MEPhI
National Research
Nuclear University

Department of
elementary particle
physics

The 7th International Conference on Particle Physics and Astrophysics (ICPPA-2024)

Separation of QCD and EWK processes of vector diboson production using machine learning algorithms with third-jet information.

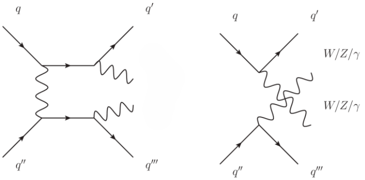
October 25, 2024

Dmitriy Zubov

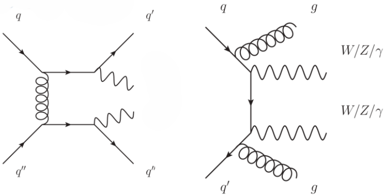
Introduction

- The production of vector boson pairs (VV, with $V = W/Z/\gamma$) provides an opportunity to perform precision studies of the electroweak sector of the Standard Model (SM), as well as the search for new physics beyond it.
- In the SM, VV may be produced at lowest order via quark-antiquark annihilation, as well as through gluon-gluon fusion via a quark loop.
- To higher order, the VV can be also produced via the vector boson scattering (VBS) process, which is crucial for probing the mechanism of electroweak symmetry breaking (EWSB) in the SM.
- VBS processes are studied by measuring the electroweak production of $VVjj$, with the corresponding QCD production of the final state $VVjj$ being the dominant irreducible background.

EWK signal



QCD background

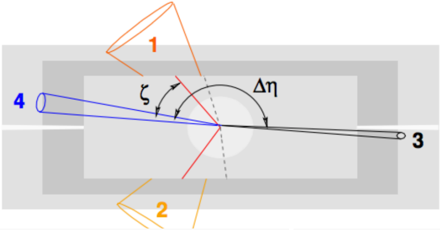


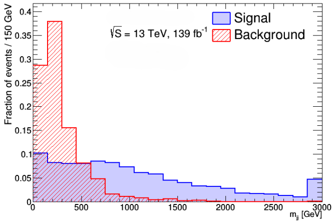
$ZZjj \to \ell\ell\nu\nu jj$

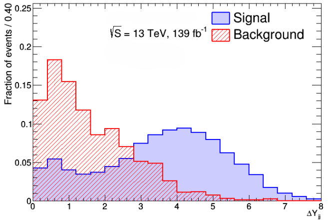
- Among all processes related to vector-boson scattering, the electroweak production of two jets and a Z-boson pair is a rare and important one.
- VBS $ZZjj$ production is uniquely sensitive to the possible anomalous interaction between four Z bosons.
  - This is forbidden at tree-level in the SM and the study of EW $ZZjj$ production is therefore a direct test of an important prediction of the electroweak theory.
- $ZZ \to \ell^-\ell^+\nu\bar{\nu}$ has a higher branching ratio than $ZZ \to \ell^-\ell^+\ell'^-\ell'^+$, but has a higher background contamination.
  - Smaller cross section precision, better sensitivity for BSM processes at extreme values of Z boson p_T.

Typical VBS topology

- Tagging jets:
  - Invariant mass — M_{jj}
  - Rapidity difference — ΔY_{jj}
- Centrality $\zeta_i = \left| \frac{y_i - \frac{y(j_1)+y(j_2)}{2}}{y(j_1)-y(j_2)} \right|$
- p_T-balance$= \frac{|\mathbf{p}_T^{\text{miss}} + \mathbf{p}_T^{Z} + \mathbf{p}_T^{j1} + \mathbf{p}_T^{j2}|}{E_T^{\text{miss}} + E_T^{Z} + p_T^{j1} + p_T^{j2}}$





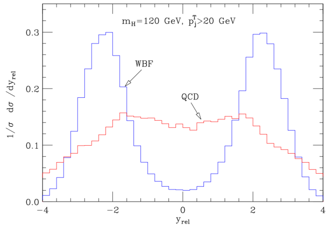


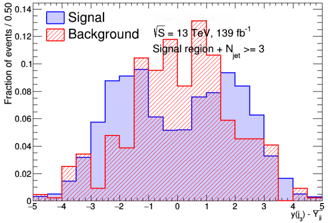
Kinematics of the third jet

Theoretical works predicted the suppression of the third jet in the central region of the detector for VBS/VBF processes compared to QCD.
[Del Duca, Frizzo, Maltoni, JHEP 05 (2004) 064]

- QCD events have higher effective scale and thus produce harder radiation than EWK.

Because of the cross section difference between the QCD and EWK processes of more than three orders of magnitude, as well as hadronisation and detector effects, direct constraints on such variables are of little effect.





Application of machine learning algorithms

The use of machine learning (ML) algorithms can improve the separation of signal events from background in cases of:

- Multiparticle final state
- The distributions of potential-separating variables overlap significantly for signal and background

This study compared the performance of ML algorithms with and without the inclusion of third jet kinematic variables.

- When training an algorithm incorporating third jet variables, the dataset was split into two categories — N_{jets}=2 and $N_{\text{jets}} \geq$3

Decision trees with gradient boosting (BDTG) based on TMVA were used as a classifier.
Signal significance was used as a metric for the final evaluation of signal-background separation:

$$Z = \sqrt{2 \times [(S + B) \times \ln(1 + (S/B)) - S]}$$

Dataset used for the study

Process: $pp \to Z(\to \ell\ell)Z(\to \nu\nu)jj$

Data:

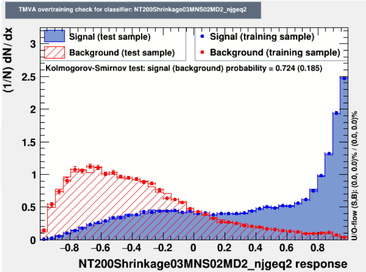
- MadGraph + Pythia8 + Delphes (with ATLAS card)
- $\sqrt{s} = 13$ TeV
- Normalized for L = 140 fb^{-1}

Final state:

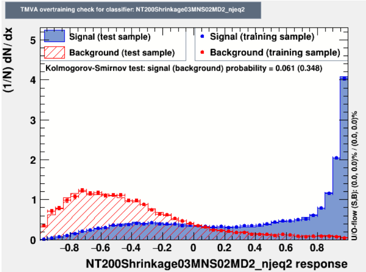
- $\geq$2 hadron jets, p_{T}>30 GeV
- Same flavour oposite charge lepton pair (e^+e^- or $\mu^+\mu^-$), leading p_{T}>30 GeV, subleading p_{T}>20 GeV
- $E_T^{miss} > 70$ GeV
- Veto on any additional lepton

Learning Results

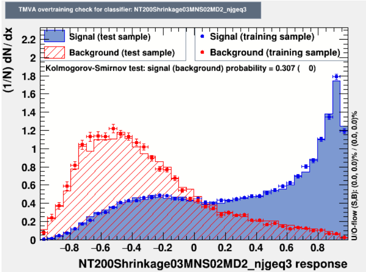
$N_{\text{jets}} \geq 2$



$N_{\text{jets}} = 2$



$N_{\text{jets}} \geq 3$

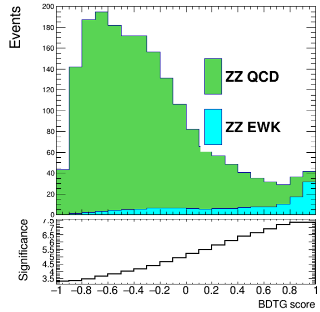


Distributions of training and test samples for signal and background on the classifier response variable.

The results of the second and third classifier were combined and compared with the results of the first classifier.

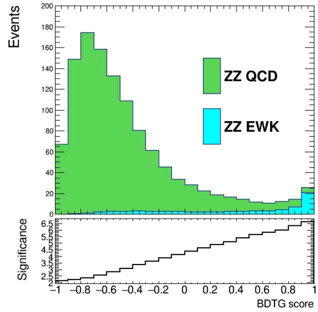
Learning Results

$N_{\text{jets}} \geq 2$



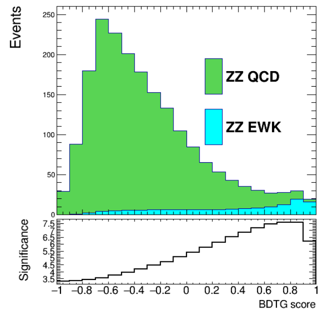
- Max. Z = 7.36±0.03
- S. = 31.4±0.4
- B. = 10.2±0.3

$N_{\text{jets}} = 2$



- Max. Z = 6.32±0.02
- S. = 28.1±0.3
- B. = 12.4±0.1

$N_{\text{jets}} \geq 3$



- Max. Z = 7.56±0.03
- S. = 35.9±0.4
- B. = 13.2±0.3

Cross- sectional and luminosity normalised distributions of signal and background on the classifier response.
And also the dependence of signal significance on the lower threshold on the classifier response.

After combining the results of the second and third classifiers:

- Total Z = 9.96±0.04; Total Signal = 56.8±0.5; Total Background = 18.1±0.3

Conclusion

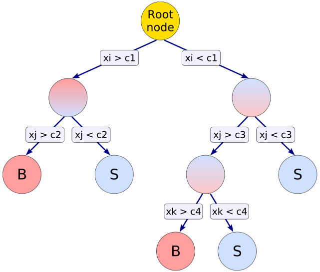
- In this study we have investigated the possibility of enhancing the separation of EWK and QCD processes of Z boson pair production.
- The use of third jet information in the machine learning algorithm increased the signalling significance.
- It is possible to use the described approach when studying other VBS/VBF processes.

Back up slides

Decision trees with gradient boosting (BDTG)

Decision tree

- A decision tree is a binary tree : a sequence of cuts paving the phase-space of the input variables
- Repeated yes/no decisions on each variables are taken for an event until a stop criterion is fulfilled
- Trained to maximize the purity of signal nodes (or the impurity of background nodes)



Advantages:

- Decision trees are independent of monotonous variable transformations
- Weak variables are ignored and do not deteriorate performance

Disadvantages:

- Decision trees are extremely sensitive to the training samples, therefore to overtraining
- Slightly different training samples can lead to radically different DT

Boosting

- Sequentially apply the DT algorithm to reweighted (boosted) versions of the training data
- Each model in the series trains upon its predecessor's mistakes, trying to correct them
- Works very well on non-optimal decision tree (small number of nodes)
- There are different boosting algorithms and in our work we use the gradient descent

Studied variables

- $Y(Z)$
- $p_{\mathrm{T}}(Z)$
- $\varphi(Z)$
- $\Delta\phi(\mathbf{E}_T^{miss}, \mathbf{p}_T^{ll})$
- E_T^{miss} significance
- E_T^{miss}/H_T
- E_T^{miss}
- p_{T}(subleading jet) (j_2)
- η(subleading jet)
- φ(subleading jet)
- p_{T}(leading jet) (j_1)
- η(leading jet)
- φ(leading jet)
- $M(j_1 j_2)$
- p_T-balance
- $\zeta(Z)$

- $\Delta Y(j_1 j_2)$
- $\Delta Y(j_1 Z)$
- $\Delta Y(j_2 Z)$
- $\Delta\varphi(j_1 Z)$
- $\Delta\varphi(j_2 Z)$

3rd jet information variables:

- $\Delta Y(j_1 j_3)$
- $\Delta Y(j_2 j_3)$
- $\Delta Y(j_3 Z)$
- $M(j_1 j_3)$
- $M(j_2 j_3)$
- p_{T}(third jet) (j_3)
- η(third jet)
- φ(third jet)
- $\zeta(j_3)$

3rd jet variables distributions