Model Independent Search for New Phenomena

Peter Renkel: Southern Methodist Uni.
on behalf of the D0 Collaboration.  August 2011
Any sign of new physics in Tevatron data?

- Do we see what we expect from Standard Model?
  - Is this excess statistically significant?
  - Do we correctly model our detector/physics?
  - New Physics?

- Look in all Tevatron data
  - Split the Tevatron data into many final states
  - For each final state, examine multiple test distributions
  - If for a particular final state/test distribution see an excess, ask questions

General, allows to analyze many final states, however not as sensitive as dedicated approaches
The D0 experiment

- Muon system
- Muon tracking
- Calorimeter
- Em objects
- Fiber and Silicon trackers
Data MC

Preselection ($p_T > \sim 15 - 35 \text{ GeV}$) corrections, splitting into multiple final states

Vista.
Looking at DATA/MC shape/number agreement for each of final states in the bulk

Search for specific new physics high $p_T$ tails, SLEUTH

Leptonic
Corrections derived from fitting MC to DATA in 7 final states

QCD from Data

1 fb$^{-1}$
Fit factors

- Fit basic distributions (like objects $p_T$, $\eta$, $\phi$) simultaneously and use more complex variables to check.
- 7 inclusive final states
  - $ee$, $e\mu$, $\mu\mu$, $e$(veto on second lepton), $\mu$(veto on second lepton), $e\tau$, $\mu\tau$
  - High $p_T$ tails are out of the fit
Vista

- Divide data into 117 exclusive final states
  - Based on high $p_T$ objects
    - Jets, $b$-jets, electrons, muons, taus, MET
  - For each final state and for each distribution, check:
    - Data/MC agreement
      - In number of events
      - In shape using Kolmogorov-Smirnov probabilities
    - Should account for large number of final states/distributions (trial factor)

Probability to see the final state as unlikely as state $i$ with probability $p_i$:

$$\tilde{P} = 1 - \prod_{i} (1 - p_i)$$

$\tilde{P} < 0.001$ corresponds to 3σ deviation
In this analysis, we analyze tenth of final states and hundreds of distributions. Therefore, the probability to observe a significant access is much larger than for a dedicated analysis. We correct for this effect taking into considerations the number of trials (final states or distributions).

\[ P = 1 - \prod_{i} (1 - p_i) \]

\[ p_i = \int \exp \left\{ - \frac{(N - N_{SM})^2}{2 \sigma_{SM}^2} \right\} dN \sum_{i=N_{data}}^{\infty} \frac{N^i}{i!} \exp\{-N\} \]

**Gaussian**: Probability that \( N \) is average, when we expect \( N_{SM} \) from SM

**Poisson**: Probability to observe at least \( N_{data} \) with average \( N \)
Vista results

- Total final states – 117
- Discrepant final states – 2 (3σ = discrepant)

- Total distributions – 5543
- Shape discrepancies - 16 (3σ = discrepant)

- Modeling issues (mostly spatial jet distributions)
  - No systematic effects are taken into account
  - Modeling jet recoil in the forward region
    - μ + 2 jets + MET
    - 4.5 σ
  - Resolutions for high $p_T$ muons
    - $\mu^+\mu^- +$ MET
    - 6.7 σ
Most discrepant Vista distributions

<table>
<thead>
<tr>
<th>VISTA Final State</th>
<th>Histogram</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu^\pm + 2$ jets + $\not{E}_T$</td>
<td>$M_T(W, j_2)$</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>$\Delta R(\mu, j_2)$</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>$M(\mu, j_2)$</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>$\Delta \eta(j_1, j_2)$</td>
<td>3.8</td>
</tr>
<tr>
<td>$\mu^\pm + 1$ jet + $\not{E}_T$</td>
<td>$p_T(W)$</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>$\Sigma p_T$</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>$p_T(\mu)$</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>$M_T(\mu^\pm, \not{E}_T)$</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>$\Delta \phi(\mu, j)$</td>
<td>3.1</td>
</tr>
<tr>
<td>$e^\pm + 2$ jets + $\not{E}_T$</td>
<td>$\Delta \eta(j_1, j_2)$</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>$M_T(j_2, \not{E}_T)$</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>$M_T(W, j_2)$</td>
<td>3.0</td>
</tr>
<tr>
<td>$e^\pm + 1$ jet + $\not{E}_T$</td>
<td>$\Delta \phi(e^\pm, j)$</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>$p_T(e^\pm)$</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>$p_T(W)$</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>$\not{E}_T$</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Shown are distributions with discrepancies $>3\sigma$

Mostly spatial distributions involving jets

Reminder: no systematics are considered
Each entry in a histogram corresponds to the deviation for a distribution.

The $\sigma$ distribution for the 117 final states

The $\sigma$ distribution for the 5543 distributions
High $p_T$ tails. Sleuth

Merge Vista final states
Lepton universality
Charge conjugation
117 Vista final states -> 31 SLEUTH final states

Cut $\Sigma p_T > C_0$ that gives the most significant excess
Correct for the trial factors

OS $e\mu$ final state

OS $e\mu + \text{MET}$ final state
Tests of the method

Are we able to re-discover $t\bar{t}$ pairs?

Remove $t\bar{t}$ MC

Run SLEUTH

Obvious discrepancy shows that SLEUTH can re-discover top pairs.

$P \sim 1.1 \times 10^{-5} \ll 10^{-3}$
Most discrepant SLEUTH
final states

<table>
<thead>
<tr>
<th>Final State</th>
<th>$\mathcal{P}$</th>
<th>$\tilde{\mathcal{P}}^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell^+ \ell^- + \not{E}_T$</td>
<td>$&lt; 10^{-5}$</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>$\ell^{\pm} + 2j + \not{E}_T$</td>
<td>$&lt; 10^{-5}$</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>$\ell^{\pm} + \tau^+ + \not{E}_T$</td>
<td>$8.9 \times 10^{-5}$</td>
<td>$0.0050$</td>
</tr>
<tr>
<td>$\ell^{\pm} + \not{E}_T + 1j$</td>
<td>0.00036</td>
<td>0.019</td>
</tr>
<tr>
<td>$e^{\pm} \mu^\mp + 2b + \not{E}_T$</td>
<td>0.0028</td>
<td>0.12</td>
</tr>
<tr>
<td>$\ell^{\pm} \tau^{\pm} + 2j + \not{E}_T$</td>
<td>0.0028</td>
<td>0.12</td>
</tr>
<tr>
<td>$\ell^{\pm} + 2b + \not{E}_T$</td>
<td>0.0077</td>
<td>0.3</td>
</tr>
<tr>
<td>$e^{\pm} \mu^\mp + \not{E}_T$</td>
<td>0.0081</td>
<td>0.31</td>
</tr>
<tr>
<td>$\ell^{\pm} \tau^{\pm}$</td>
<td>0.057</td>
<td>0.91</td>
</tr>
<tr>
<td>$\ell^{\pm} + 2b + 2j + \not{E}_T$</td>
<td>0.099</td>
<td>0.98</td>
</tr>
</tbody>
</table>

This passes the threshold of $3\sigma$ due to problems with detector modeling. Same as in VISTA.
Conclusion

- Performed Model-Independent search in D0 data
- Most states agree after trials
- The discrepant states/distributions are due to modeling issues
- SLEUTH – search for high $p_T$ tails.
  - No surprises
Backup
Preselection and Corrections

- Alpgen and PYTHIA
- Multijet from Data

- Leptonic final states
- Channel specific kinematic cuts

- Collaboration-wide corrections
  - K-factors
  - Trigger efficiencies
  - Lumi reweighting

- PYTHIA and MadEvent
- Multijet from MC

- Channel specific kinematic cuts

- Corrections later at Vista level
  - Constrained global fit
  - 43 fit parameters

CDF
PreSelection → Corrections → Leptonic → MIS fit

D0
PreSelection → Corrections → Global fit