Analysis of Full Monte Carlo Simulation in Preparation for Upgrades to LHC

Christopher Hayes
(Dated: August 9, 2013)

The luminosity upgrades to the LHC will provide a new set of challenges for data analysis in channels relevant to the Higgs boson and beyond the standard model searches. Here, I describe my analysis of CMS Monte Carlo Simulation in the Drell-Yan to ee and WW -> $2l2\nu$ decay channels under 140 pileup conditions.

I. INTRODUCTION

After the second run of the LHC ends in 2017, the machine will be further upgraded to achieve higher luminosities, with the expected luminosity being on the order of $2 * 10^{34} cm^{-2} s^{-1}$. The center of mass energy will remain the same as the expected energy to be seen in LHC Run 2 at $\sqrt{s} = 14$ TeV. The high luminosity achievable by the machine after this upgrade will come at a cost, however, as the number of events occurring per bunch crossing must increase. The expected number of pileup occurring in these events is projected to be ≈ 140 events/crossing in comparison to 50 events/crossing, the maximum pileup at the end of LHC Run 1 [1].

In preparation for these upgrades, the CMS collaboration updated its simulations to determine the respective expectations for analysis under the upgraded conditions. My project deals with the analysis of full Monte-Carlo simulation of Drell-Yan based electron pair production and WW->$2l2\nu$ events in the CMS detector under these high pileup conditions. A secondary goal was to help tune (if possible) the parameters needed to extract our relevant information from these samples. These channels are of particular importance because of their contributions to the search for and understanding of the properties of the recently discovered Higgs boson.

II. DRELL-YAN EVENTS

The Drell-Yan process is a lepton pair production process, where a quark anti-quark pair annihilate to form a Z/γ^* which then decays into two leptons of opposite charge. For this case in my research, I study the Drell-Yan processes which produce a parent Z boson that decays into an electron-positron pair that is seen by the detector. These events are of specific interest to the understanding of the Higgs boson as they are major contributors to the background events seen in the H-> ZZ channel. These events are also used as a standard for studying calibration of analysis because of the narrow Z mass resonance peak seen in the mass spectrum.

My first analysis of these 140 pileup events looked at variables dealing with regions of the CMS detector known as superclusters. The CMS detector is divided into regions based on the task that each region performs. One of these regions, the energy calorimeter or ECAL, is composed of lead tungstate ($PbWO_4$) crystals. Electron and photon showers do not typically hit a single crystal; instead, much of their energy is spread out over a gridded 3x3 area of crystals or a 5x5 area of crystal. These gridded areas are called "superclusters" and they are used throughout the reconstruction of events in the CMS detector. Thus, by looking at the various parameters that deal with the superclusters, one can determine if the energies and shower shapes match various reconstruction parameters.

Many of the variables of interest in these 140 pileup simulations matched the expectations given by Monte-Carlo samples from the 2012 conditions of the LHC. Several of the supercluster variables, however, showed a significant broadening that deviated from the expectations given by the 2012 conditions. Plots of the R9 parameter (an energy ratio, Energy 3x3/Energy 5x5), supercluster η width, and supercluster ϕ width (the detector geometric variables of pseudorapidity and angle) are presented in Fig. 1. By increasing the cut on p_T, the broadening becomes less pronounced, but these variables require more study at this time to determine the cause of this broadening at low p_T.

The next step in the analysis of these DY-> ee events was to determine the missing transverse energy (MET), E_T seen in these Drell-Yan events. Missing energy is the energy that is not seen in the detector, but is expected because energy in the transverse plane should be conserved, except in the decays which involve neutrinos and BSM physics which both involve undetected particles. Thus, in ideal Drell-Yan events, a total conservation of energy in the transverse plane should be seen. Fake MET, however, may arise in these situations that comes from misidentified or mis-reconstructed particles. In our Drell-Yan-> ee events, therefore, we expect to see a small MET that might arise from these "fake" sources.

I first compared the following two methods of MET calculation: 1) The MET calculated from the reconstruction of all particles in the events, known as the particle flow (PF) MET, 2) The MET calculated from the reconstructed particles that leave behind tracks on the silicon tracker, or track MET. Because the tracker only detects charged particles, the track method may lose some information on the MET as it does not account for photon energies or neutral hadron energies that are seen later in the ECAL and hadronic calorimeter (HCAL) respectively. I also

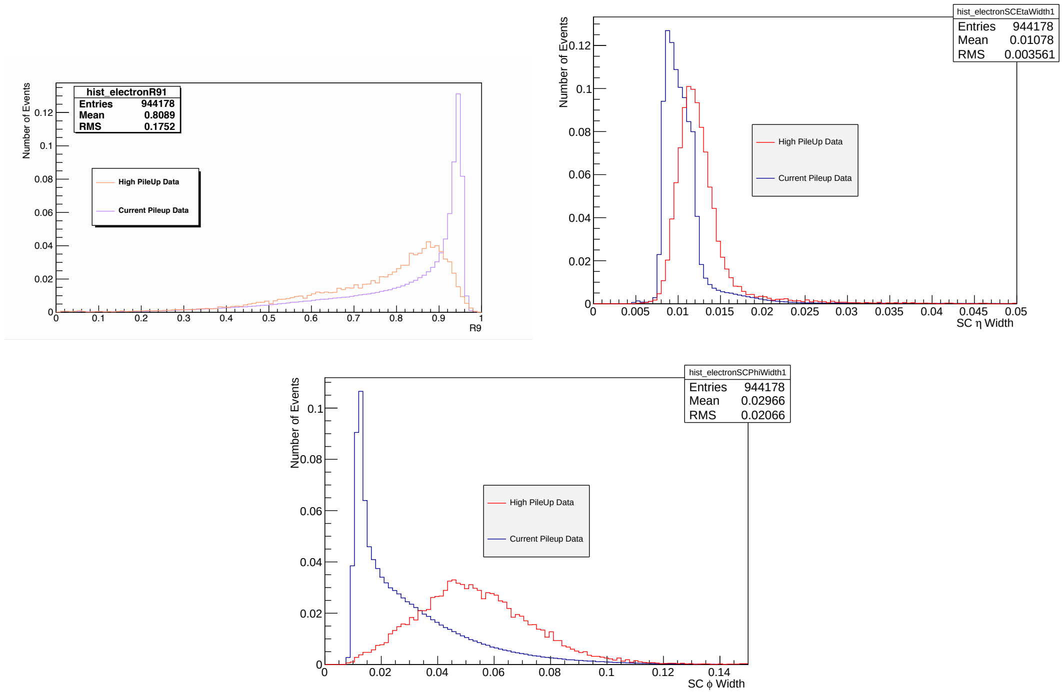

FIG. 1. From top right, clockwise: The Plots of the r9 parameter, supercluster η and ϕwidth for transverse momentum, $p_T > 20GeV/c^2$ and located in the barrel region of the detector. A significant feature is seen in the broadening of these distributions in the high pileup data. This broadening effect is seen in the end cap region as well and lessens as p_T becomes larger.

constructed a third option for the MET analysis that involves a vectorial sum of all the relevant objects seen in our events. The relevant objects in the case of our Drell-Yan events are electrons and isolated hadronic jets. The idea behind this vector-based jet MET is that some events may be mis-identified by our particle flow algorithms in these high pileup samples. Thus, by requiring further isolation constraints, I could obtain a more proper value for the MET by adding these objects under further constraints to reduce any type of fakes.

The results from the three methods are given in Fig. 2, where the Mean and RMS are plotted as a function of pileup. I find that the PF MET and vector based MET perform at about the same level. Thus, I obtain no additional discerning power or other benefits by developing this third method of MET calculation. The effect of these relatively high values of the PF MET and Track MET on the efficiency for MET recognition will be covered more in section III.

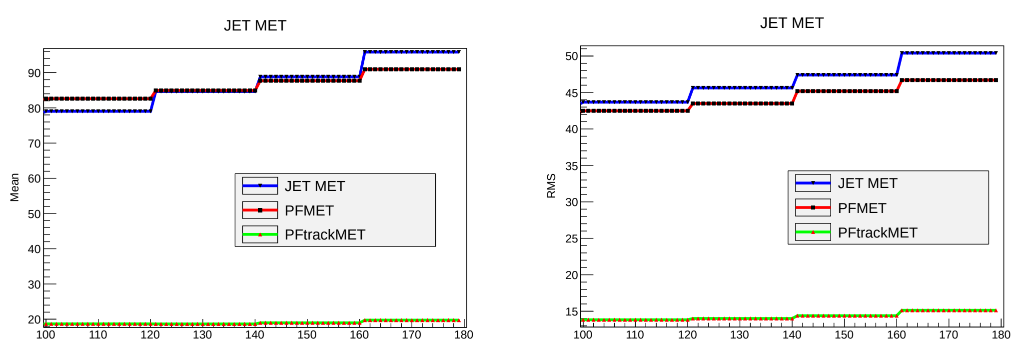

FIG. 2. A comparison of the MET methods with the mean and RMS as a function of pileup in the high pileup Drell-Yan to 2 electron samples. The vector based jet method and particle flow methods perform at similar levels, while the track MET is much lower.

My last analysis with the Drell-Yan events involves a comparison between the efficiencies of isolation seen in 2012 pileup conditions and those seen in our current 140 pileup Monte Carlo conditions. The particle flow algorithms used by CMS set specific conditions on how isolated an object such as a photon, charged particle, or neutral hadron must be to be reconstructed. These isolation parameters can thus be a measure of how effectively a particle in a signal event can be identified against the background process.

Here, I treat the DY-> ee events as a signal against a background that arises from Z boson decay into two neutrinos and a secondary process that provides the QCD background. This background accurately models electrons which arises from b-jet that are detected as "fake" electrons. By studying these events against the Drell-Yan signal, we can determine our ability to differentiate between real electrons and electron fakes. To illustrate our ability to discern the signal from background, I produced ROC curves based on the particle isolation in the two samples. The curves for the particle flow isolation efficiencies are seen in Fig. 3.

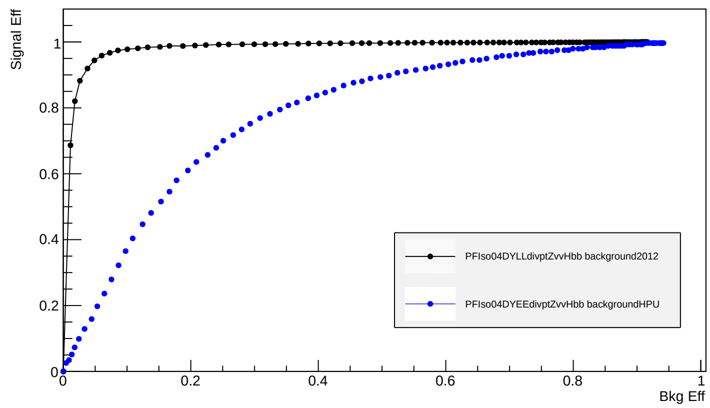

FIG. 3. A comparison of efficiency in the particle flow isolation between 2012 pileup samples and the 140 pileup samples. The 140 pileup samples have a diminished efficiency compared to the 2012 efficiencies. All cuts are made on particle $p_T > 10$ GeV. A qualitative example: 82% signal efficiency with 97% background rejection for particle flow in 2012 MC compares to the 82% signal efficiency with 65% background rejection for particle flow isolation in the 14 TeV, 140 Pileup MC.

These curves show that the isolation efficiency in the high pileup samples is diminished from the near perfect efficiency of the 2012 simulations. In order to determine which specific portions of the isolation are the most inefficient, I also determined the photon, charged particle, and neutral hadron isolations efficiency. These efficiencies can be seen in Fig. 4.

The exact cause behind the diminished charged particle and neutral hadron isolation efficiency may arise from the distributions of the background isolation spectra of the charge particle and neutral hadron backgrounds. These spectra will be further examined at a later time, but my contribution has provided a first look at this issue with particle reconstruction.

III. WW-> $2l2\nu$ BACKGROUND EVENTS

My next important contributions arose in the WW-> $2l2\nu$ channel of the Higgs decay. Due to availability of certain high pileup samples, I used samples which modeled the background WW events. As the kinematics between these samples and those which contain the Higgs share many of the same properties, I was able to apply the necessary cuts needed for the Higgs search to these samples.

My first task with the WW-> $2l2\nu$ channel dealt with the comparison of MET in these 140 pileup conditions. These events provided for an excellent extension of the studies seen in Section II as these events have actual MET arising from the 2 neutrinos that go unseen by the detector. I once again applied the three MET calculations to these samples and the results can be seen in Fig. 5. Once again, the vector based MET did not perform at a level which was beneficial to these studies. The PF MET and Track MET also perform at around the same level as seen in the Drell-Yan events. Thus, to determine the full effects of these disparities between the PF and Track MET methods, I performed efficiency studies on these WW MET samples.

Using the Drell-Yan events as our effective background, I produced the efficiency curves shown in Fig. 6. The three comparisons made in this plot include the efficiencies of the Track MET, the PF MET, and a minimization of the MET between the two methods. This minimization method proved unsuccessful, as it does not add any discerning power to the Track MET efficiency. I also found that PF MET events were severely affected by the high pileup, with almost no ability to discern between the signal and background.

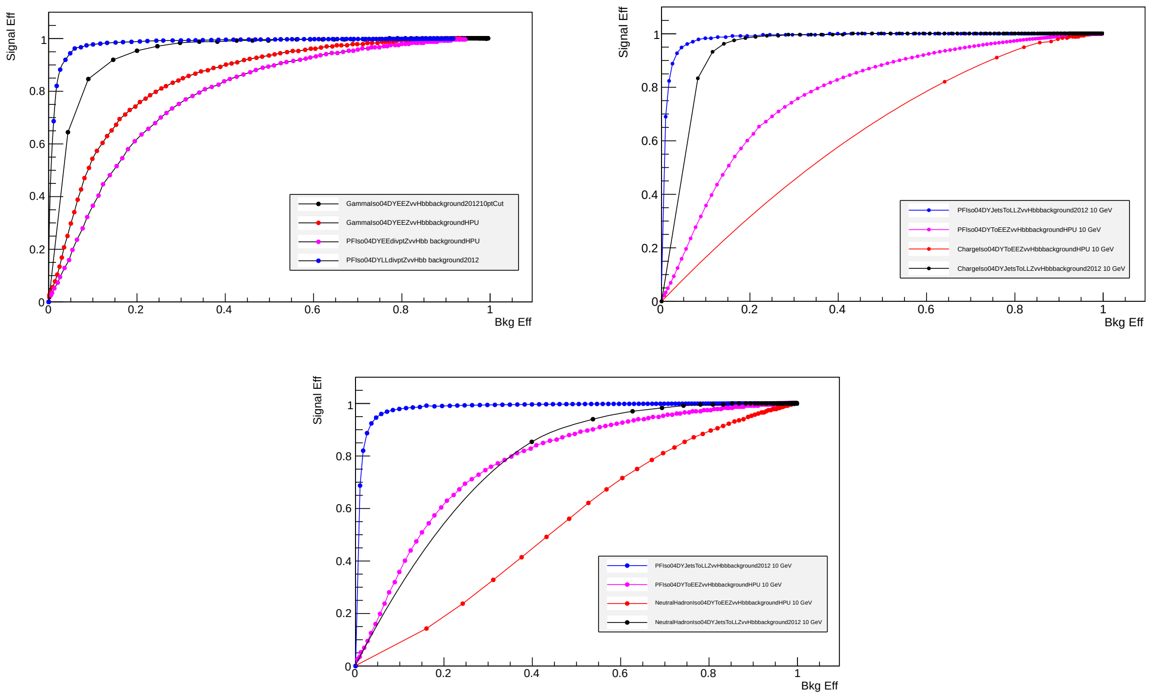

FIG. 4. From top right, clockwise: The plots of the isolation efficiency of the photons, charged particles, and neutral hadrons identified by the particle flow algorithms. The PF isolation at high pileup is labeled in purple in all plots, with the other high pileup isolation compared in red. The 2012 samples are labeled blue and black respectively. Electron cuts: $p_T > 10$ GeV.

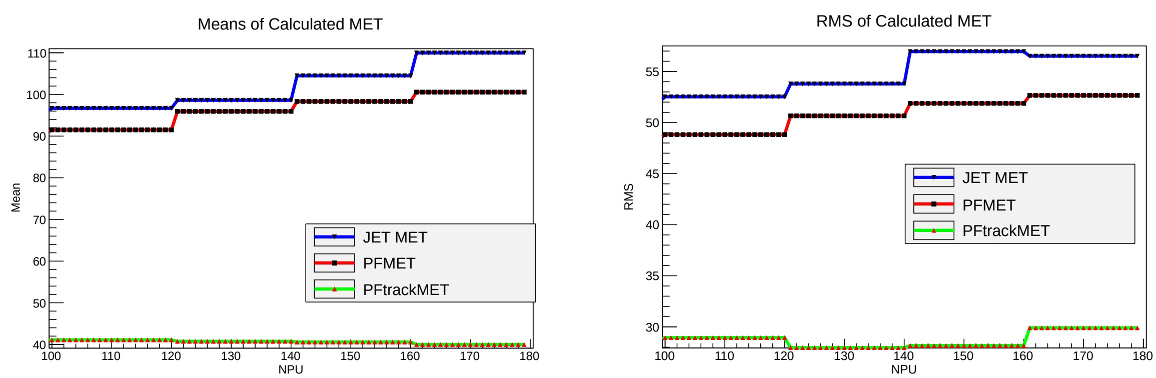

FIG. 5. A comparison of the MET methods with the mean and RMS as a function of pileup in the high pileup WW-> $2l2\nu$ samples.

All of my prior research culminated with a final contribution to a physics case that relates WW events to the search for the Higgs. By applying many of the cuts (excluding jet number in the high pileup samples) used for the 2011 and 2012 data runs of the LHC found in Ref. [2], I attempted to reproduce several of the spectra of interest to the Higgs search in this channel. The dilepton invariant mass spectra in these events must discriminate between events of the same flavor leptons and different flavor leptons because of the different physical conditions under which these events occur. The comparisons of the 140 pileup simulation and the 2012 simulation, separated in terms of flavor, are given in Fig. 7. These plots indicate that the invariant mass of the dilepton system still remains a powerful discriminating variable in these high pileup scenarios. I also show in Fig. 7 a plot of the transverse mass variable which is used as a discriminate in the Higgs selection analysis. This variable in the high pileup matches the spectrum seen in the 2012 simulation. All of these findings indicate that many of our physical analyses remain true under these high pileup conditions in this particular decay channel.

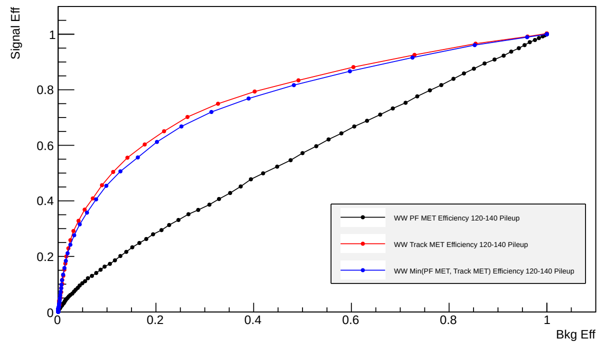

FIG. 6. The efficiencies of discerning real MET from fake MET at 120-140 PU in these high pileup Monte Carlo simulation. We find that only the track MET remains a discerning variable. The minimization method does not add any discerning power to our studies of the MET.

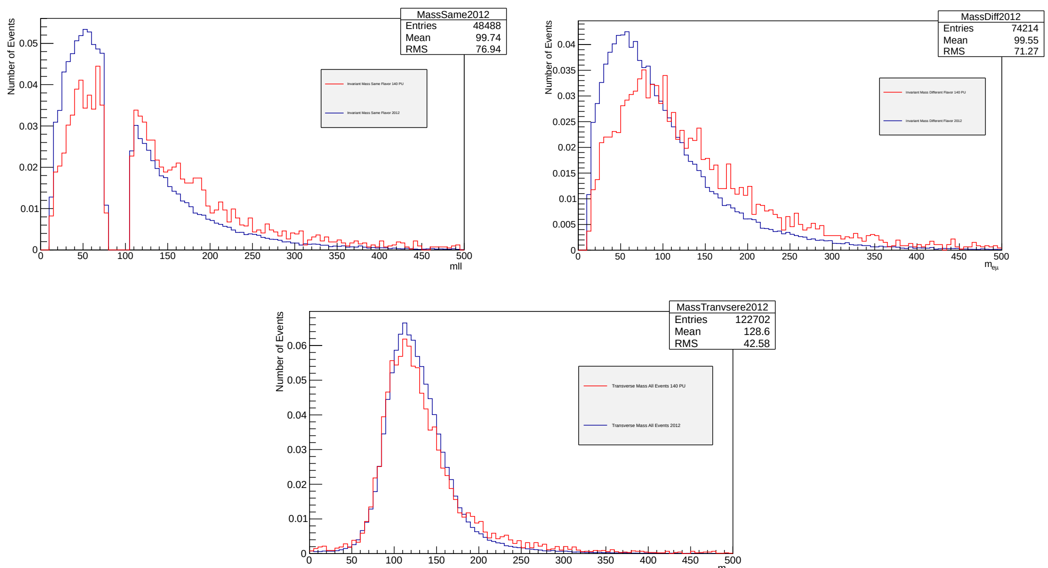

FIG. 7. Left, Clockwise: The invariant mass spectrum of WW$->2l2\nu$ with same flavor leptons. Right: The invariant mass spectrum of WW$->2l2\nu$ with different flavor leptons. Bottom: The transverse mass spectrum. All of these variables agree with the 2012 simulation expectations.

IV. CONCLUSIONS

With the completion of my analysis, progress has been made in understanding the different physics or other issues involved in these high pileup simulations. Further work will be needed to determine the exact causes of the low efficiencies seen in the isolation of electrons when compared to the 2012 simulations. From the picture presented by the WW events, the backgrounds may be similar to previous expectations. Thus, more analysis may lead to better cuts than those used in the 2012 analysis for this higher pileup and energy simulation and the proper picture of these events will be better reproduced.

[1] A. Avetisyan, J. M. Campbell, T. Cohen, N. Dhingra, J. Hirschauer, *et al.*, (2013), arXiv:1308.1636 [hep-ex].
[2] S. Chatrchyan *et al.* (CMS Collaboration), Phys.Lett., **B716**, 30 (2012), arXiv:1207.7235 [hep-ex].