Validation from NanoAODRun1

Matheus Reis, Universidade do Estado do Rio de Janeiro, Brasil

September 7, 2022

Abstract

The aim of this project was to validate the data generated in the new format called NanoAODRun1. This format is similar to NanoAOD from Run2 but for 2010, 2011 and 2012 data (Run1), being built so that it can be used in the most generic analyses. As it is in the production/validation phase, only a few samples are available, a comparison is made using these available samples with results found in other analyses. A comparison is also made between data and simulation using 2012 and 2016 data from ATLAS.

Contents

1 Introduction

NanoAODRun1 is a data format similar to the NanoAOD of run 2, but created for run 1 [1]. It can be read directly by ROOT and contains information on collision events that are necessary for the most generic analysis. About 50% of the analyzes involving Run1 data can be done using NanoAODRun1 [2]. NanoAODRun1 is an update of a data format calling NanoAODplus [2], which was used in 2021 summer school projects. Our steps are to first reproduce some analysis that has used NanoAODplus, for this we will use the study carried out in 2021 by the student Aritra Bal [3]. Then, we will do the same analysis, only using the datasets in NanoAODRun1 format, and finally we will compare the results found. We will use samples from 2010, 2011 and 2012 in this study.

2 Reproduction of the 2021 Z→2L study

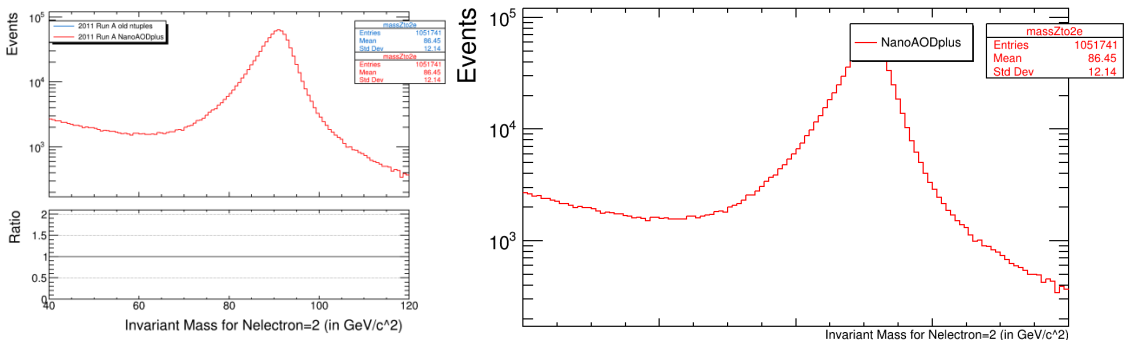
The first part of the study consists of reproducing a part of the analysis made in 2021 by Student Aritra Bal. We will acquire the invariant mass distribution of two leptons in order to find the peak mass of the Z boson. We apply the same quality and kinematic cuts made by Aritra on muons and electron. A description of the cuts can be seen below:

- $p_T > 7$ 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 p_T 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 $|SIP3D| < 4.0$ for both electrons and muons.

Furthermore, muons must be global, and both muons and electrons must be particle flow candidatess.

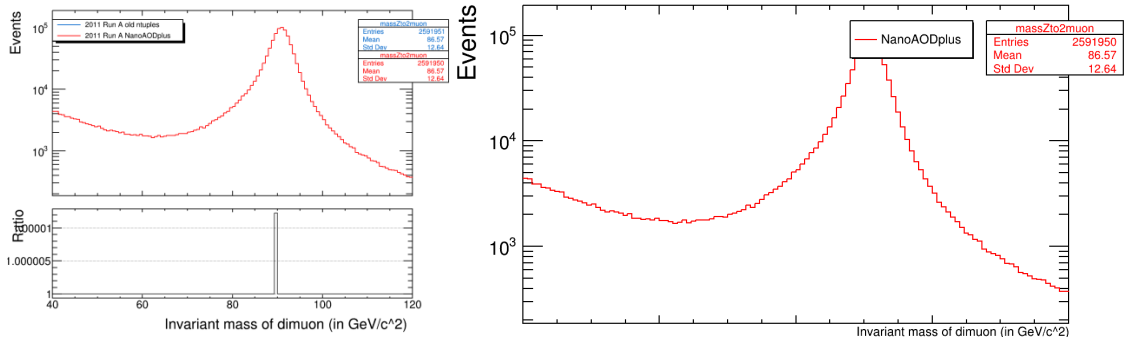
2.1 Validation of codes

A comparison was made of the distributions of the mass of $Z \to 2l$ for the year 2011 (figures 1 and 2) and for the year 2012 (figures 3 and 4). The values found were the same, thus validating the code used, as well as the NanoAODplus datasets.



(a) Reference (b) our distribution

Figure 1: $Z \to 2e$: Run A from 2011



(a) Reference (b) our distribution

Figure 2: $Z \to 2\mu$: Run A from 2011

3 Validation of format NanoAODRun1

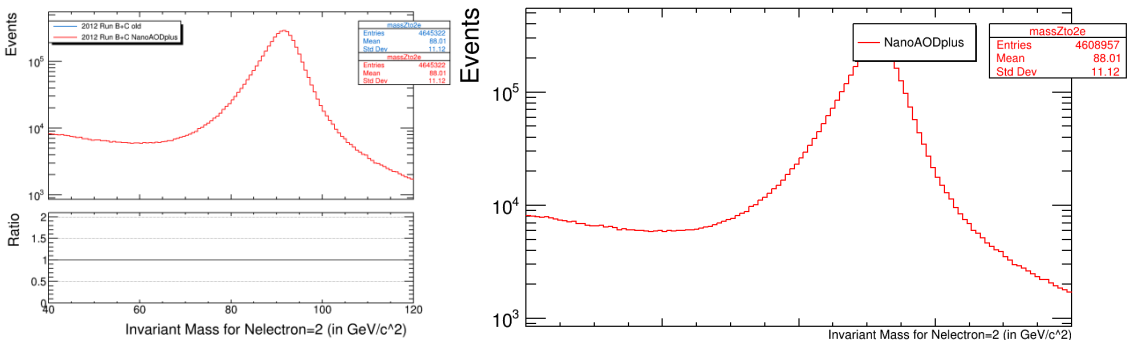
3.1 List of Samples

The validation of the NanoAODRun1 format is performed, through a direct comparison with the previous format (NanoAODplus), using different samples. In the table below are all available datasets in the NanoAODRun1 and NanoAODplus version

The name in red means that some files in NanoAODplus format are missing. The lack of some files does not prevent it from being possible to make a comparison between the different formats. The SingleMu2012C sample presents more events in the NanoAODplus version than in the NanoAODRun1 version, this behavior is not expected, and needs investigation.

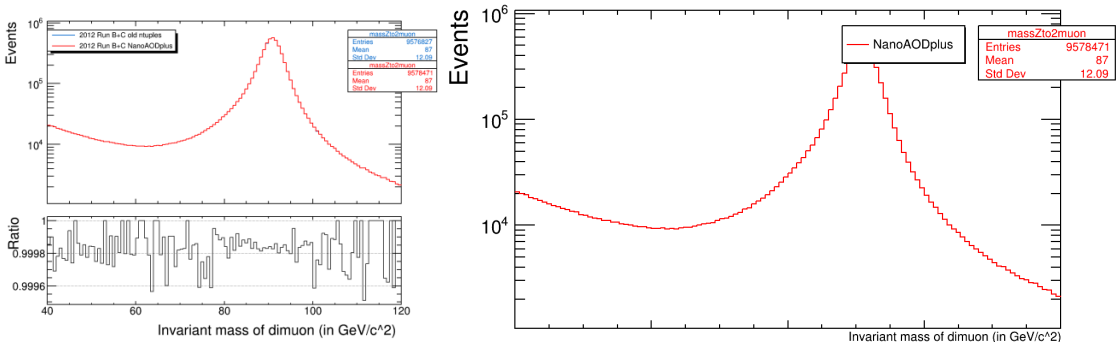
Sample	Era	Number of Entries	
		NanoAODPlus	NanoAODRun1
DoubleMu	Run2011A	35287778	35287778
	Run2011B	22334391	22334391
	Run2012B	26084708	26084708
	Run2012C	35455705	35455705
DoubleEl	Run2011A	44239644	44239644
	Run2011B	14588696	14588696
	Run2012B	26084708	26084708
	Run2012C	35455705	35455705
Drell-Yan	Run2011	34926233	34926233
SingleMu	Run2010A	20850125	20850125
	Run2010B	26718043	26718043
	Run2011A	78054209	78757263
	Run2011B	47134514	47134514
	Run2012B	53431234	53446198
	Run2012C	84700136	84151957
Electron/EG	Run2010A	47187984	47187984
	Run2010B	27856626	27856626
	Run2011A	37162334	37917536
	Run2011B	15956053	15956053
	Run2012B	60641353	62023592
	Run2012C	96942424	97927727
MuOnia	Run2010A	29583697	29583697
	Run2010B	22677534	22677534
	Run2011A	54375251	55915322
	Run2011B	25136423	26280382
	Run2012B	33196297	34025601
	Run2012C	58544497	60197132

Table 1: samples available



(a) Reference (b) our distribution

Figure 3: $Z \to 2e$: Run (B+C) from 2012



(a) Reference (b) our distribution

Figure 4: $Z \to 2\mu$: Run (B+C) from 2012

3.2 Comparison between NanoAODplus and NanoAODRun1

In figures 5,6,7 and 8 we have the distributions of the invariant mass and the relative isolation of the electrons, it is possible to see that the difference presented between the different formats resides in the way in which the isolation of the electrons is carried out. The NanoAODRun1 format differs from the NanoAODplus format in the way the electron isolation is calculated. The same behavior can be seen for different samples, specifically in figures 9,11,12,14,15,16 and 17. It is also seen a change in the way the muon isolation is done in the case of the 2010 data, being possible to see through figures 10 and 13.

3.3 Other studies using NanoAODRun1

The NanoAODRun1 format is a similar format to Run2's NanoAOD, so it should work in analyzes using the NanoAOD format. With that in mind, a test was made using the code of an analysis in development of Run2 but now using data from NanoAODRun1.

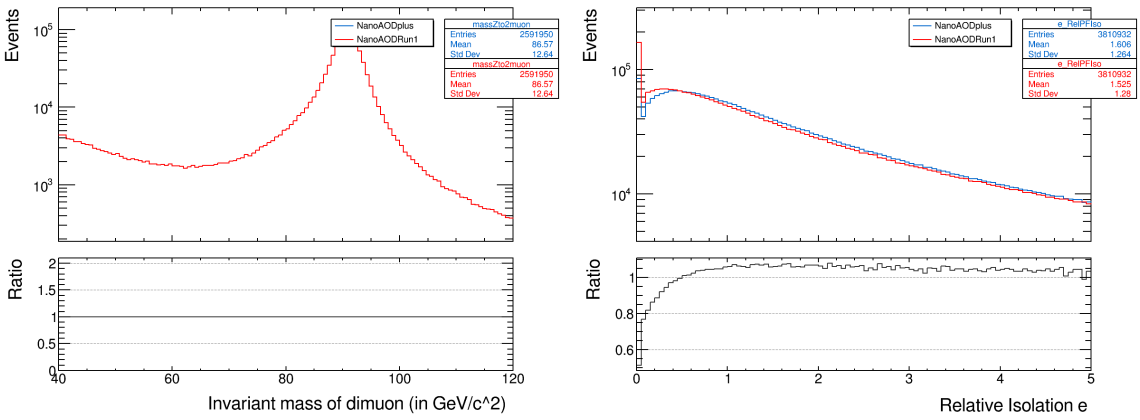


Figure 5: Sample DoubleMu2011 RunA

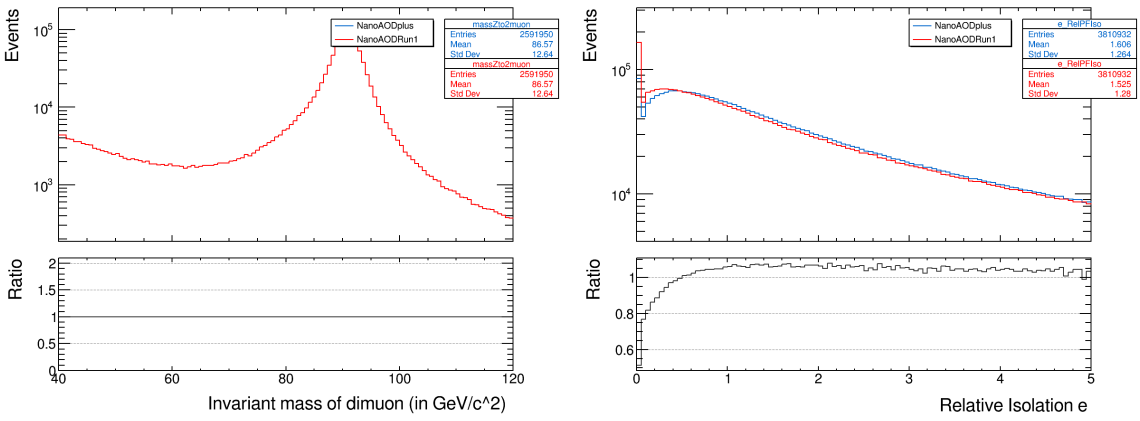


Figure 6: Sample DoubleEl2011 RunA

We selected for this study only the Drell-Yan datasets from 2011, and applied cuts to the MET and p_T of the leptons, a brief description of them can be seen below:

- MET > 50 GeV
- lepton $p_T > 20$ GeV
- leading lepton $p_T > 40$ GeV
- ΔR between the two leptons < 3.2

The code, specific to Run2, returns an error, containing the information that is missing in NanoAODRun1 compared to the NanoAOD format, some of this information is listed below:

- Secondary vertex information;

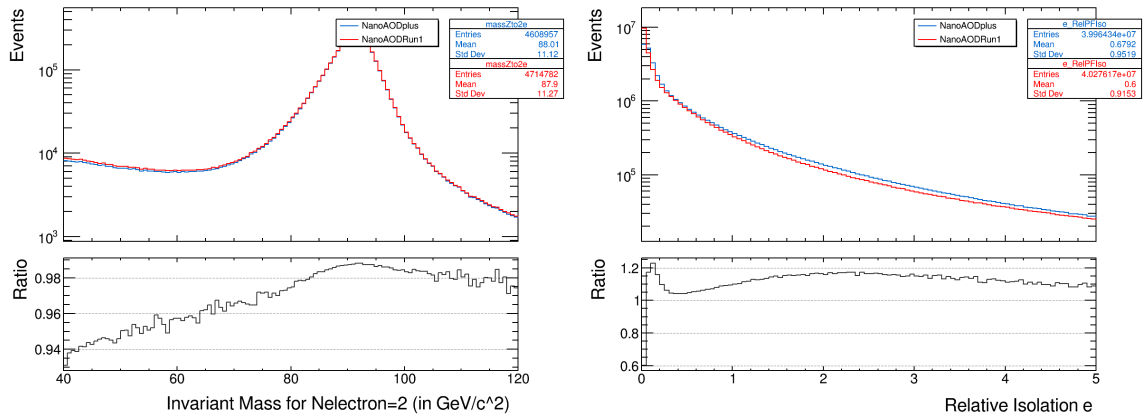


Figure 7: Sample DoubleEl2012 $RunB + C$

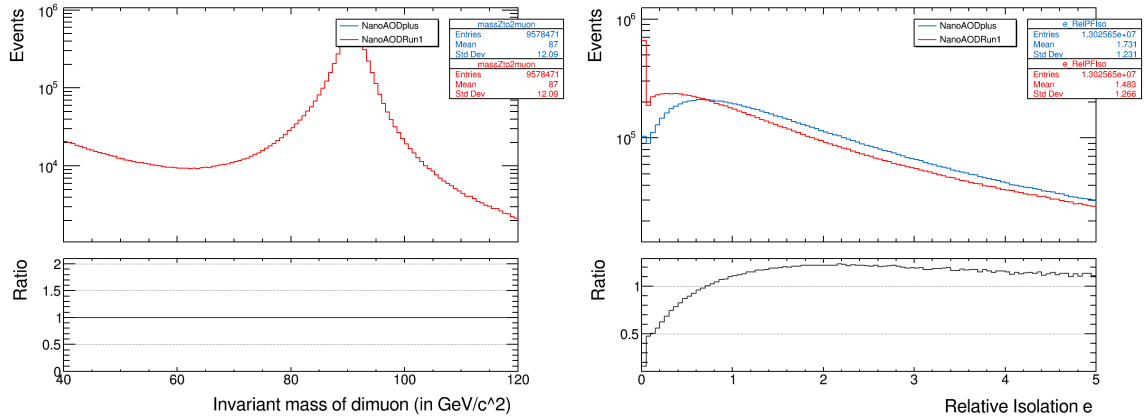


Figure 8: Sample DoubleMu2012 $RunB + C$

- IsoTrack information;

This study is not yet complete, but from it it was possible to see that there is information missing in the NanoAODRun1 format, when referring to the same information available in NanoAOD. Even with this information missing, it is possible to study the distributions, however this part has not been done yet.

4 Comparison between data and MC using ATLAS samples

A study of the Z peak distributions was also carried out using ATLAS data from 2012 and 2016. In order to carry out this study, we first pass the ATLAS data to a format similar to that of the CMS. Some cuts are already present in the Atlas data, namely:

- $p_T > 7.0$ GeV for electrons and muons.

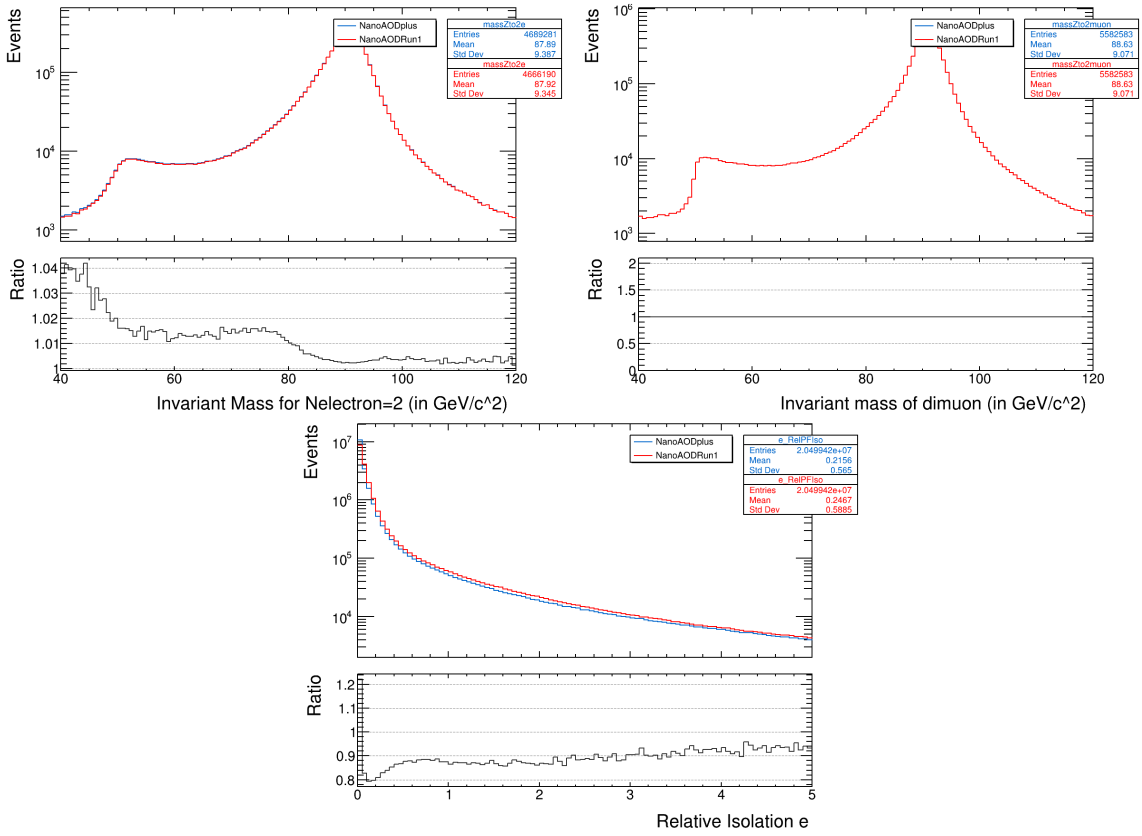


Figure 9: Sample Drell-Yan

- $|\eta| < 2.47$ 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.

The Atlas by definition enforces a veto that muons must be global and that both electrons and muons must be particles flow candidates.
It is important to point out that a normalization process was carried out, so that the area under the curve was equal to 1, in both data, since we just want to see if the shape of each sample is the same. Using data from 2016 and 2012 (Split between EGamma and DoubleMu) and the simulation Z→2e and Z→2mu we obtained the following results:

5 Conclusion

Through figures 5, 6, 7, 8 and 9 it is possible to see that the muons did not undergo any changes, acquiring a similar distribution for both formats. It is also possible to see that

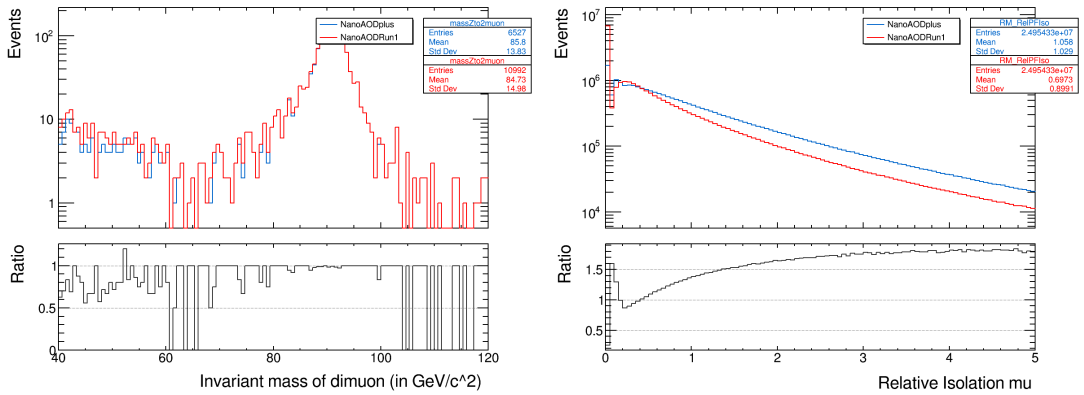


Figure 10: Sample MuOnia2010 RunA

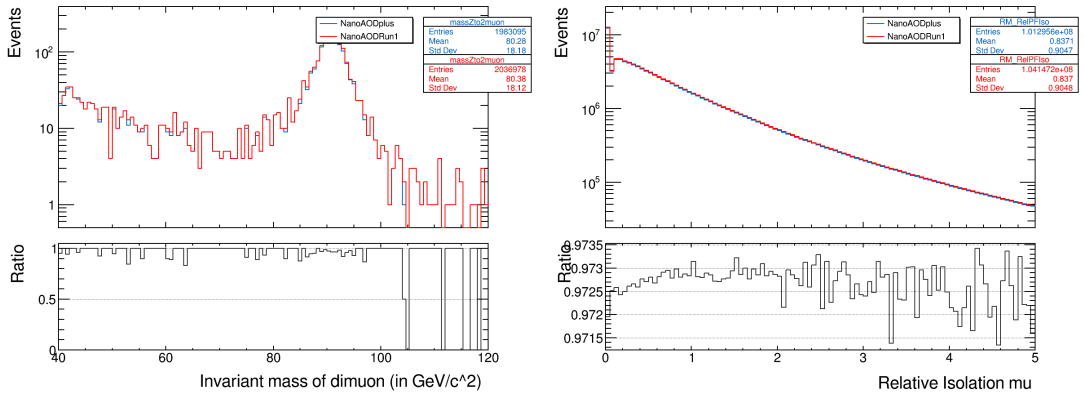


Figure 11: Sample MuOnia2011 RunA

the electrons underwent a modification in the way the relative isolation was performed, which is evident in the distributions presented.

Observing the distribution of muon isolation in figures 10 and 13, we found that there is some kind of modification in the way the muon isolation is carried out, being, in principle, only for the MuOnia and SingleMu samples from 2010.

The other distributions present an expected behavior. The difference presented in the invariant mass distributions for two muons is due to the missing files in the NanoAODplus format. Even with missing files it is possible to see that the new format was able to reproduce the results. The new NanoAODRun1 format was validated as it was able to produce distributions identical to those acquired in other older studies.

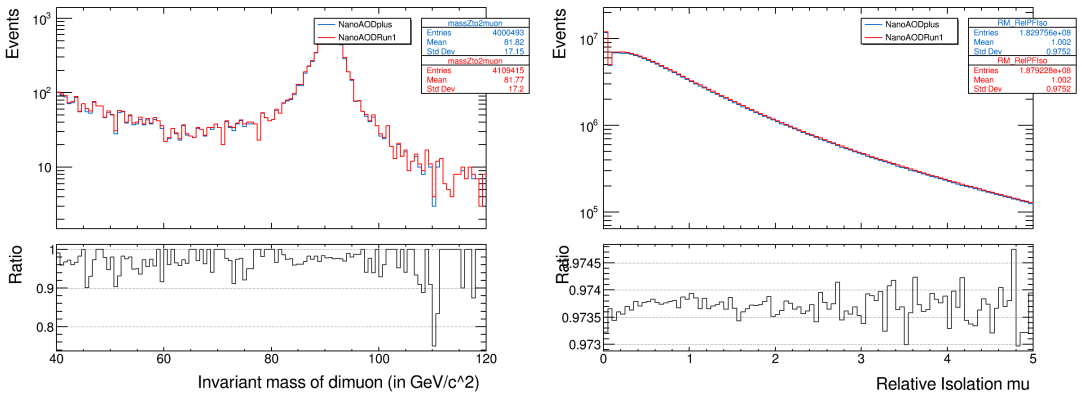


Figure 12: Sample MuOnia2012 $RunB + C$

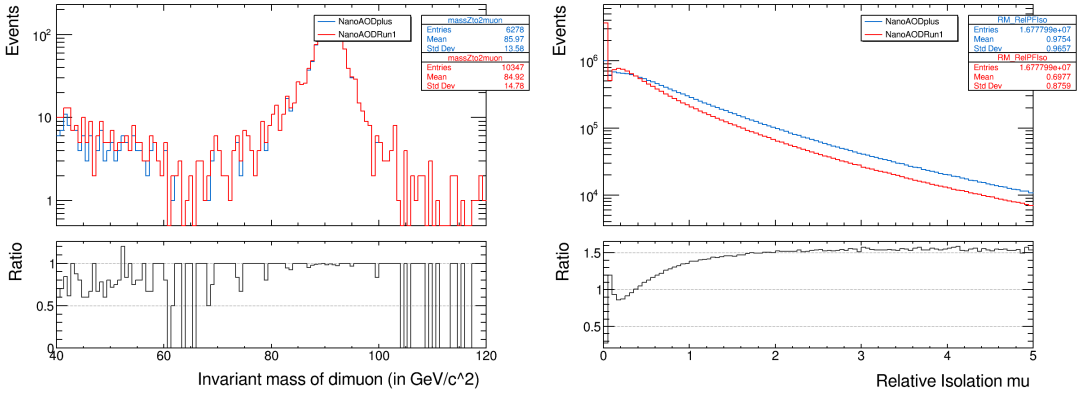


Figure 13: Sample SingleMu2010 RunA

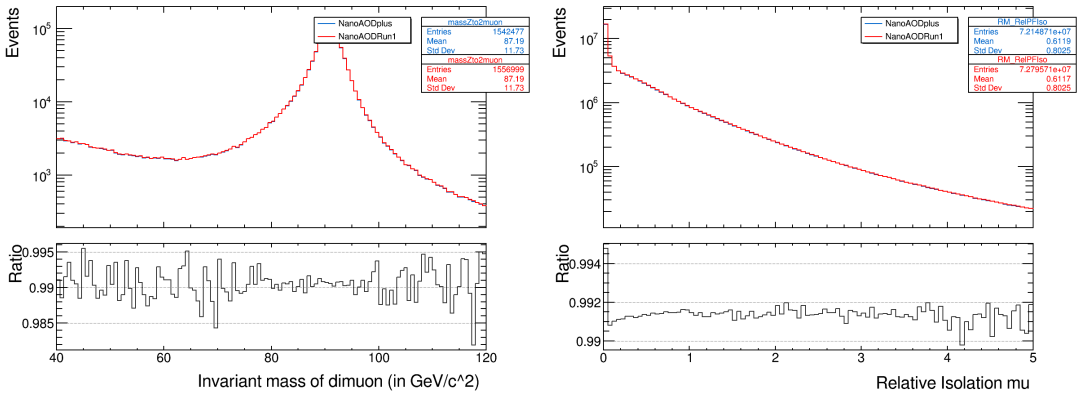


Figure 14: Sample SingleMu2011 RunA

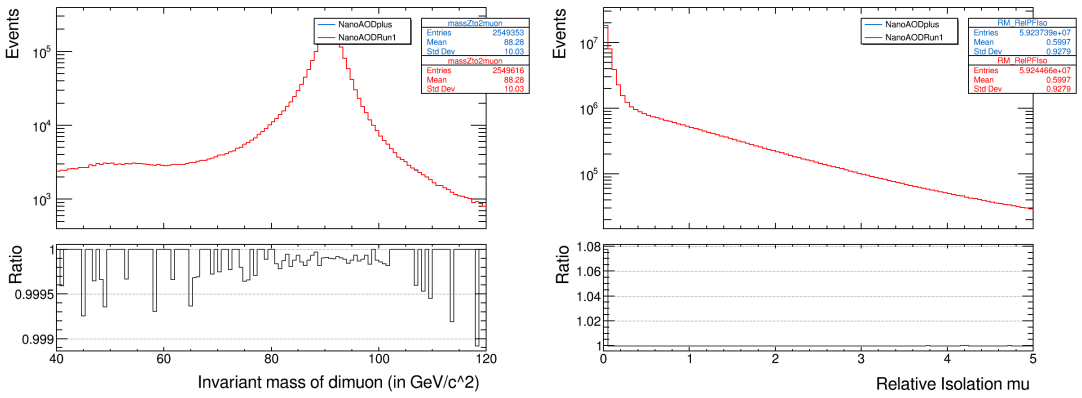


Figure 15: Sample SingleMu2012 RunB

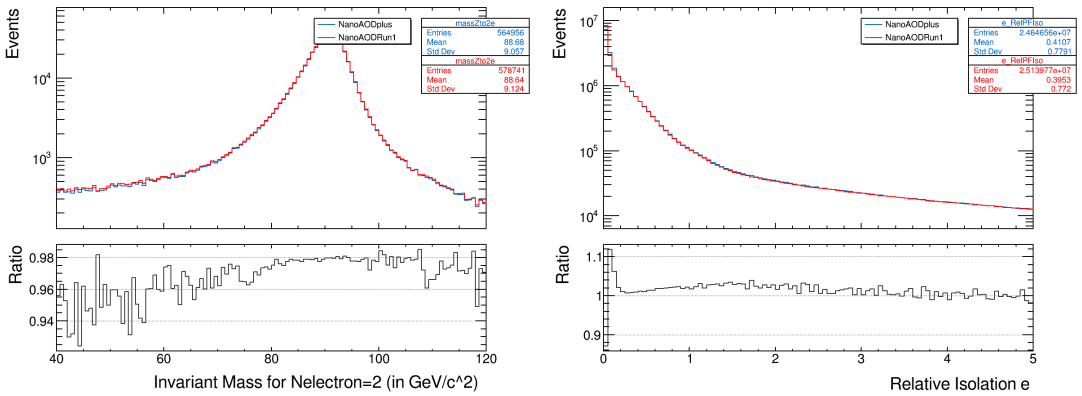


Figure 16: Sample SingleMu2012 RunA

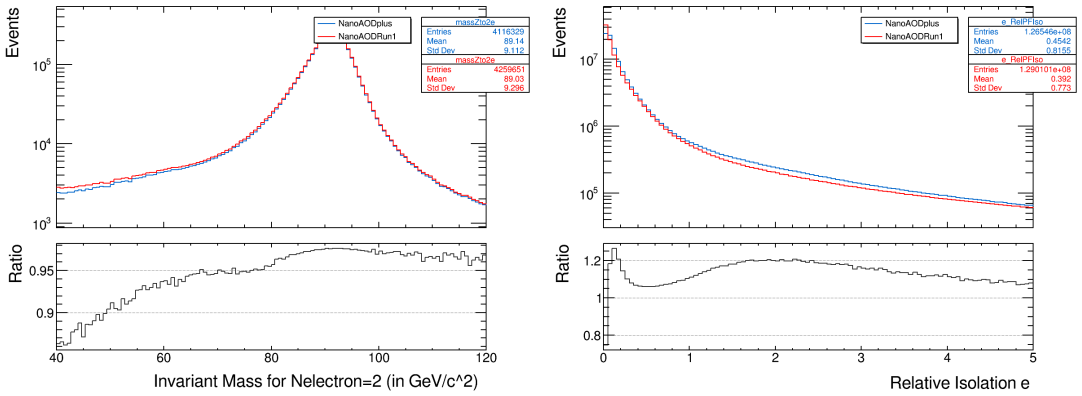


Figure 17: SampleSingleEl2011 $RunB + C$

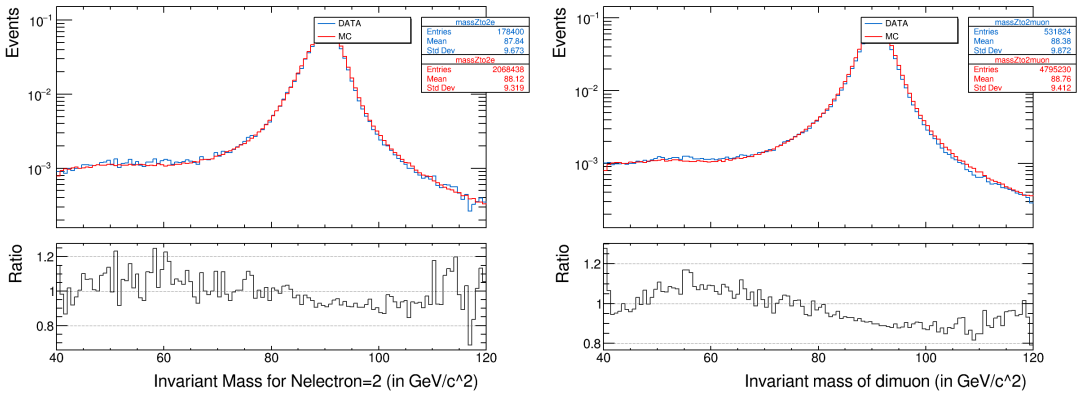


Figure 18: Comparation between data and MC using sample Z→2e and Z→2mu from ATLAs 2012

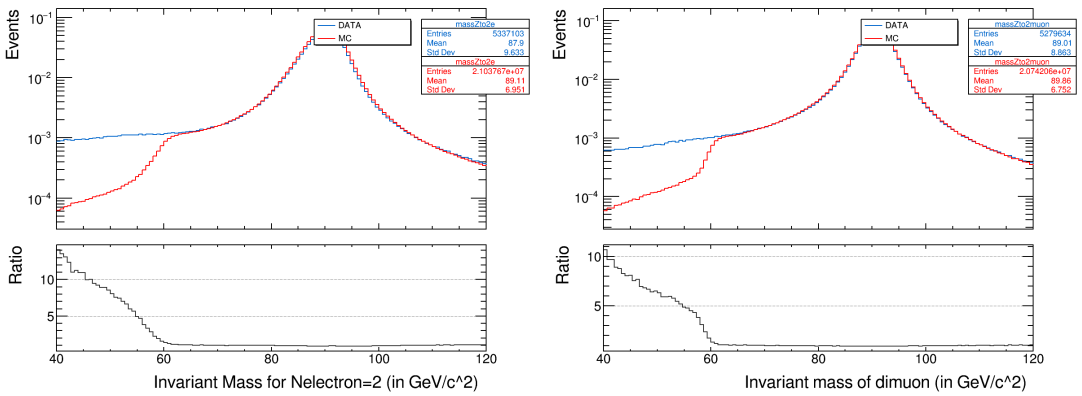


Figure 19: Comparation between data and MC using sample Z→2e and Z→2mu from ATLAs 2016

References

[1] 2019 Summer Project Report - Paula Martinez, https://www.desy.de/f/students/2019/reports/Paula.Martinez.pdf

[2] The DPOA NanoAODRun1 data tier - Achim Geiser, https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAODRun1

[3] Combined $Z \to 2L$ and $H \to 4L$ distributions using ATLAS and CMS (Open) Data - Aritra Bal, https://www.desy.de/f/students/2021/reports/Aritra.Bal.pdf