Generic Search for Deviations from Standard Model Predictions in CMS

Shahram Rahatlou

for the CMS Collaboration
Introduction

- **Generic search strategies already explored at Tevatron and Hera**
  - Provide comprehensive comparison between data and Monte Carlo Predictions
  - Complementary strategy to augment dedicated searches
  - Not optimized for specific theoretical models
  - Sensitive to unexpected signatures not covered by specific models

- **More appropriate for later stage of data taking with well understood detector response**
  - Requires a solid reference for comparison
  - Needs to assume Standard Model to be correct and look for deviation in large number of event topologies

- **Useful physics monitoring tool with early data**
  - Provide feedback on instrumental effects causing discrepancies
  - Compare different Monte Carlo simulations
  - Tuning of various MC generators with data
Model Unspecified Search in CMS (MUSiC)

- Strategy very similar to that used in H1
- Lepton triggers
  - Reduce overwhelming QCD background
- Standard reconstructed objects
  - Benefit from improved reconstruction and identification studies with first data
    - Reduce systematic uncertainty
- Three distributions sensitive to deviations
  - Scalar sum of transverse momenta $\Sigma p_T$
    - Sensitive to products from new heavy particles
  - Invariant or transverse mass of event
  - Missing energy
    - Most sensitive to new physics but unreliable with very first data

Scan Distributions for discrepancies
Physics Objects with CMS Detector

| Event Categories           | p_T     | |η| | isolation |
|---------------------------|---------|-------------|----------------|
| 1e 1µ MET                 | > 30 GeV| < 2.4       | Track based    |
| 1e 1 jet MET              | > 30 GeV| < 2.5       | Calo+track     |
| 2 jet 1e 1µ + X           | > 50 GeV| < 2.5       |                |
| 4 jet 1e 1µ + X           | > 100 GeV|             |                |

Event Categories:
- 1jet 1e 1µ + X
- 2 jet 1e 1µ + X
- 4 jet 1e 1µ + X
- 1e 1µ
- 2e 1µ
- 1γ 1µ
- 1jet 1e 1µ
- 2 jet 1e 1µ
- 4 jet 1e 1µ
Robust Background Estimate

- Key ingredient for robust and credible search of deviations
  - Any quantitative measure of deviation washed away by large uncertainties on expected background

- Monte Carlo expectations good enough for most of electroweak backgrounds
  - Main deviations in high $p_T$ regions with very small SM expectation
  - Even 30% uncertainty reasonable for such regions
  - Large MC samples can reduce

- QCD background as main source of concern
  - Predictions with large uncertainties
  - More difficulties in producing large samples

- Data driven approach used to estimate QCD background
  - Systematic uncertainty to shrink with more statistics
Data Driven Estimate of QCD Background

- Control region defined by using looser isolation criteria to estimate background in data

- Two inclusive final states used to define single scale factor for all event categories
  - Dedicated region for each of ~400 states not feasible nor practical
  - Normalization estimated from low p_T regions dominated by Standard Model
    - Systematic uncertainty of 50%

![Graph 1: Signal region extrapolated from control region](image1.png)

![Graph 2: Signal region extrapolated from control region](image2.png)
Consider any single bin or group of adjacent bins with deviations

Compute Poisson probability $p_{\text{data}}$ for $N_{\text{MC}} \pm \delta N_{\text{MC}}$ to fluctuate up or down to $N_{\text{data}}$
- Convolution with Gaussian to account for systematic uncertainties

Region with largest discrepancy (smallest $p_{\text{data}}$) called Region of Interest

Define a quantitative measure of such discrepancy with pseudo-experiments based on Standard Model predictions
Probability of discrepancy to occur in data

- Determine distribution of \( p_{\text{data}} \) for toy MC experiments based on SM expectation
- Large fraction \( \tilde{P} \) of toy experiments with \( p < p_{\text{data}} \) indication of potential deviation
- Traditional interpretation in terms of standard deviations if considering \( P \) as tail of Gaussian distribution
- Use a 3\( \sigma \) threshold in automatic search for deviations
Physics Commissioning With Early Data

- Uncertainty of 5% assumed for Jet Energy Scale (JES) and included in probability calculation

- Ignoring this uncertainty and increasing JES by 10% in pseudo-experiments cause of $4.4\sigma$ discrepancy

- Such scenario can easily occur in early data taking
  - physics monitoring

- Including systematic uncertainty reduces discrepancy to $\sim 1.6\sigma$
Comparison of Generators with MUSiC

- **ALPGEN** and **PYTHIA** have different momentum spectrum and multiplicities spectra for harder jets.

- Toy experiments of \(W+\)jets with ALPGEN compared to inclusive \(W \rightarrow e \nu\) PYTHIA MC.

Region of Interest: \(p = 6.28 \times 10^{-9}\) and \(P > 4.4 \sigma\)

Excluded due to lack of stat. for PYTHIA.
Outlook

- A model-independent tool for automatic search of deviations from standard Model tested in CMS

- Function as physics monitor in early data
  - Contribute to identifying major detector effects affecting all physics objects

- Prove robustness as more data become available
  - Improved expectation to gain confidence in automatic search outcome

- In longer term with well understood detector and under-control systematic uncertainties attempt at search for new signals
  - Missing energy will be understood and corrected for instrumental effects
  - Deviations of missing energy crucial for search of unexpected signals

- With 1 fb\(^{-1}\) of data hints of deviations due to mSUGRA (LM4) on top of Standard Model expectation in many final states
  - 36% of inclusive final states in $\Sigma p_T$ spectrum
  - 59% of inclusive final states in MET spectrum