

Towards an Improved Estimator of the W Boson Recoil at High Pileup

C. Day-Lewis, P. M. Vieira de Castro Ferreira da Silva, J. Bendavid
European Organisation For Nuclear Research (CERN), Switzerland
(Dated: October 19, 2018)

This report describes a method to estimate the hadronic recoil of the W boson from high pileup events of W decay to a lepton and neutrino, using machine learning techniques (specifically boosted decision trees). Monte Carlo (MC) simulations of the decay have their limitations and a correction to data from the CMS (Compact Muon Solenoid) experiment is needed. Data of decays of Z bosons to two leptons were used to test that the method was able to estimate the hadronic recoil. Two different techniques one involving quantile regression and another involving vertical morphing are presented and compared. Quantile regression worked well for a small number of features in the training; looking at the difference between the corrected MC data and the detector data gave rise to a χ^2 of 70.9. Whereas, quantile morphing worked well for a larger number of features; the difference here between corrected MC and detector data was found to create a χ^2 of 167.

INTRODUCTION AND THEORY

The precise measurement of the mass of the W boson has been ongoing since its discovery in 1983 [1], and it is an infamously challenging measurement. The standard model accurately predicts the mass of the vector bosons by using the Higg's vacuum expectation and the electroweak coupling as input [2]. Using this method the mass of the W boson was known to an uncertainty of 80358 ± 8 MeV [3]. This is much better than the experimentally measured world average of 80385 ± 15 MeV. In 2018, the ATLAS collaboration measured the mass of the W boson to 80370 ± 19 MeV [4]. One of the challenges of measuring the M_W (the mass of the W boson) is estimating the hadronic recoil. Hadronic recoil is the sum of the momentum of all components in an interaction apart from the bosons; it is, by definition, also equal to the momentum of the bosons.

In proton-proton collisions a potential decay mode of the W boson goes as the one shown in the Feynmann diagram in figure 1a. In this interaction the W boson decays to a lepton and a neutrino. This decay yields experimental degrees of freedom. One necessarily cannot know the individual momenta of the partons in any particular interaction. This missing initial momentum information along with the missing momentum of the neutrino means that it is not possible to directly measure the hadronic recoil. Note this is not the case for Z bosons as their decay mode is to two leptons an example of which is given in figure 1b. A way to overcome this is to look at Monte Carlo (MC) simulations of events and see how they need to be corrected in order for the simulations to match the data from detectors. This method allows for the hadronic recoil to be estimated.

In this paper, a description of a method using machine learning to correct from MC simulations to data from the CMS experiment 2016 data is presented. The starting point of this work comes from the thesis of Nicolo Foppiani and Olmo Cerri [5] [6]. They describe a method using decision trees to estimate the recoil. The principle behind the method is that the recoil although, incalculable directly due to the experimental degrees of freedom, can be ascertained from other measurable quantities in a collision. For the purpose of this paper

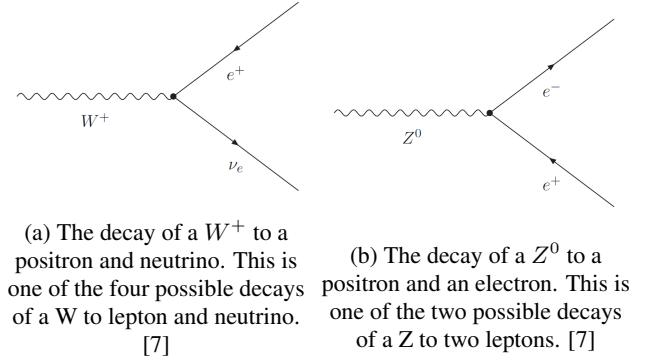


FIG. 1: Feynmann diagrams of electroweak carrier decays.

these measurable quantities we will refer to as '*features of the recoil*'. The issue is that the function to map these features to the recoil is not known and only possible to calculate analytically for 1 or at most 2 features. For example, in ATLAS' measurement of M_W they found the function analytically for two features [4]. Here it is discussed how it is possible to continue to estimate the recoil for many more features using gradient boosted decision trees.

CMS DETECTOR

Since 2009 the Compact Muon Solenoid (CMS) has been observing proton-proton collisions [8]. The detector consists of a 3.8 Tesla solenoid and 5 layers of detection [9]. The first layer is the trackers; these use silicon strips and silicon pixels to accurately track particles close to the interaction point. A small electrical signal from the silicon can be measured to identify the presence of a particle. The second layer is the electromagnetic calorimeter (ECAL) which uses scintillating $PbWO_4$ crystals to measure the energies of electrons and photons. Next in sequence is the hadronic calorimeter (HCAL); which measures the energy of hadrons using brass and steel interleaved with plastic scintillators. The fourth layer is the magnet; this is a 3.8T solenoid which bends the paths of charged particles emerging from collisions. The final layer

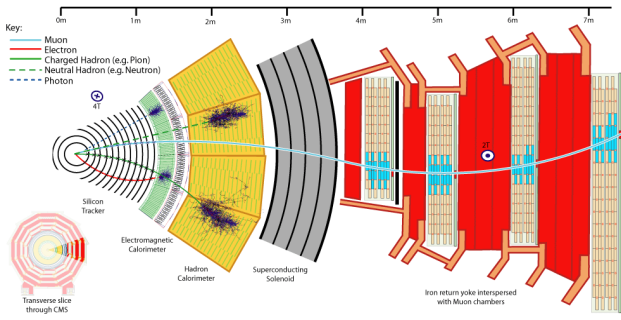


FIG. 2: Diagram showing the different components of the CMS detector [11]

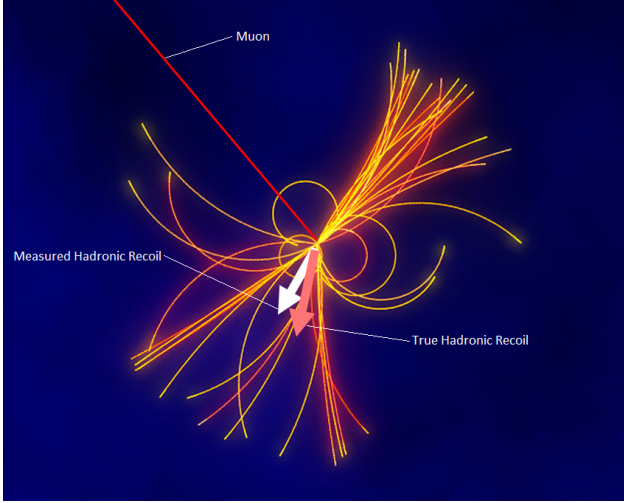


FIG. 3: Diagram showing a generic interaction. Diagram adapted from: [12]

is the muon detectors; these are comprised of drift tubes, cathode strip chambers and resistive plate chambers [9]. The structure of the CMS detector is shown in figure 2.

INITIAL ATTEMPTS AT CORRECTION

The diagram in figure 3 shows the need for a correction as the measured hadronic recoil is different from the true hadronic recoil.

Original attempts at a correction were performed using linear regression techniques. Two correction variables were created. One, e_1 , corresponds to the scale of difference between the measured hadronic recoil and the true hadronic recoil, whereas the other, e_2 , corresponds to the direction. e_1 and e_2 are event by event corrections, meaning that their values change depending on the event. But there is an intrinsic resolution in the correction. Plotting the intrinsic resolution against $\log_{ht}$ and sphericity, as in figure 4, shows that for low values of sphericity the value of σ is large. Plots showing the distributions of e_1 and e_2 are in the appendix.

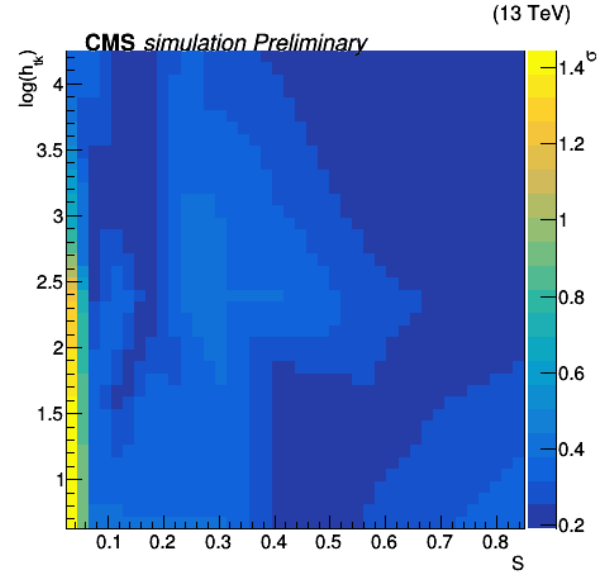


FIG. 4: 2D histogram showing how sigma varies for e_2 .

The logarithm of the scalar transverse momentum of an event, $\ln(h_{Tk})$, is here referred as $\log_{ht}$. Sphericity is the measure of how isotropic the jets are distributed from the center of an interaction point. It is defined as $S = \frac{3}{2}(\lambda_2 + \lambda_3)$ ($0 \leq S \leq 1$). Where $\lambda_1, \lambda_2, \lambda_3$ are the normalised eigenvalues of the full momentum tensor of the event [10].

In the case of the Z boson decay, shown in figure 1b, there are no experimental degrees of freedom, in other words it is possible to measure the hadronic recoil and allow us to determine whether our method for correcting simulation to data is working.

GRADIENT BOOSTING

To understand how decision trees work, one can imagine the case where a large amount of data on people is stored. And we wish to determine a likelihood that a certain person uses social media daily. By asking simple yes or no questions of the data, we can start to quantify how each question performs in predicting whether someone uses social media daily. Using these questions simple trees can be created. Figure 5 shows two examples of a tree and a number at each branch which quantifies the likelihood that answer corresponds to a person using social media daily. These simple trees are referred to as weak learners. By combining lots of weak learners, one can create a strong learner which perform much better. It is worth noting one doesn't have to ask yes or no questions as a tree can have more than two branches. In fact, typically in boosted decision trees we allow the structure of the strong learner to develop itself. Although we can specify things such as the number of trees, the maximum and minimum number of branches etc.

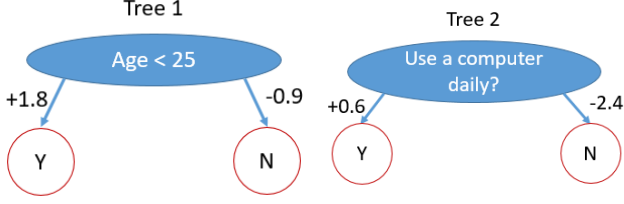


FIG. 5: Two examples of weak learner decision trees.

Specifically, we are using gradient boosted decision trees. A boosted decision tree is one that converges to a point using a loss function. A gradient boosted decision tree is one where the loss function is differentiable and we iteratively descend towards a minimum.

QUANTILE REGRESSION

During this project, two methods to calculate a recoil correction were explored; quantile regression and classifier training. Quantile regression is when one has two distributions and we wish to correct one distribution to the other, so we integrate both distributions then divide them up into quantiles. After dividing them up into quantiles a BDT (boosted decision tree) is used to correct each quantile. A graph demonstrating this is given in the appendix. The correction is generalised for many different features. A feature is a variable of the data set that we wish to use in the training; examples include: the energy density, ρ , the transverse momentum, p_T or the sphericity etc.

The software package used in to provide the gradient boosted decision tree algorithm is from Scikit-Learn [13].

CLASSIFIER TRAINING

Classifier training is when there are two distributions and a re-weighting for the probability distribution function is found so that the two distributions will match. Then this is generalised for all the features using a BDT with classifier training (a graph demonstrating this is given in the appendix). To do that the simulation and data are shuffled together and assigned either a 1 or a 0, the training then runs and it attempts to work out what it is about certain events that makes it either simulation-like or data-like. In our case this was performed twice. The first classifier training was used to correct the kinematics. This means that the first features used were p_T and η . Allowing a set of probability weights to be used to correct the simulation so that the kinematics in both the simulation and data were the same. From then it was possible to run training for many more features using the corrected simulation data. The same software package was used to provide the classifier training algorithm [14].

The loss function used in this decision tree is a cross-entropy loss function. The cross entropy between two dis-

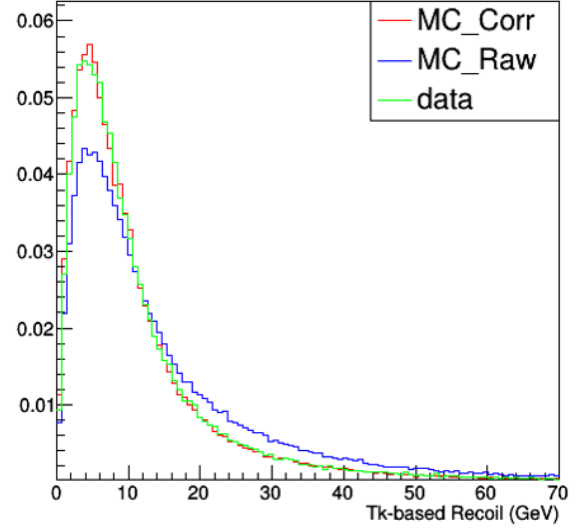


FIG. 6: Graph showing the overall recoil distribution following quantile regression training. The raw MC, the detector data and the corrected MC are displayed.

tributions is the average amount of information required to identify random events from the distribution [15].

It is possible that the quantile regression method will simply learn the function. As opposed to learning how each feature needs to be corrected in order to match simulation to detector data. This is not particularly useful, even if the result will match the data. However, when using the classifier training we correct the kinematics first. We make sure that the kinematics match the detector data and simulation. Then we can correct the features with the knowledge that we are correcting one data set to another and the kinematics for both are the same. For this reason it is speculated that classifier training is a better method in principle.

RESULTS

In collecting the results there are two important things to look at when quantifying the success of a correction. One must consider the overall recoil distribution to see if the correction has worked to bring the raw MC data closer to the detector data. And one must look at the correction of individual features of the correction by plotting the profile of a 2D histogram of a feature against the track p_T .

The results here show the training and correction being applied to Z bosons decay into two leptons. In order to ascertain whether the machine learning was finding a sensible correction for the hadronic recoil, it is possible to use the Z boson decay, as the W and Z bosons are both carriers of the weak interaction and have similar masses.

Figure 6 shows the results after quantile regression (the overall recoil distribution). The corrected MC simulation is overlaying the detector data; the p-value for this is 10.1% and

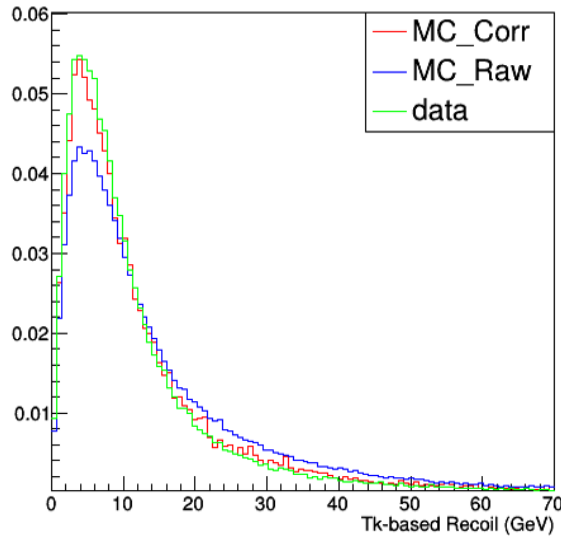


FIG. 7: Graph showing the overall recoil distribution following classifier training. The raw MC, the detector data and the corrected MC are displayed.

the χ^2 is 70.9. The features used in this quantile regression are $Z_{PT}(\text{vis_pt})$ and $Z_\eta(\text{vis_eta})$.

Figure 7 shows the the overall recoil distribution after classifier training. This time a comparison of the corrected MC simulation and the detector data yeilds a p-value of 0% and a χ^2 of 167. The features used in this quantile regression are: (tkmet_logpt, ntnpv_logpt, npvmet_logpt, tkmet_phi, tkmet_n, tkmet_sphericity, ntnpv_sphericity, npvmet_sphericity, absdphi_ntnpv_tk, dphi_puppi_tk, rho, nvert, mindz, vz, nJets).

Figure 8 shows the results after the same quantile regression training that was performed in figure 6. Here, however, we display the profile for the vertex multiplicity (nvert) median. Figure 9 shows the 68% confidence level of the same 2D histogram. Both appear to show a strong correlation between corrected MC and the detector data. It is sensible to recognise that values for high vertex multiplicity are less significant as there are far fewer events in the histograms at these high values; for this reason we need not worry about the fact that the corrected MC and the detector data lines do not match as well. Note that the mean is flat for vertex multiplicity which implies that there it has no correlation to the recoil.

Figure 10 shows the results after the same classifier training that was performed in figure 7. As before, however, we show the profile for the vertex multiplicity (nvert) median and the 68% confidence level is shown in figure 11. The corrected MC doesn't overlay the detector data as well as it does for the quantile regression training. Nonetheless, the correction does appear to go in the right direction.

A similar analysis is displayed for the feature ρ for the same training's: in figure 12 the profile of the median are shown for quantile regression; in figure 13 the profile of the 68% confi-

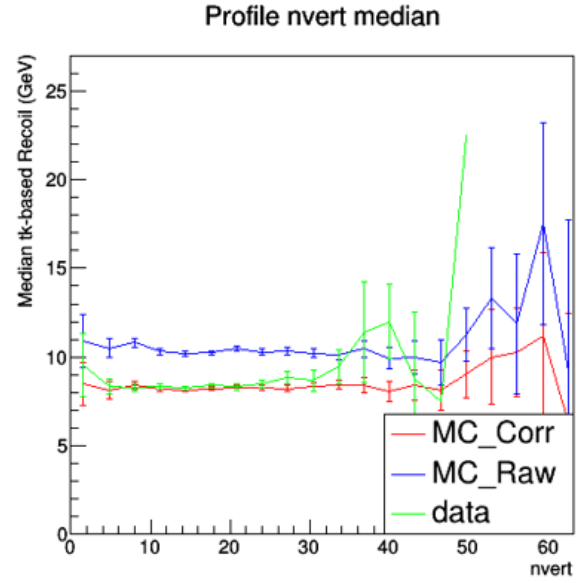


FIG. 8: Graph showing the profile of the median of a 2D histogram of vertex multiplicity against track p_T . Monte Carlo correction performed by quantile regression training.

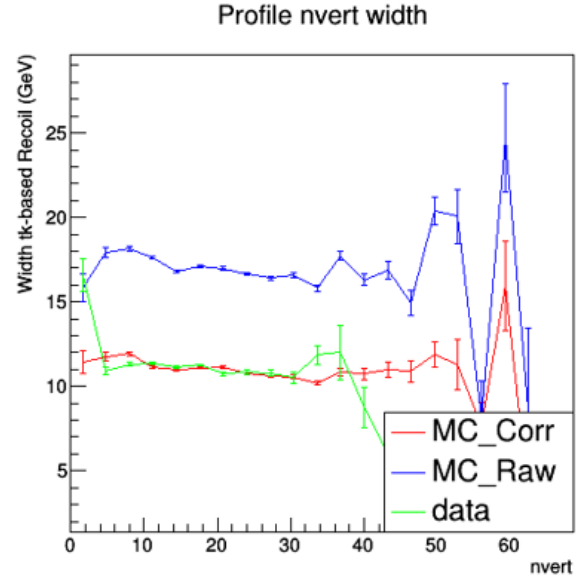


FIG. 9: Graph showing the profile of the 68% confidence level of a 2D histogram of vertex multiplicity against track p_T . Monte Carlo correction performed by quantile regression training.

dence level are shown for quantile regression; in figure 14 the profile of the median are shown for classifier training; and in figure 12 the profile of the 68% confidence level are shown for classifier training. Note the gradient in figures 12 and 14; these imply that ρ is correlated to the recoil. The results here highlight the same story that the quantile regression gives rise to a better corrector than classifier training.

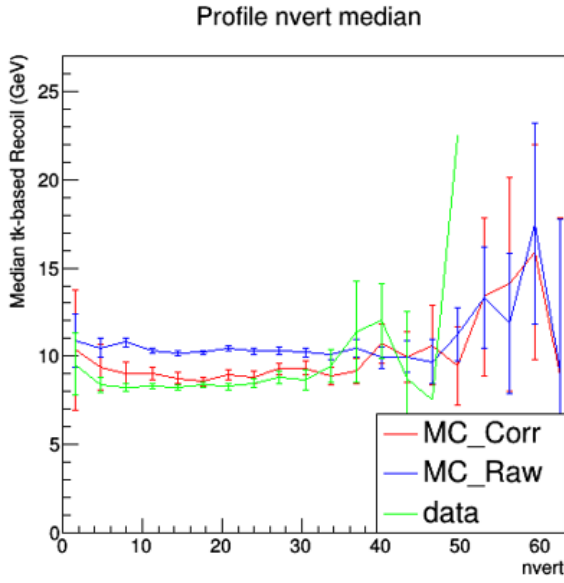


FIG. 10: Graph showing the profile of the median of a 2D histogram of vertex multiplicity against track p_T . Monte Carlo correction performed by classifier training.

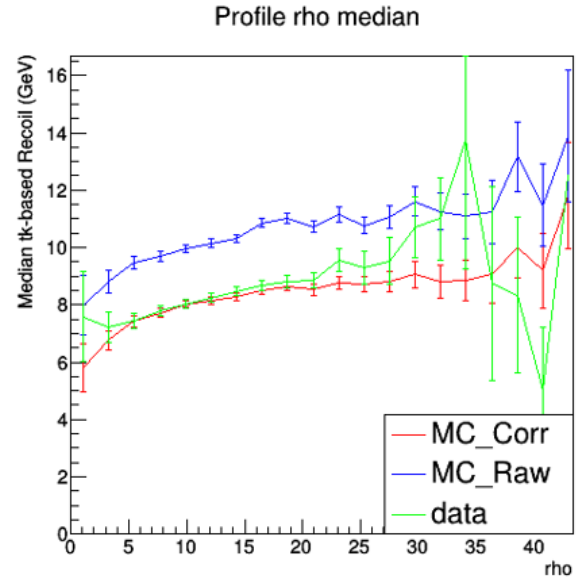


FIG. 12: Graph showing the profile of the median of a 2D histogram of ρ against track p_T . Monte Carlo correction performed by quantile regression training.

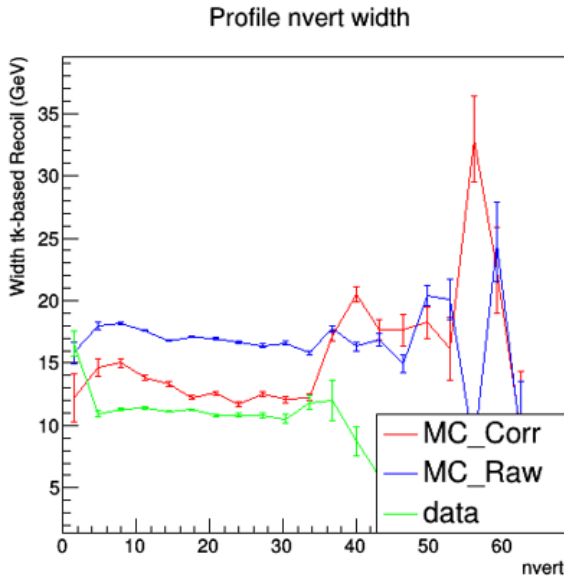


FIG. 11: Graph showing the profile of the 68% confidence level of a 2D histogram of vertex multiplicity against track p_T . Monte Carlo correction performed by classifier training.

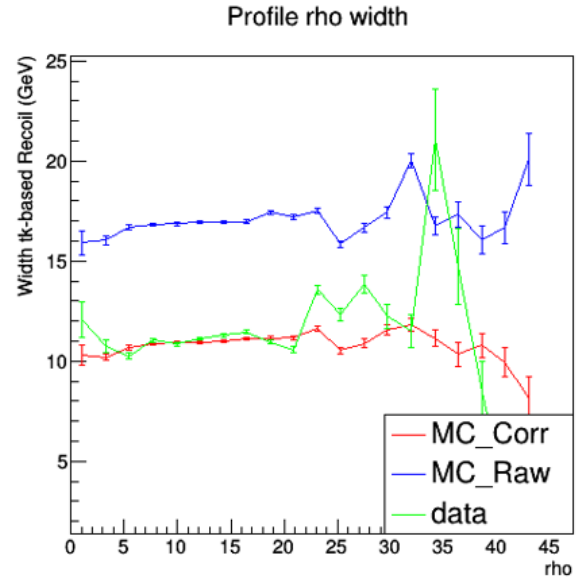


FIG. 13: Graph showing the profile of the 68% confidence level of a 2D histogram of ρ against track p_T . Monte Carlo correction performed by quantile regression training.

Earlier in this paper it was stated that the classifier training method was better in principle, yet, in practise, it didn't perform as well as the quantile regression method. There are two possible reasons for this. The discrepancy here could be due to the kinematics weights not being applied correctly. Analysing the weights applied for the kinematics correction showed that they did not have a median of 1, which is an issue. Another reason may be that the best possible parameters to apply to the

boosted decision tree learning have not yet been found as this is a work in progress.

In the appendix graphs showing the difference between the different data taking eras show that it is acceptable to use the same method on all of the different data taking eras. Also in the appendix graphs showing a comparison of the W and Z boson corrected MC are evidence that the same training used on Z decay data can also be applied to W boson decay data.

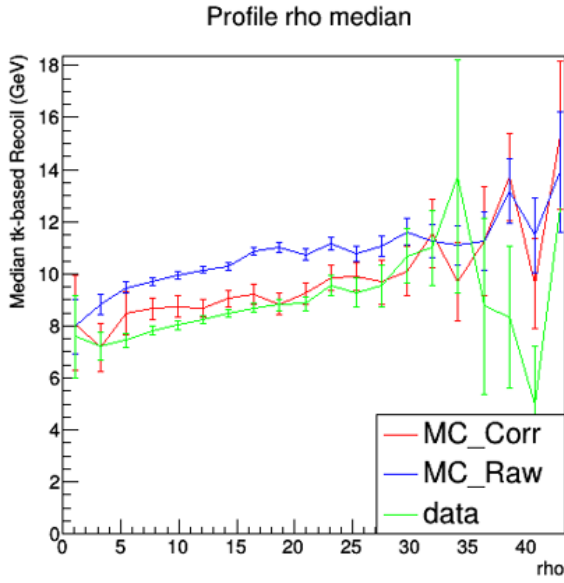


FIG. 14: Graph showing the profile of the median of a 2D histogram of ρ against track p_T . Monte Carlo correction performed by classifier training.

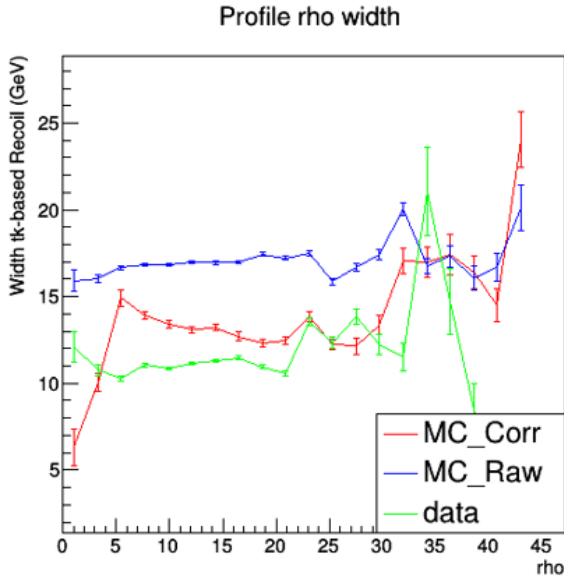


FIG. 15: Graph showing the profile of the 68% confidence level of a 2D histogram of ρ against track p_T . Monte Carlo correction performed by classifier training.

CONCLUSIONS

It was found that the difference between data and corrected Monte Carlo for quantile regression is very small; however, this method may be incorrect as the kinematics have not been correctly dealt with. The difference between data and corrected Monte Carlo for the classifier training method is larger

than when using quantile regression; however, this could be due to an issue with correcting the kinematics. The next step would be to solve this issue and to find the optimal parameters for classifier training. Then in depth analysis of this method applied to W boson decays would allow the hadronic recoil estimator to be calculated.

ACKNOWLEDGEMENTS

I would like to give a special thank you to my supervisor Pedro Da Silva for his enormous help during this project and patience in explaining ideas. The inspiration for this project came from work by Nicolo Foppiani and Olmo Cerri and their doctoral supervisor Luigi (Gigi) Rolandi. I was lucky to receive help from Pasquale Musella who has a breadth of experience in using decision trees in this way. I massively enjoyed my project and my time at CERN and wish to thank the summer student team for providing this opportunity.

REFERENCES

- [1] UA1 Collaboration, G. Arnison et al, "Experimental observation of isolated large transverse energy electrons with associated missing energy at $s = 540$ GeV". Phys. Lett. B 122, 103116, (1983)
- [2] A. Salam, J.C. Ward, "Electromagnetic and weak interactions", Phys. Lett. 13, 168171, (1964)
- [3] M. Baak et al, "The global electroweak fit at NNLO and prospects for the LHC and ILC", Eur. Phys. J. C 74, 3046, (2014)
- [4] ATLAS Collaboration, "Measurement of the W-boson mass in pp collisions at $s = 7$ TeV with the ATLAS detector", Eur. Phys. J. C, 78:110, (2018)
- [5] N. Foppiani, L. Rolandi, "Definition and calibration of the hadronic recoil in view of the measurement of the W boson mass with the CMS experiment", CERN-THESIS-2017-125, (2017)
- [6] O. Cerri, L. Rolandi, "Hadronic recoil in the W boson production at LHC for a W mass measurement with the CMS experiment", CERN-THESIS-2017-157, (2017)
- [7] http://kjende.web.cern.ch/kjende/en/wpath_hcphysics2.htm
- [8] <https://home.cern/scientists/updates/2013/02/lhc-report-run-1-final-flurry>, (2013)
- [9] CMS Collaboration, D. Acosta, et al, "CMS Physics Technical Design Report Volume I: Detector Performance and Software", CERN/LHCC 2006-001, (2006)
- [10] ATLAS Collaboration, "Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at $s = 7$ TeV", CERN-PH-EP-2012-119, (2012)
- [11] D. Barney, et al, "Interactive Slice of the CMS detector", CMS-doc-4172-v2, (2004)
- [12] <http://ivonisio.blogspot.com/2012/07/a-particula-de-deus.html>, (2012)
- [13] <http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html>
- [14] <http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>
- [15] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", MIT Press, (2016)

APPENDIX

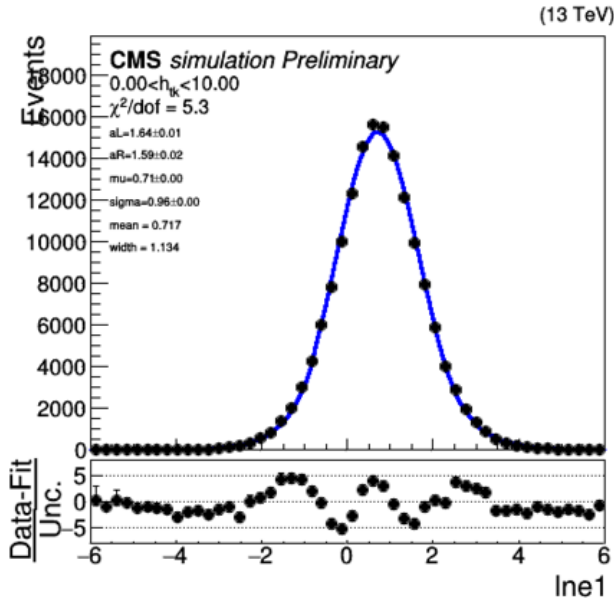


FIG. 16: Graph showing $e1$ as originally used with a Gaussian fit in the centre and a exponential fits for the tails.

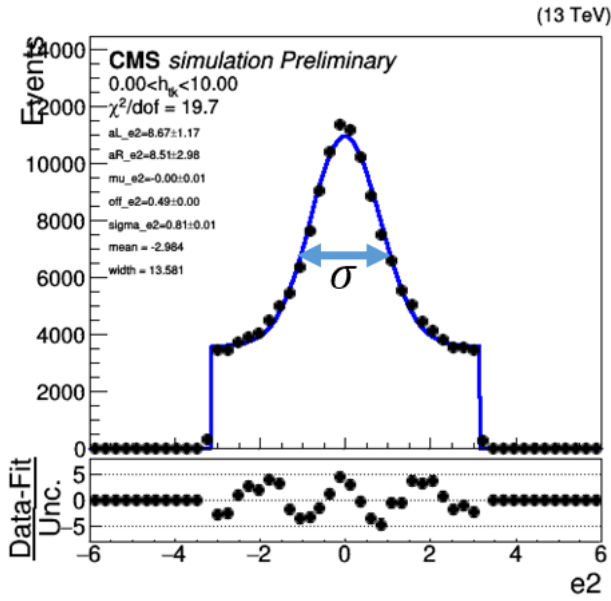


FIG. 17: Graph showing $e2$ as originally used with a Gaussian fit in the centre and a exponential fits for the tails and an offset.

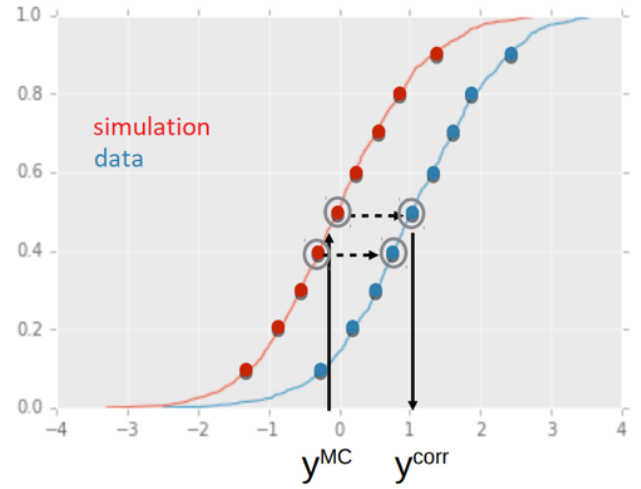


FIG. 18: Graph demonstrating how quantile regression is performed.

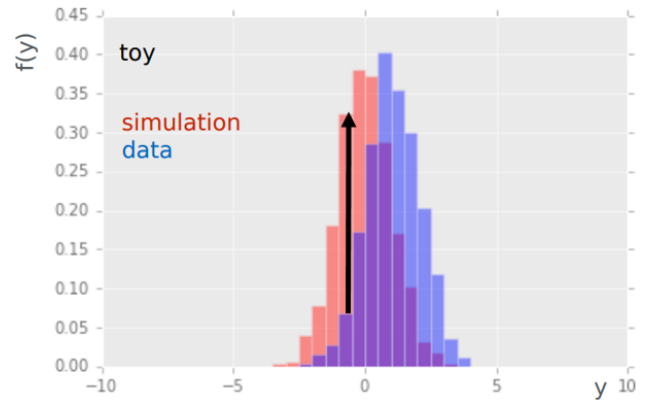


FIG. 19: Graph showing how classifier training is performed.

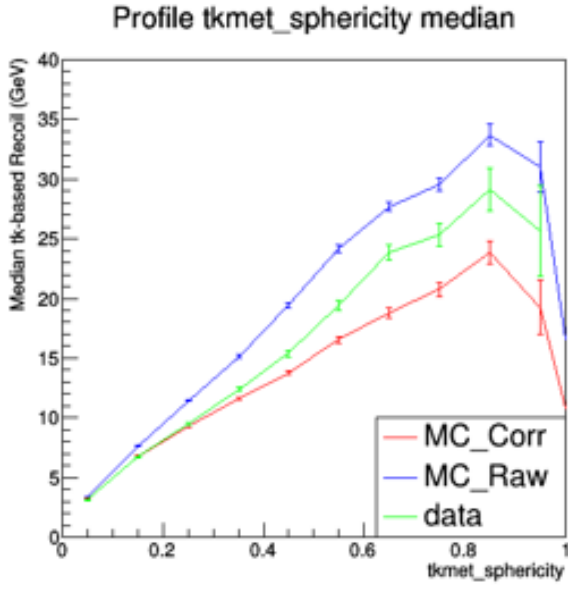


FIG. 20: Graph showing the profile of the median for sphericity. Monte Carlo correction performed by quantile regression training.

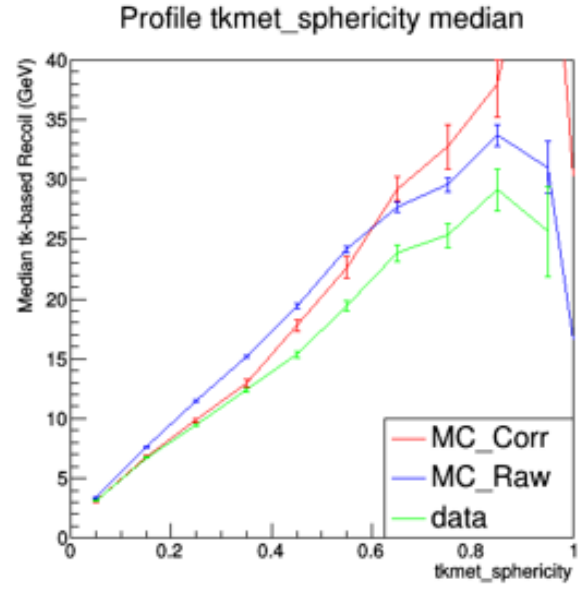


FIG. 22: Graph showing the profile of the median for sphericity. Monte Carlo correction performed by classifier training.

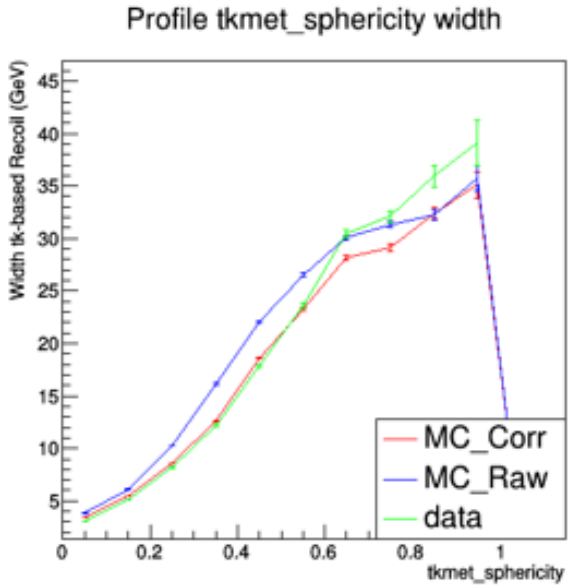


FIG. 21: Graph showing the profile of the 68% confidence level for sphericity. Monte Carlo correction performed by quantile regression training.

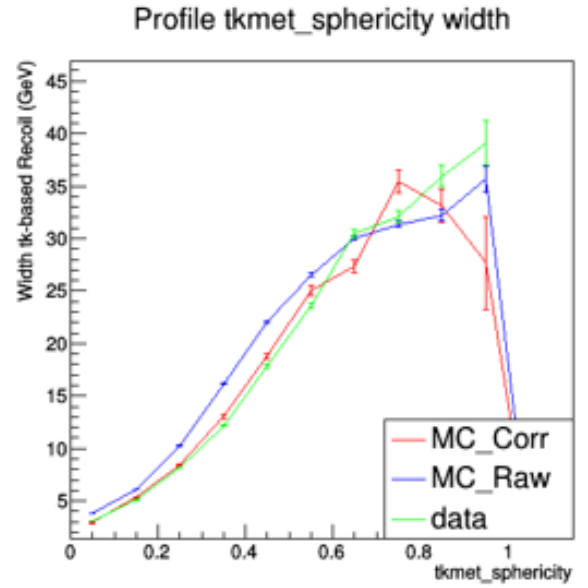


FIG. 23: Graph showing the profile of the 68% confidence level for sphericity. Monte Carlo correction performed by classifier training.

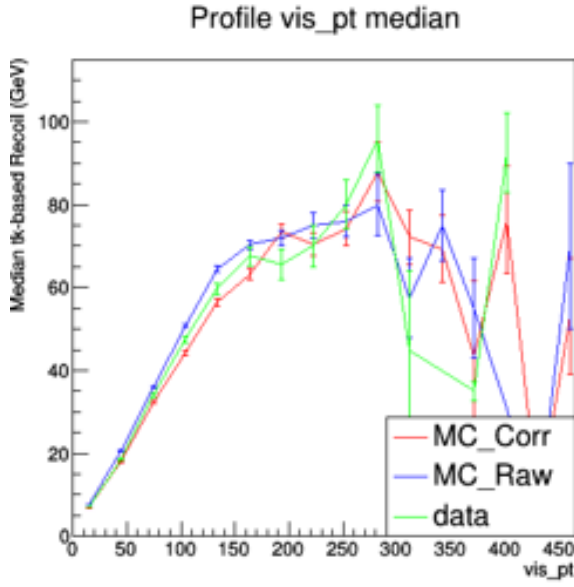


FIG. 24: Graph showing the profile of the median for Z_{pT} . Monte Carlo correction performed by quantile regression training.

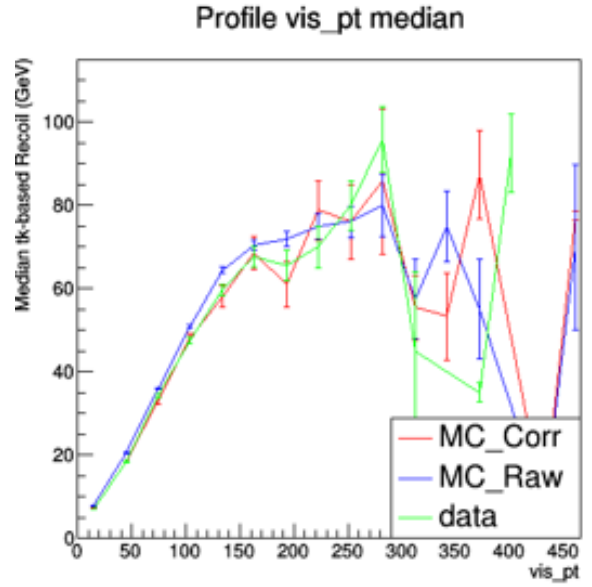


FIG. 26: Graph showing the profile of the median for Z_{pT} . Monte Carlo correction performed by classifier training.

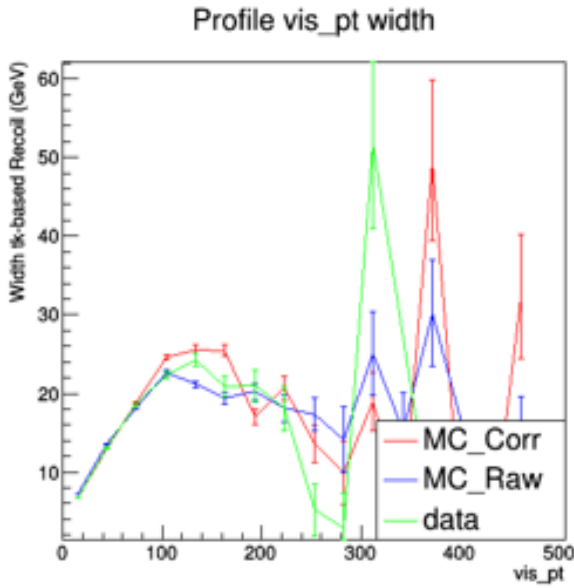


FIG. 25: Graph showing the profile of the 68% confidence level for Z_{pT} . Monte Carlo correction performed by quantile regression training.

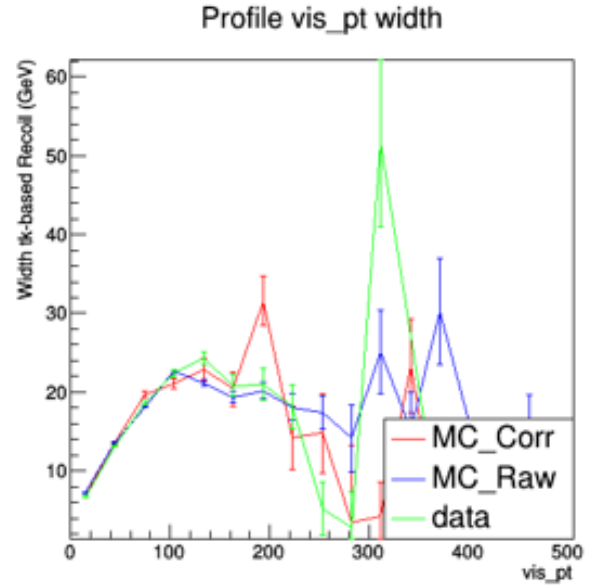


FIG. 27: Graph showing the profile of the 68% confidence level for Z_{pT} . Monte Carlo correction performed by classifier training.

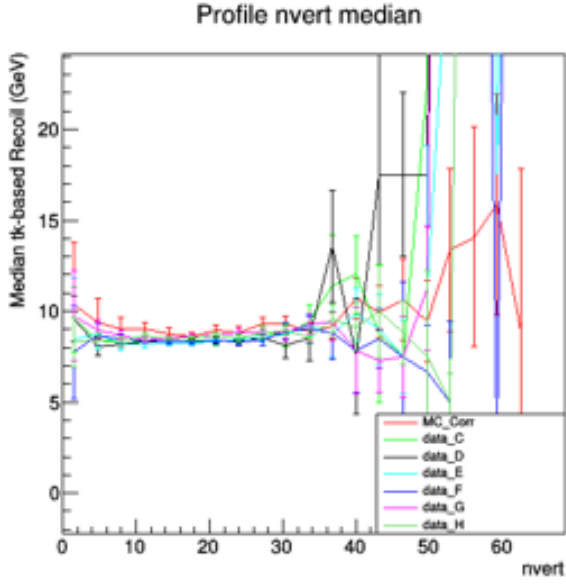


FIG. 28: Graph showing the difference between data taking eras.

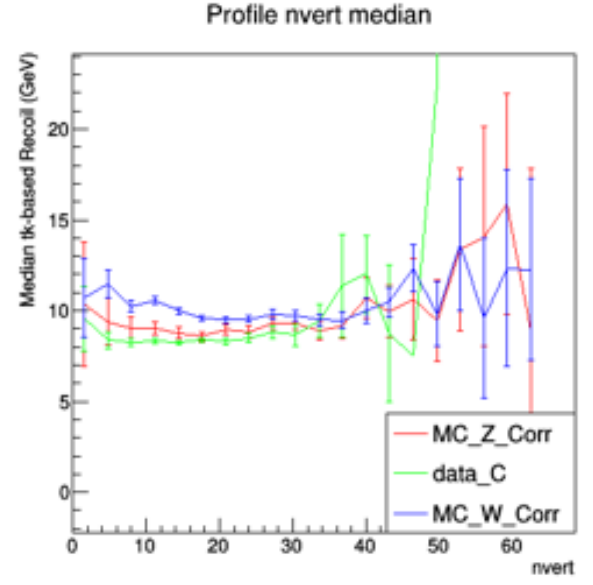


FIG. 30: Graph showing the profile of the median for 2D histogram of the vertex multiplicity plotted against track p_T with the same classifier training being applied to W boson data and Z boson data. Again, the Z boson expectedly performs better but this is proof of concept that the same method will work for the W boson.

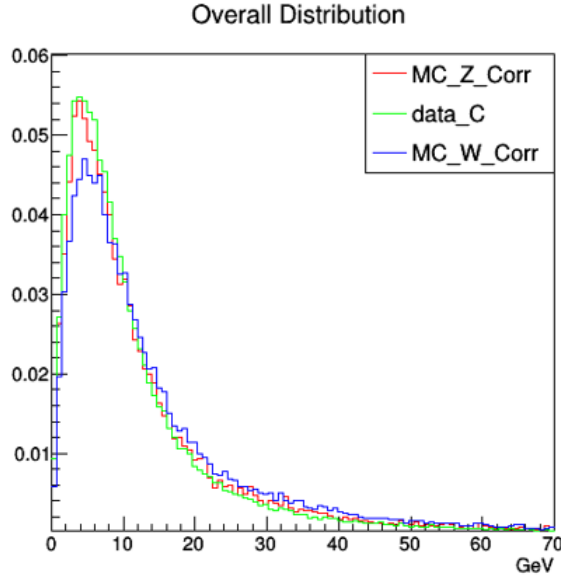


FIG. 29: Graph showing the overall recoil distribution with the same classifier training being applied to W boson data and Z boson data. The Z boson expectedly performs better but this is proof of concept that the same method will work for the W boson.