Data-driven Estimation of SM Backgrounds for SUSY Searches at the LHC

Takayuki Yamazaki (University of Tokyo)

on behalf of the ATLAS Collaborations & CMS Collaborations
Outline

[1] Introduction
[2] Data-driven Estimation
   (1) One Lepton Mode
   (2) No Lepton Mode
[3] If SUSY exists …
[4] Summary
[1] Introduction
SUSY Particle at the LHC

Gluinos/squarks are produced via strong interactions.

Event Topology
- Large missing $E_T$
- High $P_T$ multi-jet
  (leptons)

Promising search mode
- No lepton mode
- One lepton mode
- Opposite-sign dilepton mode
- Same-sign dilepton mode

R-parity conservation is assumed. 

\[ R = (-1)^{3(B-L)+2s} \]

\[ \rightarrow \text{LSP is stable.} \]
Discovery Potential

\[ \tan \beta = 10, A_\sigma = 0, \mu > 0 \]
with systematics
\[ m_h = 120 \text{ GeV} \]

\[ \sim 1 \text{ TeV in } 1 \text{ fb}^{-1} \]

3 TeV in much integrated luminosity

Discovery of SUSY is promising at the LHC.
If SUSY exists, clear excess can be observed @ 1fb⁻¹.

Since SUSY signal is observed as an excess from SM background in missing \( E_T \) distribution, deep understanding of SM background is essential for the SUSY searches and SM background should be estimated from the data in the early stage of collision.

“SU3 point” : \( m_0=100\text{GeV} \), \( m_{1/2}=300\text{GeV} \)

gluino mass \( \sim 700\text{GeV} \), squark mass \( \sim 650\text{GeV} \)

QCD = light flavor, bb, cc
(1) One Lepton Mode
Selection Cuts

- missing $E_T > 100\text{GeV}$
- missing $E_T > 0.2\text{Meff}$
- at least 4 jets with $P_T > 50\text{GeV}$
- at least 1 jet with $P_T > 100\text{GeV}$
- Transverse Sphericity $S_T > 0.2$
- one lepton with $P_T > 20\text{GeV}$
- $M_T(\text{lepton, } E_T) > 100\text{GeV}$

$M_T(\text{lepton, } E_T) > 100\text{GeV}$ is required to reduce $t\bar{t}+\text{jets}$ and $W^\pm +\text{jets}$. 
Data-driven

- To estimate the SM BG from data, “control sample” is used.

As a control sample for one lepton mode, MT < 100GeV events are selected. (The other conditions are the same as the selection cuts.)

\(\bar{t}t\) + jets and \(W^\pm\) + jets dominant one lepton control sample is obtained.

Missing \(E_T\) distribution is independent of MT.
**Result**

The normalization factor is obtained with the event numbers of the signal region and the control sample in $E_T=100-200\text{GeV}$.

In low $E_T$ region, SUSY contribution is expected to be small.

<table>
<thead>
<tr>
<th></th>
<th>$mE_T &gt; 100\text{GeV}$</th>
<th>$mE_T &gt; 300\text{GeV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True BG</strong></td>
<td>203 +/- 6</td>
<td>12.4 +/- 1.6</td>
</tr>
<tr>
<td><strong>Estimated BG</strong></td>
<td>190 +/- 8</td>
<td>9.4 +/- 0.7</td>
</tr>
<tr>
<td>Ratio (Estimated/True)</td>
<td>0.93 +/- 0.05</td>
<td>0.76 +/- 0.11</td>
</tr>
</tbody>
</table>
(2) No Lepton Mode
Selection Cuts

- missing $E_T > 100\text{GeV}$
- missing $E_T > 0.2\text{Meff}$
- at least 4 jets with $P_T > 50\text{GeV}$
- at least 1 jet with $P_T > 100\text{GeV}$
- Transverse Sphericity $S_T > 0.2$
- no lepton with $P_T > 20\text{GeV}$
- $\Delta \phi(E_T - \text{jet } i) > 0.2$ ($i = 1, 2, 3$)

The reasons of QCD BG are

1. neutrinos emitted from semileptonic decays of b/c (real missing $E_T$)
2. mis-measurement of jet energies (fake missing $E_T$)

In both cases, the direction of $E_T$ is approximately pointing to the jet direction.

$\Delta \phi(E_T - \text{jet } i) > 0.2$ cut reduces QCD BG by $\sim 80\%$. 

ATLAS

12 March 2008 Moriond QCD 2008
Data-driven (QCD BG)

- QCD BG can be estimated from multi-jet events without missing $E_T$.

multi-jet events without missing $E_T$

1st Jet

2nd Jet

multi-jet events

$\left(N_{50\text{GeV}} \geq 4 \& \& N_{100\text{GeV}} \geq 1 \& \& E_T < 100\text{GeV}\right)$

Apply $E_T$ function to 1st and 2nd leading jets

Reproduced semi-leptonic decay events

True QCD

Reproduced QCD

ATLAS
Data-driven (top / W BG)

- top / W processes contribute to the background in the no lepton mode when the lepton emitted in the $W^{\pm} \rightarrow l \nu$ process is not identified.

The main reasons of missing lepton are
1. tau->hadrons
2. out of acceptance ($P_T < 20\text{GeV}$)

Since the kinematics of the control sample and that of top / W BGs are almost the same except for the existence of lepton, top / W BGs can be estimated with the control sample.
Result (QCD / top / W)

The normalization factor can be obtained from data, but its detail is in the backup slides.

<table>
<thead>
<tr>
<th></th>
<th>mET &gt; 100GeV</th>
<th>mET &gt; 300GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>True QCD/top/W</td>
<td>8077+/-90</td>
<td>300+/-17</td>
</tr>
<tr>
<td>Estimated QCD/top/W</td>
<td>8158+/-273</td>
<td>327+/-28</td>
</tr>
<tr>
<td>Ratio (Estimated/True)</td>
<td>1.01+/-0.04</td>
<td>1.09+/-0.11</td>
</tr>
</tbody>
</table>
Data-driven (Z BG)

- Z (->νν) BG in the no lepton mode can also be estimated.

**Control sample**

Z (->νν) BG

**Z->νν BG**

Replace Method

Two opposite-sign same flavor leptons with \( P_T > 20 \text{GeV} \)
\( \not{E}_T < 30 \text{GeV} \)
\( M_Z - 10 \text{GeV} < M_{ll} < M_Z + 10 \text{GeV} \)

The standard no lepton SUSY cuts are applied after replacing \( ll \) with \( νν \).

uncertainty \( \sim 15\% \)

12 March 2008 Moriond QCD 2008
If SUSY exists, SM backgrounds will be overestimated because of SUSY contamination in the control samples and its contribution to normalization region ($E_T=100-200\text{GeV}$), but SUSY can be discovered @ 1fb$^{-1}$ since the excess of SUSY signal is much larger.
[4] Summary

• \( \sim 1 \text{TeV} \) SUSY can be discovered in 1\( \text{fb}^{-1} \) at the LHC
• If SUSY exists, clear excess can be observed in \( \slashed{E}_T \) distribution.
• For the SUSY searches, SM background should be estimated from real data.
• By data-driven estimations, SM background can be estimated.
• If SUSY exists, SM background will be overestimated because of the SUSY contamination, but the overestimation is much smaller than the excess of the SUSY signal, and clear excess can be observed.
Backup
SUSY Particle at the LHC

LHC: proton-proton collider

Squarks and gluinos are produced via strong interactions.
Normalization of QCD BG

QCD BG is concentrated in min $\Delta \Phi = 0.0-0.2$. On the other hand, top, W, and Z BGs are flat. So, we can estimate the number of top, W, and Z BGs in min $\Delta \phi = 0.0-0.2$ from all BGs in min $\Delta \Phi = 0.4-0.6$ region. By subtracting the estimated number of top, W and Z BGs from the number of all BGs, the number of QCD BGs in min $\Delta \Phi = 0.0-0.2$ can be obtained.

We can estimate the number of QCD BG in this region.

• We chose min $\Delta \phi = 0.4 \sim 0.6$ to avoid QCD tail effect.
• The number of QCD BG after $\Delta \phi$ cut is obtained from the following equation.

$$\frac{\text{# of QCD BG (min } \Delta \phi = 0.0 - 0.2)}{\text{# of QCD BG (min } \Delta \phi = 0.0 - 0.2)} \times \left[ \frac{\text{# of All BG (min } \Delta \phi = 0.0 - 0.2)}{\text{# of All BG (min } \Delta \phi = 0.4 - 0.6)} \right]$$

We use MC information only to obtain this factor.
Normalization of top and W BG

- We can normalize the control sample using the number of the BG events in \( m_{\text{ET}}=100-200 \text{GeV} \). This number is obtained by the following equation.

\[
N(\text{top and W BG}) = N(\text{All BG}) - N(\text{QCD BG}) - N(\text{Z BG})
\]

- The number of QCD BG can be obtained from the equation in the last page.
- The number of Z BG can be obtained with “MC method” or “Replace method” and its error is about 20% @ 1fb\(^{-1}\).
- When SUSY exists, \( N(\text{top and W BG}) \) is overestimated since SUSY signal contributes to \( N(\text{All BG}) \).
OS Dilepton Mode

In a similar way, the BG in the OS dilepton mode can be estimated from data.

<table>
<thead>
<tr>
<th></th>
<th>mET &gt; 100GeV</th>
<th>mET &gt; 300GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>True BG</td>
<td>98.8 +/- 9.9</td>
<td>6.0 +/- 2.5</td>
</tr>
<tr>
<td>Estimated BG</td>
<td>92.1 +/- 10.7</td>
<td>4.2 +/- 0.5</td>
</tr>
</tbody>
</table>
Systematic Errors
- No lepton mode -

<table>
<thead>
<tr>
<th></th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>jet &amp; mET energy scale</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>lepton energy scale</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>lepton efficiency</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>generator (ALPGEN-&gt;MC@NLO)</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>ALPGEN generation parameter</td>
<td>&lt; 5%</td>
</tr>
</tbody>
</table>

Data-driven estimation is stable against various systematic uncertainties.
Systematic Error

- OS dilepton mode -

<table>
<thead>
<tr>
<th></th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>jet &amp; mET energy scale</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>lepton efficiency</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>generator (ALPGEN-&gt;MC@NLO)</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>ALPGEN generation parameter</td>
<td>&lt; 5%</td>
</tr>
</tbody>
</table>
One Lepton Mode

<table>
<thead>
<tr>
<th></th>
<th>mET $&gt;$ 100GeV</th>
<th>mET $&gt;$ 300GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>True BG</td>
<td>203$\pm$6</td>
<td>12.4$\pm$1.6</td>
</tr>
<tr>
<td>Estimated BG</td>
<td>190$\pm$8</td>
<td>9.4$\pm$0.7</td>
</tr>
<tr>
<td>Ratio (Estimated/True)</td>
<td>0.93$\pm$0.05</td>
<td>0.76$\pm$0.11</td>
</tr>
</tbody>
</table>

12 March 2008

Moriond QCD 2008
No Lepton Mode

\[ \min \Delta \Phi @ \text{mET}=100 \sim 200\text{GeV} \]

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>0.0 \sim 0.2</th>
<th>0.2 \sim 0.4</th>
<th>0.4 \sim 0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>All BG</td>
<td>11826</td>
<td>4720</td>
<td>1002</td>
<td>838</td>
</tr>
<tr>
<td>QCD BG</td>
<td>5025</td>
<td>4002</td>
<td>289</td>
<td>129</td>
</tr>
</tbody>
</table>

Truth QCD (\(\min \Delta \Phi < 0.2\)) = 4002 +/- 63

Estimated QCD (\(\min \Delta \Phi < 0.2\)) = 3881 +/- 75

\[ \frac{\# \text{ of QCD BG (min } \Delta \phi > 0.2)}{\# \text{ of QCD BG (min } \Delta \phi < 0.2)} = 0.256 \pm 0.009 \]

Numbers @ \(\min \Delta \Phi > 0.2 \& \& \text{mET}=100 \sim 200\text{GeV}\)

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>QCD</th>
<th>Z</th>
<th>top/W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth</td>
<td>7106 +/-84</td>
<td>1023 +/-32</td>
<td>560 +/-24</td>
<td>5523 +/-74</td>
</tr>
<tr>
<td>Estimation</td>
<td>-</td>
<td>992 +/-40</td>
<td>560 +/-112</td>
<td>5554 +/-146</td>
</tr>
</tbody>
</table>

12 March 2008
Moriond QCD 2008
No Lepton Mode

<table>
<thead>
<tr>
<th></th>
<th>mET &gt; 100GeV</th>
<th>mET &gt; 300GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>True QCD/top/W</td>
<td>8077 +/-90</td>
<td>300 +/-17</td>
</tr>
<tr>
<td>Estimated QCD/top/W</td>
<td>8158 +/-273</td>
<td>327 +/-28</td>
</tr>
<tr>
<td>Ratio (Estimated/True)</td>
<td>1.01 +/-0.04</td>
<td>1.09 +/-0.11</td>
</tr>
</tbody>
</table>
MC Method (No Lepton Z BG)

- Z (->vv) BG in the no lepton mode can also be estimated.

The normalization factor is obtained with the event numbers of the MC sample and the control sample.

Since Z->ll events are statistically limited, we rely on the MC sample (Z->vv) for the shape of Z->vv BG distributions and just obtain the normalization factor with the event numbers of the MC sample (Z->ll) and the control sample (Z->ll).

Control sample

Two opposite-sign same flavor leptons with $P_T > 20$ GeV

$E_T < 40$ GeV

$M_Z - 10$ GeV $< M_{ll} < M_Z + 10$ GeV

The standard no lepton SUSY cuts are applied after replacing ll with vv.

uncertainty $\sim 20\%$
Because of SUSY contamination, the control sample becomes harder.
One Lepton Mode

-> We can discover SU3 @ 1fb⁻¹.

<table>
<thead>
<tr>
<th></th>
<th>mET &gt; 100GeV</th>
<th>mET &gt; 300GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>True QCD/top/W</td>
<td>203+/−6</td>
<td>12.4+/−1.6</td>
</tr>
<tr>
<td>Estimated QCD/top/W</td>
<td>296+/−10</td>
<td>33.3+/−1.4</td>
</tr>
<tr>
<td>True QCD/top/W + SUSY</td>
<td>653+/−8</td>
<td>245+/−4</td>
</tr>
</tbody>
</table>
No Lepton Mode

min $\Delta \Phi$ @ mET=100~200GeV

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>0.0~0.2</th>
<th>0.2~0.4</th>
<th>0.4~0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUSY + All BG</td>
<td>12540</td>
<td>4785</td>
<td>1068</td>
<td>903</td>
</tr>
<tr>
<td>QCD BG</td>
<td>5025</td>
<td>4002</td>
<td>289</td>
<td>129</td>
</tr>
</tbody>
</table>

Truth QCD (min $\Delta \Phi < 0.2$) = 4002 +/- 63
Estimated QCD (min $\Delta \Phi < 0.2$) = 3881 +/- 75

$\# \text{ of QCD BG (min } \Delta \phi > 0.2) / \# \text{ of QCD BG (min } \Delta \phi < 0.2) = 0.256 \pm 0.009$

Numbers @ min $\Delta \Phi > 0.2$ && mET=100~200GeV

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>QCD</th>
<th>Z</th>
<th>top/W</th>
<th>SUSY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth</td>
<td>7755+/-88</td>
<td>1023+/-32</td>
<td>560+/-24</td>
<td>5523+/-74</td>
<td>649+/-25</td>
</tr>
<tr>
<td>Estimation</td>
<td>-</td>
<td>992+/-40</td>
<td>560+/-112</td>
<td>6203+/-148</td>
<td>(0)</td>
</tr>
</tbody>
</table>
No Lepton Mode

![Graph showing event distribution vs. Missing ET and Effective Mass](image)

<table>
<thead>
<tr>
<th></th>
<th>mET &gt; 100GeV</th>
<th>mET &gt; 300GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>True QCD/top/W</td>
<td>8077 +/- 90</td>
<td>300 +/- 17</td>
</tr>
<tr>
<td>Estimated QCD/top/W</td>
<td>9726 +/- 312</td>
<td>804 +/- 52</td>
</tr>
<tr>
<td>True QCD/top/W + SUSY</td>
<td>12925 +/- 114</td>
<td>3000 +/- 55</td>
</tr>
</tbody>
</table>

-> We can discover SU3 @ 1fb^{-1}.

12 March 2008

Moriond QCD 2008
**Improved MT Method**

**Improved MT Method**
- To subtract SUSY contamination from the control sample $M_T < 100\text{GeV}$, signal region $M_T > 100\text{GeV}$ times $0.6$ is subtracted from the control sample.

\[
\frac{\text{# of SUSY signal in the control sample } M_T < 100\text{GeV}}{\text{# of SUSY signal in the signal region } M_T > 100\text{GeV}} \quad \text{MC information}
\]

- To reduce SUSY contribution in the normalization region, we use the number of events in $\not{E}_T=100$-150GeV.
One Lepton Mode – Improved MT-

<table>
<thead>
<tr>
<th></th>
<th>$\not{E}_T &gt; 100\text{GeV}$</th>
<th>$\not{E}_T &gt; 300\text{GeV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True BGs</td>
<td>203+/-6</td>
<td>12.4+/-1.6</td>
</tr>
<tr>
<td>Estimated BGs</td>
<td>212+/-11</td>
<td>12.3+/-1.0</td>
</tr>
<tr>
<td>True BGs + SUSY</td>
<td>653+/-8</td>
<td>245+/-4</td>
</tr>
</tbody>
</table>

12 March 2008 Moriond QCD 2008
No Lepton Mode - Improved MT -

<table>
<thead>
<tr>
<th></th>
<th>mET &gt; 100GeV</th>
<th>mET &gt; 300GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>True QCD/top/W</td>
<td>8077+/-90</td>
<td>300+/-17</td>
</tr>
<tr>
<td>Estimated QCD/top/W</td>
<td>8419+/-405</td>
<td>397+/-67</td>
</tr>
<tr>
<td>True QCD/top/W + SUSY</td>
<td>12925+/-114</td>
<td>3000+/-55</td>
</tr>
</tbody>
</table>

12 March 2008 Moriond QCD 2008