Higgs to four leptons using nanoAODplus

Paula Martínez Suárez
(supervised by Achim Geiser and Nur Zulaiha Jomhari)

DESY Summer Student Programme 2019

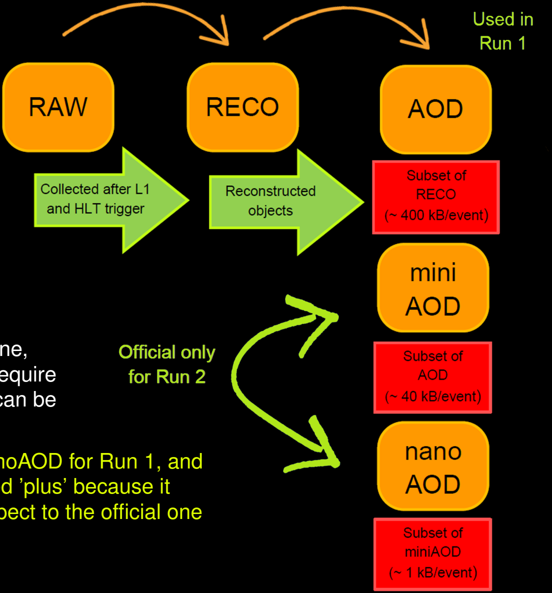
What is nanoAODplus?

The AOD format is a subset of all the reconstructed data (RECO), with enough information to carry out most of the Physics analyses.

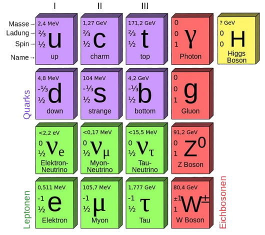
Due to the lack of computational resources and the increasing of the collected data, CMS started developing more simplified data formats for Run 2, like miniAOD and nanoAOD.

The nanoAOD format is the lightest one, and its is the only one that does not require a CMSSW environment to be run. It can be read using only ROOT.

nanoAODplus is the equivalent of nanoAOD for Run 1, and it is still under development. It is called 'plus' because it contains additional variables with respect to the official one (from Run 2).



Why is the Higgs boson important?



The Standard Model by itself does not allow the vector (spin 1) bosons to have masses, due to gauge invariance.

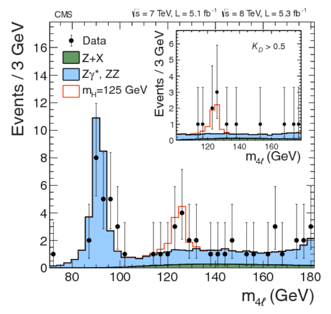
The photon does not have mass, but the Z and W bosons do $\rightarrow$ electroweak symmetry breaking.

To describe this symmetry breaking we need to introduce a new scalar (spin 0) field, which leads to a new particle, the Higgs boson.

This field is also responsible for the masses of the fermions, whose coupling to the Higgs boson is proportional to their masses.

The Higgs mechanism was theorized in the 1960s, but its existence was not confirmed until 2012, in the CMS and ATLAS experiments at the LHC (CERN).

(Right) Distribution of the four-lepton invariant mass for the $H \rightarrow ZZ \rightarrow 4\ell$ analysis. The points represent the data, the filled histograms represent the background, and the open histogram shows the signal expectation for a Higgs boson of mass $m_H = 125$ GeV, added to the background expectation.

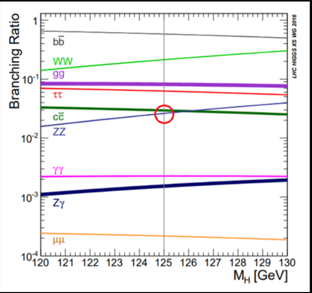


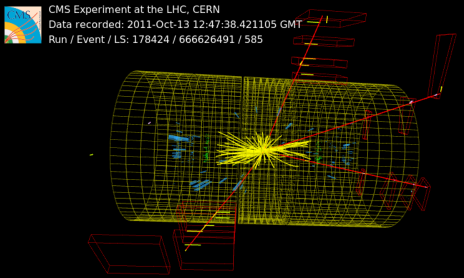
The four leptons decay channel

The $H \to ZZ \to 4\ell$ decay channel (along with $H \to \gamma\gamma$) is one of the best options to reconstruct the mass peak of the Higgs boson, since the leptons can be measured with high precision and therefore the resolution is good.

There are decay channels with higher cross sections, but they have disadvantages:

- The hadronic channels have more background due to the nature of the proton-proton collisions at the LHC.
- The WW channel contains neutrinos, which can not be detected.



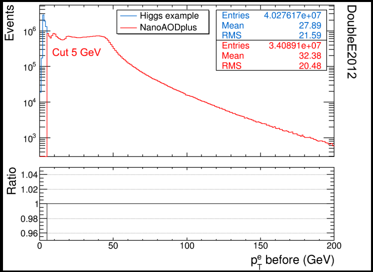


The CERN Open Data Portal

- CERN Open Data Portal is a platform which allows public access to the data produced by the different research activities at CERN. For CMS it contains simplified datasets, reconstructed data and simulations, and the necessary analysis software.
- A research level example for the Higgs to four leptons process using 2011 and 2012 data is provided. It can be used to produce the invariant mass of the four leptons from AOD data.

The aim of this project is to produce the same plot, as similar as possible, using the nanoAODplus format.

This is not a trivial task, since nanoAOD format contains additional cuts, used to reduce the size per event and make it more efficient.

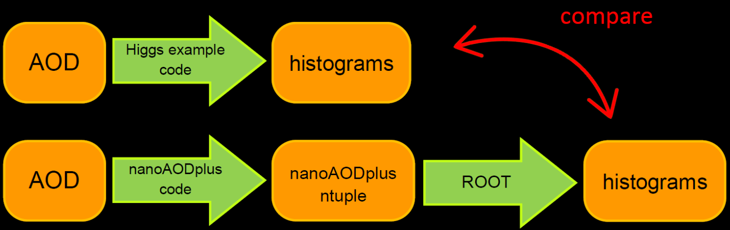


p_T of electrons before aplying any kinematic cut (2012 dataset).

The challenge

The steps that we need to follow are:

- Matching the variables from AOD and nanoAODplus format.
- Verify that they are correctly defined and identify internal cuts in the nanoAODplus.
- Find the cuts in the Higgs example and apply them to our nanoAODplus ntuple.
- Produce the same histograms and compare the final results.



Selection cuts

Quality cuts

Global muons
Particle flow muons and electrons

Kinematic cuts $\rightarrow$ to save only the good muons/electrons

Main cuts
$p_T > 5$ GeV for muons, $p_T > 7$ GeV for electrons
$|\eta| < 2.4$ for muons, supercluster $|\eta| < 2.5$ for electrons

For both muons and electrons
$|$Impact parameter significance$| < 4$
$|$Distance in xy to the vertex$| < 0.5$
$|$Distance in z to the vertex$| < 1$
Relative isolation < 1

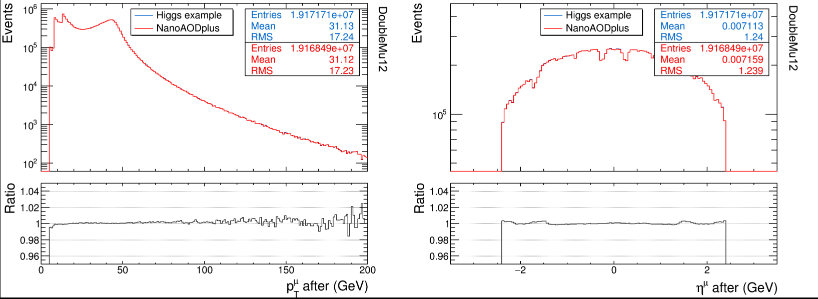
Only for electrons
Misshits ≤ 1
Within barrel or endcap acceptance

Further cuts

If we want to keep only $H \rightarrow ZZ \rightarrow 4\ell$ we need:

4 good ℓ
Total charge = 0
Charge of each pair (that comes from a Z boson) = 0

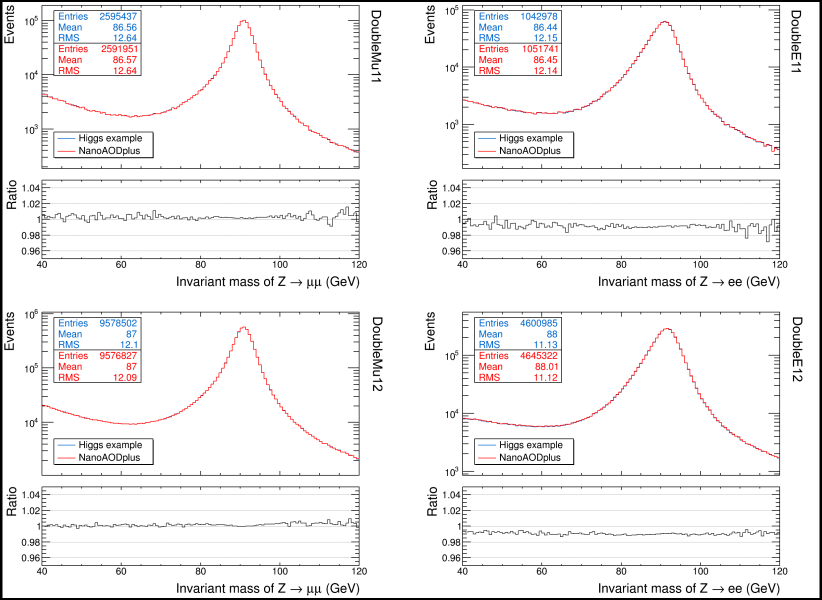
Control plots



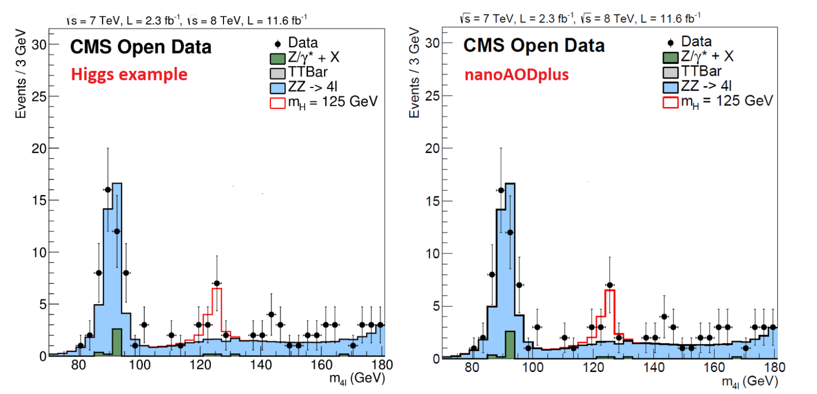
Example from DoubleMuon dataset (2012), transverse momentum and eta of the muons. The agreement between the Higgs Example (AOD) and the nanoAODplus (nanoAOD-like) after applying the cuts is good.

(Next slide) Reconstructed mass of the Z boson using the highest p_T lepton pairs for DoubleMuon 2011 and 2012 and DoubleElectron 2011 and 2012 datasets.

Control plots



The Higgs to four leptons plot



Higgs to four leptons plot produced from AOD data (left) and nanoAODplus data (right). The agreement is good.

Conclusions and outlook

- The agreement between the AOD and the nanoAODplus is good enough to reproduce the results, but it still needs to be improved.
- This differences are not a problem as long as we understand the cause.
- Next step: using the same variables as the official nanoAOD to test our code with Run 2 data.

Thank you for your attention :-)

Backup

Datasets

Data 2011	DoubleMu DoubleE
Monte Carlo 2011	HZZ ZZto4mu ZZto2mu2e ZZto4e DY1011 DY50TuneZ11 TTBar11
Data 2012	DoubleMuParked DoubleEParked
Monte Carlo 2012	HZZ ZZto4mu ZZto2mu2e ZZto4e DY1012 DY50TuneZ12 TTBar12

Same as in the Higgs example.