Top-quark background estimation for physics BSM in the di-lepton and jets final state with the ATLAS detector

Kgomotso Monnakgotla
jeremiah.kgomotso.monnakgotla.cern.ch

School of Physics and Institute for Collider Particle Physics
University of the Witwatersrand

SAIP Conference 2019 Polokwane

July 11, 2019
Content

Motivation

The Madala Hypothesis

Object Selection

Analysis Strategy

Top background estimation

Summary
(Some) ATLAS and CMS Results
Motivation
Multi-lepton anomalies at the LHC


A wide range of analysis showing excesses in the multilepton production at the LHC has been studied.

The profile likelihood ratios for each of the individual fit results with that of their combination

Results
- The combined excesses corresponds to a significance of $8\sigma$
- These excesses can be explain by the Madala Hypothesis
- it is clear that di-lepton data is not well understood
Measurement of lepton differential distributions and the top quark mass with CMS @13 TeV: JHEP 10 (2018) 117

Analysis Selection
- Select dileptonic $t\bar{t}$ events: $e^\pm\mu^\mp$ final state
- Inclusive analysis: no requirement on jets!

Results
- $m_T$ distribution well described at high mass region
- High discrepancy for $m_T$ at the peak
- Discrepancy at low $m_{\ell\ell}$
* Is this a top MC mismodeling?
* Are there other non-top results observing these features?
Motivation
Measurement of the WW xsec in ATLAS

ATLAS @ 8 TeV: JHEP 09 (2016) 029
ATLAS @13 TeV: arXiv:1905.04242

Analysis Selection
- Select $\ell^\pm \ell^\mp$ and vetoing events with jets
  - $ee/\mu\mu$ (top-row)
  - $e\mu$ (middle and bottom row)

Results
- High discrepancy for $m_{\ell\ell} < 100$ GeV
- High discrepancy for lepton $p_T$ around 100 GeV
  * very unlike to be produced by fake leptons
  * New physics or poor modelling?
The Madala Hypothesis
The Madala Hypothesis

References:  

$gg \rightarrow H \rightarrow Sh$

- $H$: Heavy Scalar with mass range $[240, 2m_t]$ GeV
- $S$: Higgs-like Scalar with between $m_h$ and $m_H$
- $h$: Standard Model (SM) Higgs boson with mass $m_h = 125$ GeV

$S$ boson: BRs decays depend on its mass

- $> 135$ GeV: $S \rightarrow WW$ becomes the dominant decay

$H \rightarrow Sh$ produces a wide range of possible final states, dominated by leptons and jets
Fits to LHC data

The SM+BSM fit results for the di-lepton invariant mass spectrum from ATLAS in Run 1 (left) and CMS in Run 2 (right).

Results

- It is clear that there exists an excess of events at $m_{\ell\ell} < 100$ GeV
- Fit the $H \rightarrow Sh$ model to CMS data and ATLAS data in the region where $m_{\ell\ell} < 110$ GeV
- Fit of data results from arXiv:1901.05300
Object Selection

The **final state** for this search is characterized by **two** oppositely charged **leptons** and **jets**

### Requirement for Leptons

<table>
<thead>
<tr>
<th>Leptons:</th>
<th>Muons:</th>
</tr>
</thead>
<tbody>
<tr>
<td>electrons:</td>
<td>muons:</td>
</tr>
<tr>
<td>- $p_T^e &gt; 15$ GeV</td>
<td>- $p_T^\mu &gt; 15$ GeV</td>
</tr>
<tr>
<td>- $</td>
<td>\eta</td>
</tr>
</tbody>
</table>

### Requirement for Jets

<table>
<thead>
<tr>
<th>Jets:</th>
<th>b-tagged jets:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- $p_T &gt; 25$ GeV</td>
<td>- MV2C10 @ 85 %</td>
</tr>
<tr>
<td>- $</td>
<td>\eta</td>
</tr>
</tbody>
</table>

Select events with leading lepton $p_T > 27$ GeV, Sub-leading lepton $p_T > 15$ GeV and $\geq 2$ jets
Signal Characterization: Jet multiplicities

Number of b-tagged Jets ↓

Signal Features

- $N_{(b)jet}$: Veto events with $b$-jets: 
  - Reduce Top background
- Select events with ≥ 2 jets: High Signal/Background

Highest Signal/Background

- $N_{jets} \geq 2$+ $b$-jet veto
- Dominanted by Top background
- top purity: 60 %
Top background estimation
The Top CR and VR are both defined to be orthogonal to the signal region selection and required to be signal depleted.

<table>
<thead>
<tr>
<th>Observable</th>
<th>SR</th>
<th>Top-VR</th>
<th>Top-CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leading lep $p_T$</td>
<td></td>
<td>$&gt; 27$ GeV</td>
<td></td>
</tr>
<tr>
<td>Sub-leading lep $p_T$</td>
<td></td>
<td>$&gt; 15$ GeV</td>
<td></td>
</tr>
<tr>
<td>Number of jets</td>
<td></td>
<td>$\geq 2$</td>
<td></td>
</tr>
<tr>
<td>Number of $b$-tagged jets</td>
<td>$=0$</td>
<td>$=2$</td>
<td>$=0$</td>
</tr>
<tr>
<td>$m_{bb}$</td>
<td></td>
<td>$&gt; 150$ GeV</td>
<td></td>
</tr>
<tr>
<td>$m_{\ell\ell}$</td>
<td></td>
<td></td>
<td>$&gt; 150$ GeV</td>
</tr>
</tbody>
</table>

ATLAS Simulation work in progress

$\sqrt{s} = 13$ TeV, $\int L dt = 36$ fb$^{-1}$

$H \rightarrow Sh \rightarrow h\phi + \mu\nu e$

Plot: "CutBjet_TopSR/Mll"
Top Control Region: 0-\textit{b}jets+$m_{\ell\ell}$>150 GeV

\begin{center}
\begin{tabular}{l|c|c|c|c}
 & Higgs & WW & Other VV & $t\bar{t}$ \\
\hline
0 $b$-jets+$M_{ll}$ > 150 GeV & 3.19 ± 0.08 & 1355.89 ± 7.01 & 146.37 ± 1.90 & 3534.73 ± 17.30 \\
\end{tabular}
\end{center}

- Define Top CR by 0-$b$-jet + $m_{\ell\ell}$ > 150 GeV
- top-quark is the dominant background, mostly $t\bar{t}$
- data/MC agreement : 0.97
- Top($t\bar{t}$ + $Wt$) purity: 72%

\begin{itemize}
  \item $p_T^0$: Leading lepton $p_T$ ↑
  \item $p_T^1$: Sub-leading lepton $p_T$ ↑
\end{itemize}
The contribution from non-resonant $WW$ (see Lebohang Mokoena’s talk) is about 25%.

The contribution from $W+$jets, other $VV$ and $Z/\gamma^*$ processes add up to 3%.

Good agreement data/MC: 0.97

The residual mismodelling is due to missing EW corrections.
Define Top VR by 2-\textit{b}-jet + \textit{m}_{bb} > 150 \text{ GeV}

Top-quark enriched region

data/MC agreement : 0.97

Top(\textit{t}\textit{t} + \textit{Wt}) purity: 99%

The MC residual mismodelling is due to missing EW corrections
Top Validation Region: $2$-jets+$m_{bb}>150$ GeV

- Other Background contributions add up 1%
- Good agreement data/MC: 0.97
- $m_T$ and $p_T^{\ell\ell}$ are well described by the MC
- Observe a MC overshoot at low $\Delta \eta^{\ell\ell}$ and high $\Delta \phi^{\ell\ell}$
Summary

- Several results from ATLAS and CMS observe discrepancy between MC and data.
- The Madala hypothesis allows to better describe the data in top and WW SM results at LHC.
- The search for Madala Hypothesis is performed with two opposite charged leptons and jets in the final state.
- Top background is the dominant SM process to the $H \rightarrow Sh$ search, so it is crucial to validate the top MC.
- Top-CR has good data/MC agreement of 0.97 with 72% top purity.
- The top-VR has good data/MC of 0.97 with 99% top purity.
- The MC works very well in the regions of the phase-space where the signal is not expected.
Thank You!