VBF H->bb update

Anton Kapliy
University of Chicago

FTK meeting
July 13 2006
Algorithm

1. Identify two b-jets, two tagging jets (defined as a highest Pt jet in forward and backward region)
2. Apply a variety of cuts (Pt, deltaEta, Mbb etc).
3. Apply a threshold cut to the remaining events. Accept each event with a weight given by probability to pass a threshold.
4. Predict how many SIG and BG pass after 300 fb^-1 of LHC data.
5. Roughly estimate the significance via #sig/sqrt(#bg), with errors.
6. *Improve it by explicitly evaluating sigma{BG} via a toy MC model. The improvement mainly comes from our knowledge of approximate Mbb.*
7. Plot the combined SIG and MC BG plot (with each signal first properly normalized, taking into account pass efficiency and Xsection).
8. NOTE: using MC BG instead of BG, because BG has too little statistics!
9. Fit the combined curve to exponent.
10. Subtract the fit to obtain Higgs signal.
1. Get enough BG data to deduce its shape. Suppose it follows a falling exponent from 60 GeV to 200 GeV. Guess the functional form [blue curve]

2. Generate random events that follow this curve [black histogram]. # of events generated = # of BG events expected after 300 fb^-1, which is calculated using Xsection and pass probability.

3. Fit the histogram to the given functional shape and obtain errors on the parameters [red curve]

4. Propagate these errors to the error on the integral under this curve over the region where we expect the Higgs signal.

5. To get “error on the error” - i.e. sigma{delta[#BG]}, we repeat the above algorithm, changing the (#of events on input) by one sigma in both directions. Then we average the two numbers.

In the case shown: #[expect]=52M
If we integrate between 85 and 135 GeV:
delta[#BG] = 2941 ; sqrt(N) = 7201
Choosing cuts, part 1

<table>
<thead>
<tr>
<th>Cuts:</th>
<th>No cuts</th>
<th>DeltaEta{between forward jets} &gt; 2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot:</td>
<td><img src="image" alt="Plot" /></td>
<td><img src="image" alt="Plot" /></td>
</tr>
</tbody>
</table>

| #SIG/ sqrt(#BG): | 3.3 +/- 0.4 | 3.4 +/- 0.4 |

DeltaEta between tagging jets, sig and bg

Entries: 19030
Mean: 3.851
RMS: 1.564

Entries: 25093
Mean: 3.009
RMS: 1.565

Entries: 15600
Mean: 4.355
RMS: 1.222

Entries: 15830
Mean: 3.994
RMS: 1.17
Choosing cuts, part 2

<table>
<thead>
<tr>
<th>Cuts:</th>
<th>DeltaEta{between forward jets} &gt; 2.4</th>
<th>DeltaEta{between forward jets} &gt; 2.4 Trigger 4j40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot:</td>
<td><img src="image1.png" alt="Graph 1" /></td>
<td><img src="image2.png" alt="Graph 2" /></td>
</tr>
<tr>
<td>#SIG/\sqrt(#BG):</td>
<td>3.4 +/- 0.4</td>
<td>3.0 +/- 0.4</td>
</tr>
</tbody>
</table>

Recall that $y_t=50\text{GeV}$. Note that now it would be good to apply a Pt cut.
Choosing cuts, part 3

Cuts: DeltaEta{between forward jets} > 2.4
      Trigger 4j40

Plot:

#SIG/\sqrt(#BG): 3.0 +/- 0.4

With these cuts, let us look at the Mbb histograms
Mbb plots, using data from root files

SIG:
total: 100k
pass cuts: 5500
pass trig: 1931.49
efficiency: 1.93%

BG:
total: 3.5M
pass cuts: 3210
pass trig: 214.5
efficiency: 0.006%

SIG and BG added with proper weights (taking into account efficiency and Xsecs). Looks exactly like BG, of course.
Mbb extrapolated to 300 fb^-1
Extrapolating BG shape to MC BG to get rid of low-statistic jitter:

Note: we are NOT using MC-generated shape for SIG, as we have sufficient statistics:

1. Mass of Higgs reconstructed from b-bbar jets in atifact

Entries 8710
Mean 70.78
RMS 44.81

Entries 2.570219e+07
Mean 72.01
RMS 29.95

Note new range: 40..200
Combined Mbb and final higgs mass

Comment:

The quality of the final SIG histogram is hardly satisfactory. However, I did not spend a lot of time tweaking my fitting algorithm. I believe the result can be significantly improved by properly configuring TMinuit.

More on this on the next slide.
Same analysis, but assuming $x10 \, \text{XSec}\{\text{sig}\}$

Extracted from the combined histo:

<table>
<thead>
<tr>
<th>Mbb after BG subtraction (excl)</th>
<th>Entries</th>
<th>Mean</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.570219e+07</td>
<td>72.1</td>
<td>29.91</td>
</tr>
</tbody>
</table>

Original signal:

<table>
<thead>
<tr>
<th>1. Mass of Higgs reconstructed from $b$-$b\bar{b}$ jets</th>
<th>Entries</th>
<th>Mean</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>h_higgs_mass1</td>
<td>5500</td>
<td>103.2</td>
<td>24.09</td>
</tr>
</tbody>
</table>

Comment:
The algorithm I use to extract the SIG Mbb histogram from the combined histogram works well in this case: we can even see the substructure around the center of the peak.

Apparently my signal extraction algorithm need some tuning to work better with smaller signal on top of large background.
Numbers (for realistic Xsection)

SIG Xsec = 3.15+-10% pb
BG Xsec = 1.38E6+-10% pb
Integrated Luminosity = 300 fb^-1

Given 300 fb^-1 of data, we have:
#SIG = 18k +- 2k  // Note: sqrt(22k) = 130
#BG = 26M +- 3M

Xsec[SIG]*efficiency: 0.06 pb
Xsec[BG]*efficiency: 86 pb

#SIG/sqrt(#BG) = 3.6 +- 0.4

sqrt(#BG) = 5069
delta{BG} = 1910 +- 101  -  from a toy MC, requiring 75<Mbb<135
#SIG2 (with 75<Mbb<135) = 14043 +- 1439 (76% of total signal)
#SIG2/delta{BG} = 7.3 +- 0.8
TODO & known problems

Statistics:
• We have ~30M 2->2 and 2->3, but it is not enough to deduce the shape of BG curve.
• Dropping 2->2 (90%), which don't pass the threshold cut anyway, *does* give enough statistics. However, there are SHERPA problems with currently available BG that drops 2->2.
• In this study, I used the 2->3 ONLY yt=50GeV sample, which is known to have problems. But for now there's nothing else I can effectively use. However, it takes no work to rerun the algorithm once we have more correct BG.

Using MC BG:
• When I make a combined SIG + BG plot, I use BG generated via a toy MC that follows a function that I subjectively choose by looking at the BG histogram (more on this later). This should be OK as long as in the actual BG doesn't have any other peaks (from QED processes or elsewhere).

Need to improve the signal extraction algorithms:
It works well if we assume x10 higher Xsection for the signal. Need to improve it so it works better with more realistic Xsections.