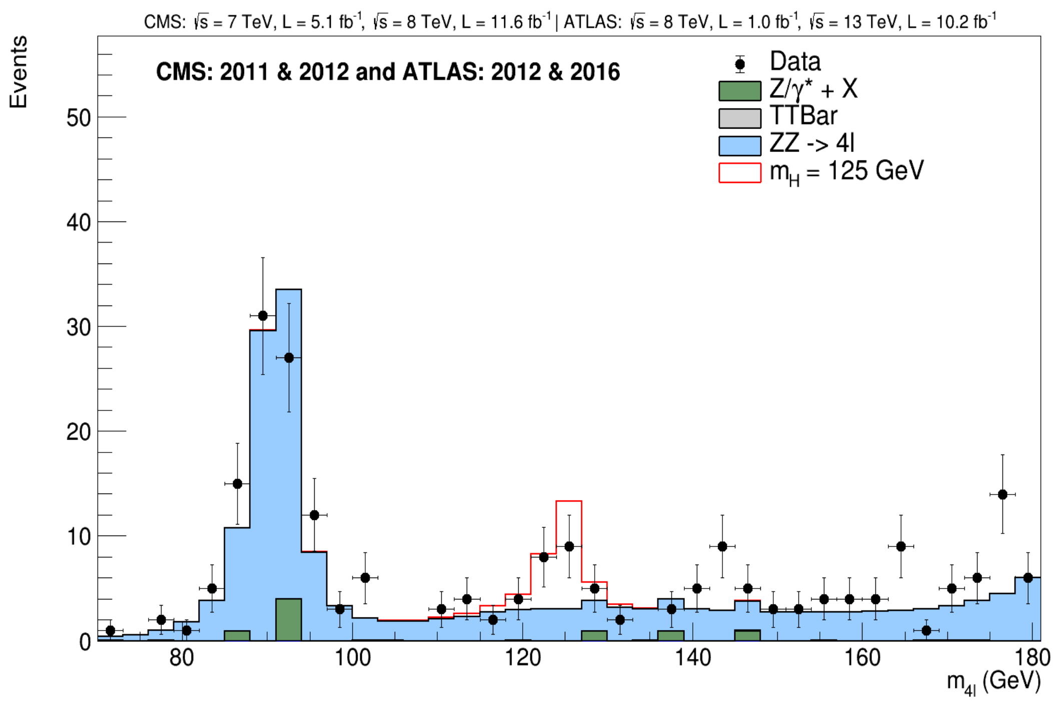


Figure 6: 4L mass spectrum: **Open Data** from ATLAS and CMS

4.2 CMS Run 2 Data

With the addition of CMS Run 2 Data (not yet available publicly), statistics can be increased greatly. We have included Run 2 Data from 2016, with $\sqrt{s} = 13\ TeV$ and $L = 38\ fb^{-1}$. Scaling the MC by an appropriate factor, the final mass spectrum was obtained as shown in 7.

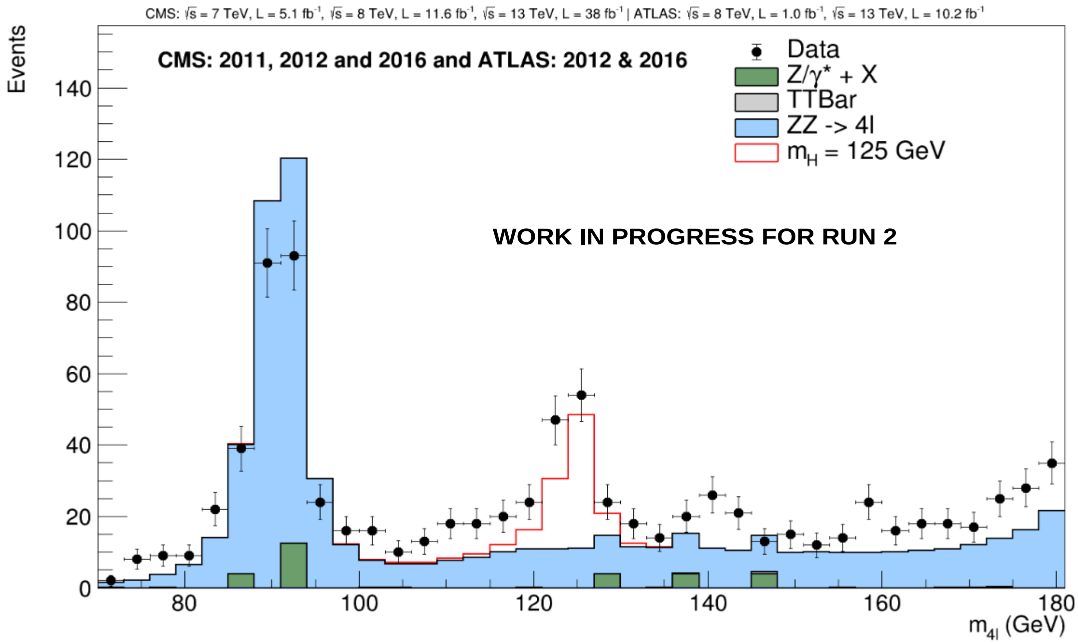


Figure 7: 4L Mass spectrum: CMS and ATLAS Open Data, and CMS Run 2 2016 Data

The analysis presented here uses the cuts and histogram binning from CMS Run 1.

For Run 2 Data from 2016-18, with total $L = 137\ fb^{-1}$, the analysis has been carried out for the $H \to ZZ \to 4\mu$ decay, with comparison against Figure 5 from [8] shown in 8.

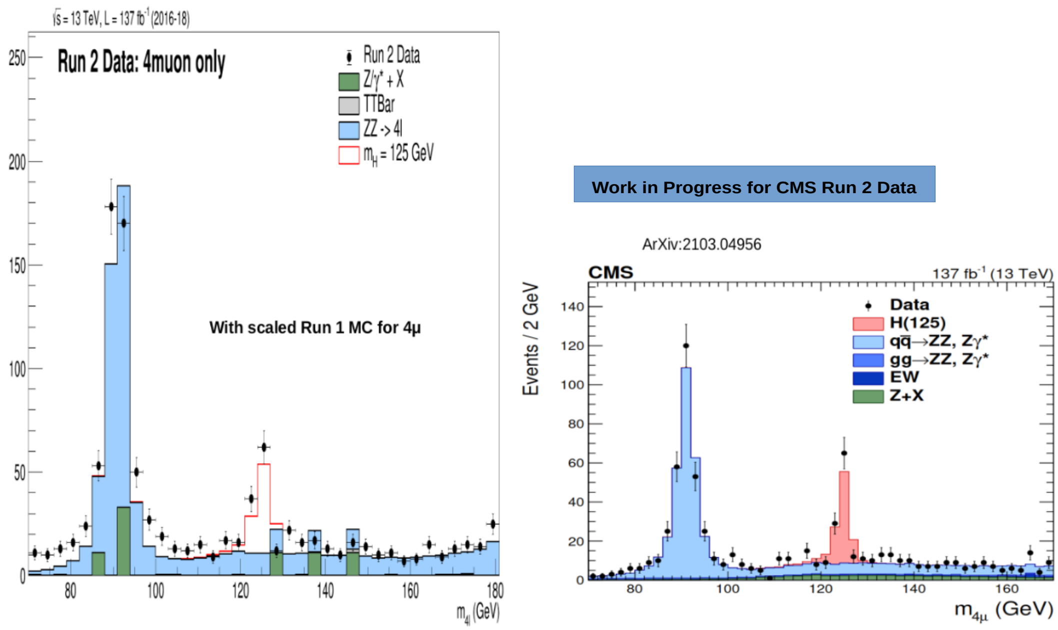


Figure 8: $H \to 4\mu$ decay mass spectrum: CMS Run 2 Data

5 Conclusions

The newly generated nanoAODplus nTuples have been validated to produce results identical to the old samples from the 2019 Summer Project. For Monte Carlo, new nanoAODplus samples have now been included in the analysis, including for the background Drell-Yan and $t\bar{t}$. In each case, they have been validated against the old samples.

Statistics have been greatly increased for the $Z \to 2L$ and $H \to 4L$ mass spectra with the inclusion of the openly available CMS Run 1 and ATLAS 8 TeV and 13 TeV data. A preliminary investigation has also been carried out using the CMS Run 2 data, for further increase in statistics.

Acknowledgements

I would like to thank my supervisors Achim and Yewon for their guidance and help throughout the project, and for helping me to learn a lot in the process. Special mention also to my colleagues Leonardo, Murillo and Raphael during these eight weeks, for being a great team to work with.

References

[1] CMS NanoAOD Work Book
https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAOD

[2] 2019 Summer Project Report - Paula Martinez.
https://www.desy.de/f/students/2019/reports/Paula.Martinez.pdf

[3] CMS NanoAODplus Documentation.
https://www.desy.de/~geiser/nano/nanoAODplusv0.8.html

[4] Full list of Branches and Variables: ATLAS 13 TeV Open Data dataset.
http://opendata.atlas.cern/release/2020/documentation/datasets/dataset13.html

[5] Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC
https://arxiv.org/pdf/1207.7235.pdf

[6] Summer Student Project B12 Report - Leonardo Olivi, 2021.

[7] Review of the ATLAS 13 TeV Open Data dataset.
https://cds.cern.ch/record/2707171/files/ANA-OTRC-2019-01-PUB-updated.pdf

[8] Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13\ TeV$.
https://arxiv.org/pdf/2103.04956.pdf

Appendix

Next page.

ATLAS 2012+2016 CONVERSION TABLE

ATLAS	2012 Data Type	2016 Data Type	CMS	Data type
runNumber	Int_t	Int_t	run	UInt_t
eventNumber	Int_t	Int_t	event	ULong64_t
mcWeight		Int_t	Generator_weight	Int_t
lep_n	UInt_t	UInt_t	nMuon, nElectron	UInt_t
lep_charge	Float_t[]	Vector<Float_t>	Muon_charge, Electron_charge	Int_t[]
lep_pt/1000.0	Float_t[]	Vector<Float_t>	Muon_pt, Electron_pt	Float_t[]
lep_eta	Float_t[]	Vector<Float_t>	Muon_eta, Electron_eta	Float_t[]
lep_phi	Float_t[]	Vector<Float_t>	Muon_phi, Electron_phi	Float_t[]
lep_E	Float_t[]	Vector<Float_t>	/	/
lep_z0/10.0	Float_t[]	Vector<Float_t>	Muon_dz, Electron_dz	Float_t[]
lep_d0/10.0	Float_t[]	Vector<Float_t>	Muon_dxy, Electron_dxy	Float_t[]
lep_sd0/10.0	Float_t[]	Vector<Float_t>	Muon_dxyErr, Electron_dxyErr	Float_t[]
lep_charge	Float_t[]	Vector<Float_t>	Muon_charge, Electron_charge	Float_t[]
lep_type	Int_t[]	Vector<Int_t>	$11 = e, \quad 13 = \mu, \quad 15 = \tau$	-
lep_ptcone30/lep_pt	Float_t[]	Vector<Float_t>	Muon_pfRelIso03_chg, Electron_pfRelIso03_chg	Float_t[]
(lep_etcone20/lep_pt) ×9.0/4.0	Float_t[]	Vector<Float_t>	Muon_pfRelIso03_all	Float_t[]
(lep_etcone20/lep_pt) ×9.0/4.0	Float_t[]	Vector<Float_t>	Electron_pfRelIso03_all	Float_t[]
(lep_etcone20/lep_pt) ×16.0/4.0	Float_t[]	Vector<Float_t>	Muon_pfRelIso04_all	Float_t[]
lep_isTightID	/	Vector<Bool_t>	Muon_tightID	Bool_t[]
$\sqrt{\text{lep\_z0}^2 + \text{lep\_d0}^2/10.0}$	Float_t[]	Vector<Float_t>	$\text{Muon}_ip3d, Electron_ip3d$	Float_t[]
lep_n	UInt_t	/	nTau	UInt_t
lep_charge	Float_t[]	/	Tau_charge	Float_t[]
lep_pt/1000.0	Float_t[]	/	Tau_pt	Float_t[]
lep_eta	Float_t[]	/	Tau_eta	Float_t[]
lep_phi	Float_t[]	/	Tau_phi	Float_t[]
lep_z0/10.0	Float_t[]	Vector<Float_t>	Tau_dz	Float_t[]
lep_d0/10.0	Float_t[]	/	Tau_dxy	Float_t[]
tau_n	/	Vector<Float_t>	nTau	Float_t
tau_pt/1000.0	/	Vector<Float_t>	Tau_pt	Float_t
tau_eta	/	Vector<Float_t>	Tau_eta	Float_t
tau_phi	/	Vector<Float_t>	Tau_phi	Float_t
tau_charge	/	Vector<Int_t>	Tau_charge	Int_t
met_et/1000.0	Float_t[]	Vector<Float_t>	Met_sumEt	Float_t[]
met_phi	Float_t[]	Vector<Float_t>	Met_phi	Float_t[]
jet_n	/	Int_t	nJet	UInt_t[]
alljet_n	Int_t	/	nJet	UInt_t[]
jet_pt	Float_t[]	Vector<Float_t>	Jet_pt	Float_t[]
jet_eta	Float_t[]	Vector<Float_t>	Jet_eta	Float_t[]
jet_phi	Float_t[]	Vector<Float_t>	Jet_phi	Float_t[]
jet_m	Float_t[]	/	Jet_mass	Float_t[]
scaleFactor_PILEUP	Float_t[]	Float_t[]	sf_pileup	Float_t[]
scaleFactor_ELE	Float_t[]	Float_t[]	sf_ele	Float_t[]
scaleFactor_MUON	Float_t[]	Float_t[]>	sf_muon	Float_t[]
scaleFactor_BTAG	Float_t[]	Float_t[]	sf_btag	Float_t[]
scaleFactor_TRIGGER	Float_t[]	Float_t[]	sf_trigger	Float_t[]
scaleFactor_JVFSF	Float_t[]	Float_t[]	sf_jvfsf	Float_t[]
scaleFactor_ZVERTEX	Float_t[]	Float_t[]	sf_zvertex	Float_t[]
trigE	Bool_t[]	Bool_t[]	Trig_goodMuTrigger	Bool_t[]
trigM	Bool_t[]	Bool_t[]	Trig_goodETrigger	Bool_t[]
passGRL	Bool_t[]	/	GoodLumisection	Bool_t[]
hasGoodVertex		/	PVtx_isGood = *true*	
hasGoodVertex	Bool_t[]	/	PVtx_isMain = *true*	Bool_t[]
hasGoodVertex		/	PVtx_isValid = *true*	
-	-	-	Muon_isPFcand = *true*	Bool_t[]
-	-	-	Muon_isGlobal = *true*	Bool_t[]
-	-	-	Muon_softID = *true*	Bool_t[]
-	-	-	Muon_isTracker = *true*	Bool_t[]
-	-	-	Electron_isPFcand = *true*	Bool_t[]
-	-	-	$\text{Muon}_{softID} = true$	Bool_t[]
if $\lvert\text{lep\_eta}\rvert<1.37$	-	-	Electron_isEB = *true*	Bool_t[]
if $\lvert\text{lep\_eta}\rvert>1.52$	-	-	Electron_isEE = *true*	Bool_t[]
-	-	-	Muon_sip3d = 0.0	Float_t[]
-	-	-	Electron_sip3d = 0.0	Float_t[]
-	-	-	Electron_lostHits = 0	UChar_t[]
-	-	-	Muon_mass = 0.105658 GeV/c^2	Float_t[]
-	-	-	Electron_mass = 0.0005109989 GeV/c^2	Float_t[]
-	-	-	Tau_mass = 0.177686 GeV/c^2	Float_t[]

lep_d0 is actually lep_trackd0pvunbiased.
lep_sd0 is actually lep_tracksigd0pvunbiased.