Search for diboson production in lepton+MET+$b\bar{b}$ channel

M Trovato and C Vernieri

1INFN-Pisa, Polo Fibonacci Largo B. Pontecorvo, 3 - 56127 Pisa, Italy
E-mail: mtrovato@FNAL.GOV, cvernier@FNAL.GOV

Abstract. The study of associated boson production in the final state with a lepton, neutrino, and a $b\bar{b}$-pair is important since the event topology of this process is the same as expected for associated production of a $W$ and the Standard Model light-Higgs boson. Thus, the search for boson production can be considered as a preliminary step towards the Higgs discovery. Here we present a search for $WZ/ZZ$ in events with a lepton, missing transverse energy, and $b$-quark jets. Besides looking at the sample where two exclusive jets are found, we investigate the sample with 3 jets which contains 40% of the signal events. In this sample the invariant mass of the two $E_{T}$-leading jets would normally be chosen to signal the $Z$ boson. We describe an alternative procedure to reconstruct the $Z$-invariant mass in this sample. This procedure exploits the information carried by the third jet in the sample.

1. Introduction

Measurements of diboson production at the Tevatron have been used to test of the electroweak sector of the Standard Model (SM) and to place limits on models of physics beyond the SM [1]. Diboson measurements are also instrumental for searches for the SM light Higgs boson at the Tevatron. By choosing to focus on the final state where a $Z$-boson decays into $b\bar{b}$-pairs, the event topology would be the same as expected for associated production of a $W$ and a light Higgs boson ($m_{H} < 135 \text{ GeV/c}^{2}$). At the Tevatron, the process $WH \rightarrow Wb\bar{b}$ has an expected $\sigma \cdot BR$ about five times lower than $WH \rightarrow Wb\bar{b}$ for $m_{H} \simeq 120 \text{ GeV/c}^{2}$. Therefore, observing this process would be a benchmark for the even more difficult search for the light Higgs in the $WH \rightarrow Wb\bar{b}$ process.

This paper describes a preliminary study towards improving the signal acceptance in the cross-section measurement of $WZ$ and $ZZ$ associated productions in a final-state with a lepton, missing transverse energy, and up to three jets, where two jets will be required to carry $b$-flavour ("$b$-tagged jets"). A first mandatory step of such a measurement is to validate how the simulations models the data on the aforementioned final-state sample, prior to any $b$-tagging requirements. This sample, named "pretag sample", benefits a larger statistics, and is therefore sensitive even to small mis-modelings. In Sec. 2 simulations will be compared to data in the pretag sample, where we apply the standard kinematical cut at CDF which requires the number of jets to be exclusively two. However, simulations show that only about half of the signal events pass this exclusive two-jets cut. In Sec. 3 we will investigate the sample with three jets in the final state. The problem of how best to reconstruct the correct $Z$-boson mass will be addressed.

1 $BR$ being the Branching ratios of $Z/H \rightarrow b\bar{b}$

Published under licence by IOP Publishing Ltd
2. Two jets pretag sample
In order to spot possible mis-modelings in our simulations the pretag sample is investigated.

2.1. Sample selection
We investigate the finale state in which a W or Z boson decays leptonically and the other Z-boson hadronically. The experimental signature involves the presence of a charged lepton (electron or muon) a neutrino (identified through the missing transverse energy, $E_{T,\text{miss}}$) and large-$E_T$ jets. The jet $E_T$ is corrected for known instrumental and physics effects. The sample we investigate is selected by applying the following cuts:

- exactly two jets with $E_T > 15$ GeV and $|\eta| < 2$
- an isolated triggered electron with $|\eta| < 1.1$ and $E_T > 20$ GeV
- or muon with $|\eta| < 1.$ and $P_T > 20$ GeV
- $E_{T,\text{miss}} > 20$ GeV
- $E_{T,j1,j2} > 25, 15$ GeV and $|\eta_{j1,j2}| < 2$; $j1$, $j2$ being the first and second leading $E_T$-jets
- $M_W^W > 10$ (30) GeV if the triggered lepton is a muon (electron), $M_W^W$ being the W-invariant mass in the transverse plane
- $E_{T,\text{miss}}$-significance > 1.8 if the triggered lepton is an electron. The definition of $E_{T,\text{miss}}$-significance is given in [3]

2.2. SM processes composition of the selected sample
The following processes would contribute to a data sample selected sample within our cuts:

- **Electroweak and top (EW):** $WW$, $WZ$, $ZZ$, $Z+jets$, $t\bar{t}$, single-top. Each of these processes can mimick the signal signature, with one detected lepton, large $E_{T,\text{miss}}$, and two jets. The contamination of these processes in the selected data sample is estimated by using their accurately predicted cross sections [4]. The shapes (templates) of a number of observables are obtained from ALPGEN+Pythia, Pythia MC after the simulation of the CDF detector.

- $W(\rightarrow l\nu)+jets$, $l = e, \mu, \tau$. Due to the presence of real leptons and neutrinos, the $W+jets$ background is the hardest to be reduced. Templates are obtained from ALPGEN+Pythia MC, while the rate normalization is obtained from data (see Fig.1)

- **QCD:** multi-jet production with a jet faking the lepton and fake $E_{T,\text{miss}}$. Since the mechanism for a jet faking a lepton or for fake missing transverse energy is not expected to be well modeled in MC events both rate normalization (see Fig.1) and templates are obtained from data. The way howackground shapes are obtained for QCD events is explained elsewhere [7]

2.3. Shape validation
In this section we compare a number of significant observables to data and expectations (Fig.2, 3, 4) in order to validate the used MC, the QCD templates, and the expected rates in the pretag sample.

We see some mis-modeling of the data, in the lowest jet $E_T$ bins. We are planning to investigate how other generators for $W+jets$ events would model our data. Other QDC models are under investigation as well.
Figure 1. Fit on $M^W_T$ for the QCD and $W$+jets rate in data collected with the high-$P_T$ muon trigger. The dotted curve show the result of the fit.

Table 1. Predicted and observed number of events in the pretag sample. $W$+jets and QCD rates are estimated from the fitting data. The expected rates are separated for different triggered lepton type. By construction the expected number are equal to observed numbers. The calculation of uncertainties on these rates is still on-going.

| Process       | Rate (Electrons) | Rate (Muons) |
|---------------|------------------|--------------|
| Signal($WZ+ZZ$) | 125.7            | 120.8        |
| $WW$          | 745.2            | 544.6        |
| $EW$          | 1213.7           | 2062.9       |
| $W$ + $jets$  | 14874.9          | 12353.7      |
| QCD           | 1937.0           | 216.8        |
| Total Observed| 18990            | 15497        |

Figure 2. Transverse energy of the leading $E_T$-jet when the triggered lepton is a muon (a) or a central electron (b).
3. Jet studies on the $WZ\to \ell\nu3j$ sample

We would like to select a signal sample among these candidates, and at the same time preserve a fair $Z$ mass resolution, by making use of the information carried by all the jets in the event. In previous studies of the $WZ$ process at CDF the $Z$ mass was defined as $M_{j_1j_2}$, $j_1$, $j_2$ being the two leading-$E_T$ jets. In the sample where 3 jets are found $M_{j_1j_2}$ has a degraded resolution (Fig.5).

3.1. Sample selection

Besides the selection reported in Sec. 2.1 we allow for a third jet with $E_T > 15$ GeV and $|\eta| < 2$.

3.2. Adopted strategy

Improving the resolution in such a sample means to choose the correct jet combination coming from $Z$ for building the $Z$ mass. Our goal is to investigate the origin of the additional jet, which may be initiated by gluon(s) radiated from the interacting partons (ISR) or from the $Z$-decay products (FSR). In order to do that:

- Jets are ordered in decreasing $E_T$: $j_1$, $j_2$, $j_3$
- Jets are matched in direction to quarks from $Z$ decay ("MJ")
• When a number of matches different from 2 is found the event is not considered
• Investigate at generator level the origin of the not-matched jet ("NMJ") in order to determine the right jet combination ("RJC")
  (i) NMJ = j3 is from ISR $\rightarrow$ RJC = j1j2 - 40% of events
  (ii) NMJ = j2 is from ISR $\rightarrow$ RJC = j1j3 - 18% of events
  (iii) NMJ = j1 is from ISR $\rightarrow$ RJC = j2j3 - 9% of events
  (iv) NMJ is from FSR $\rightarrow$ RJC = j1j2j3 - 19% of events

In 14% of cases two or more jets are not matched. We are planning to rescue these events by implementing a more efficient matching algorithm. This algorithm will search for hadrons rather than quarks in the jet cone and will trace back the origin of the hadrons in order to understand if they were produced by a $Z$-decay.

Four different Neural Networks (NNs) are trained in MC signal events to isolate each of the above cases. These NNs use the information carried by the kinematical variables and should possibly make us able to decide event by event which is the proper jet combination to be used for building the $Z$-mass.

3.3. The importance of knowing the correct jet pair
In Fig. 5 we compare $M_{j1j2}$ distribution with the distribution built by using the event-by-event RJC (Sec. 3.2). These distributions have been built in the sample where two jet-to-quark matches are found.

![Figure 5](image-url)

In black, dijet mass built with the two leading jets, while the distribution in blue is built using the RJC.

The low and high mass tails affecting the $M_{j1j2}$ distribution are drastically reduced by choosing the correct combination.

3.4. Look for $M_{j1j2}$
In this paper only the NN whose goal is to isolate RJC = j1j2 from the rest ($NN_{12}$) will be shown. Results on the other NNs are still preliminary. In order to isolate events when $RJC = j1j2$ we investigate some kinematical variables. We divided the sample into two parts:

• RJC= j1j2
• Other jet combinations (RJC = j1j3, j2j3, etc.) which we name "OJC"

Below is the list of the variables used:
(i) \( m_{jj'}/m_{j1j2j3}^2 \)
(ii) \( \gamma_{jj'} = (E_j + E_{j'})/m_{jj'} \)
(iii) \( d\eta_{jj'} = |\eta_j - \eta_{j'}| \)
(iv) \( dR_{jj'} = \sqrt{d\eta_{jj'}^2 + d\phi_{jj'}^2} \)
(v) \( dR_{j1j2,j3} \), \( dR \) between the third jet and vectorial sum of two leading jets.
(vi) \( dR_{j1j2,j3}, dR \) between the third jet and vectorial sum of all the three jets.
(vii) llr, is the likelihood ratio used for discriminating a quark from a gluon jet. For further details see [8].

A subset of the aforementioned observables are shown in Fig. 6. Some additional variables defining jet \( \eta \) and \( p_T \) distributions are being studied:

- \( p_T(j_1+j_2)/p_T(j_3) \)
- \( \eta(j_1+j_2)/\eta(j_3) \)

which would distinguish between quark jets and gluon jets.

**Figure 6.** Angular distribution between jet pairs and llr jet distributions. \( RJC = j1j2 \) is compared to OJC. See text for symbol significance

\(^2\) \( jj' \) refers to the three possible combinations: \( j1j2, j1j3 \) and \( j2j3 \).
3.5. NN_{_{12}} output

In order to avoid background (W+jets, t\bar{t}, etc...) being sculpted later, the input variables are weighted. Weights applied are calculated such that the \( M_{j1j2} \) distribution in the OJC sample become approximately the same of the one in the \( RJC = j1j2 \) sample. By doing this we will decorrelate the NN_{_{12}} output from the numerical value of \( M_{j1j2} \) and make it sensitive only to the kinematical distributions of the involved variables. The variables described above are weighted accordingly and are used for training a Neural Network. We employ the MLP method, (see [9] for details). The NN_{_{12}} response is shown in Fig. 7.

3.6. Results

In order to understand the impact of this method on the sensitivity of the measurement we apply the NN_{_{12}} selection on the major sources of background (W+jets, Z+jets, ttbar and single top) events and compare them to WZ events. In Fig. 8 the standard and the new mass distributions are shown in both signal and background events. The signal is multiplied by 70 in order to facilitate a visual comparison.

![Figure 7. NN_{_{12}} output](image)

![Figure 8. (a) Mj1j2. (b) Mj1j2 if we require NN_{_{12}} \geq 0.5](image)

We note the better separation when a cut on the NN_{_{12}} output is applied. We note that as a consequence of the above mentioned decorrelation procedure, the \( M_{j1j2} \) distribution in the background events doesn’t get sculpted. Therefore the S/B ratio increases, where S, B are calculated in the selected mass window [70,110] GeV/c^2.
Table 2. S, B are calculated in the selected mass window [70,110] GeV/c². “G” gives the $\sigma/\mu$ ratio in the mass window, where $\sigma$ and $\mu$ are the width and the mean of the gaussian distribution used for the fit.

|       | std        | if NN$_{12}$ >0.5 |
|-------|------------|-------------------|
| S     | 79.59      | 6.52              |
| B     | 10322      | 329               |
| S/ B  | 0.008      | 0.02              |
| G for Evt/Wind | 16.22 % | 12.3%              |

3.7. Conclusions

After the NN$_{12}$ output cut, we select 25% of WZ events in which Z peak is appreciably narrower. The resolution is improved by 25% and signal over background is improved by a factor of 2. We plan to repeat the procedure to isolate events where RJC = j1j3; j2j3; j1j2j3 and use the information carried by these four NNs for deciding event by event which is the proper combination for building the Z mass.

If successfull, the technique will significantly increase the signal acceptance of several important analyses (Diboson, Higgs, etc.). We note that when applied to the b-tagged sample for the light Higgs search, the performance of this method is expected to change because of the different background composition (mostly $t\bar{t}$) and of the different fragmentation properties of b-flavored jets.

Acknowledgments

We are grateful to Prof. Giorgio Bellettini, Dr. Giuseppe Latino, Dr. Vadim Rusu and Dr. George Velev for many fruitful discussions and suggestions.

References

[1] Hagiwara K et al 1990 Phys. Rev. D 41 2113
[2] Abazov V M et al (D0 collaboration), arXiv:0904.0673v1.
[3] Culbertson R et al Search for Anomalous Production of di-photon+MET Events in 2/fb of Data CDF 9184
[4] Junk T et al Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 6.7 $fb^{-1}$ CDF 10241
[5] Mangano M L, Moretti M, Piccinini F, Pittau R, and Polosa A 2001 J. High Energy Phys. 07 001.
[6] Sjöstrand T, Edén P, Friberg C, Lönnblad L, Miu G, Mrenna S and Norrbin E 2001 Computer Phys. Commun. 135 238.
[7] B. Cooper, A. Messina Estimation of the Background to W→ ev+n jets events CDF 6636.
[8] Ketchum W, Rusu V, Alawi Z, Young-Kee K Search for ZW/ZZ Production in Leptons + Jets channel CDF 10061.
[9] Hoecker A, Speckmayer P, Stelzer J, Therhaag J, von Toerne E, Voss H 2009 TMVAUsersGuide http://tmva.sourceforge.net