



# Statistical Tools for Dijet Mass Search

Kalanand Mishra

*Fermilab*

*(on behalf of dijet analysis team: R Demina, S Esen, E Halkiadakis, A Harel, R Harris, J Hirschauer, S Jabeen, C Jeong, K Kousouris, S Lee, D Mason, D Miner, K Mishra, S Ozturk, J St.-John, K Yi, M Zielinski, ...)*

*Exotica Multijets meeting  
(April 8, 2010)*



# Aims & overview of the new tools

## Aims

- A simple lightweight tool to compute CL or to set limit.
- Build on top of RooFit data modeling framework
- Use RooStats – the common statistics tools for LHC
- Accessible via cvs. Easy to understand, adopt, deploy.

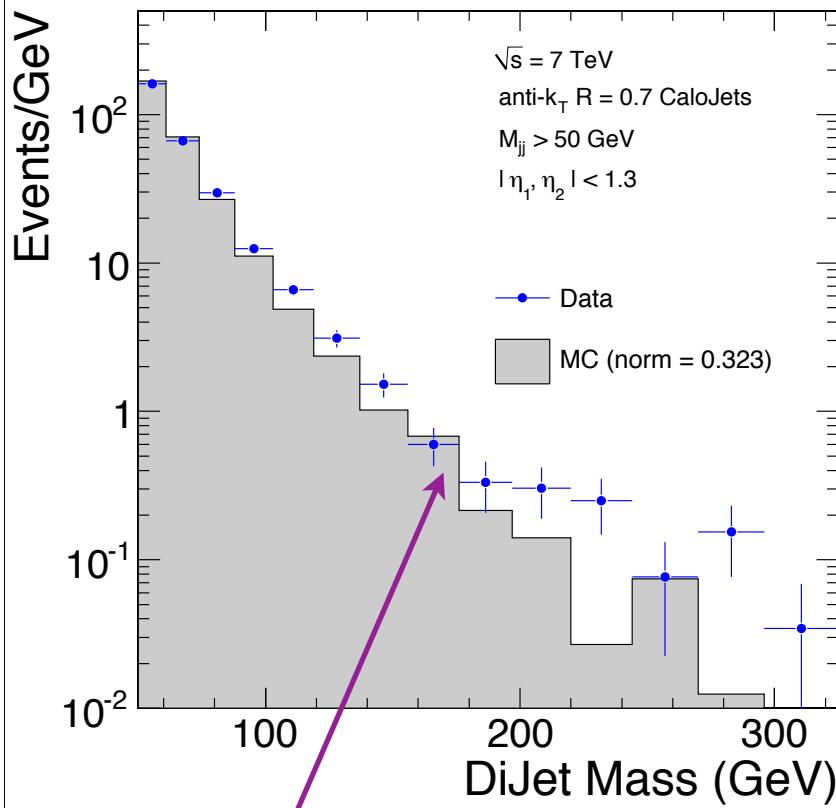
## Why ?

- ◆ Get the basic stuff done right, use gut of ROOT/RooFit fitting functionality
  - ▶ put simple example scripts actually used in our analysis in one place
    - e.g., code to fit the shape of the QCD dijet mass spectrum
    - code which uses MC templates to fit for new physics signal
      - making use of template-morphing technique
    - code to generate and fit pseudo experiments → estimate syst., biases
- ◆ RooStats: emergent common machinery to compute likelihood, significance, p-value for a hypothesis, or confidence level (CL)
  - blessed by CMS Statistics committee, being used by other experiments
  - easy to compare, cross check, and combine results
- ◆ No need to reinvent the wheels → most of what we need is already there!
  - put use case scripts in cvs: easier to vet the code and reproduce results

# Necessitated by immediate analysis needs

No fancy stuff, but need to have all the bells and whistles in right place for ICHEP...

Dijet mass distribution,  $d\sigma/dm$ , in data



Need to fit this spectrum to estimate background for dijet resonance search

- ◆ Need ability to fit data for Signal and Background (QCD) shapes and yields
  - Compute likelihood ratio:  $L(S+B)/L(B)$
  - convert to significance/CL
  - generate 100's of pseudo experiments of real data size, count outliers & bias
  - show both numerical value of the NLL minimum & visualization of LH contour

- ◆ For most of these tasks there is an appropriate class in RooFit/RooStat. We can focus on analysis task at hand
  - study background shape, find acceptable parametrization of the QCD dijet mass spectrum, signal model, systematic bias in shape and normalization, ....

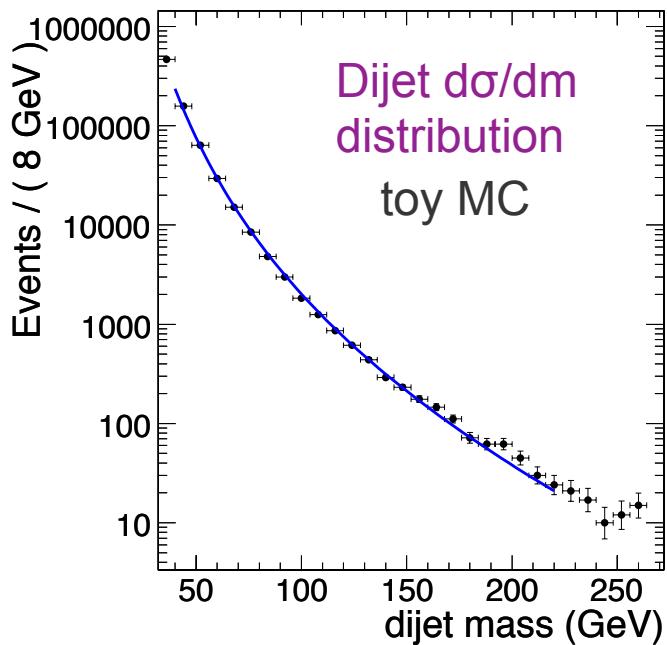
Many of these are repetitive → we would like to automate them and reuse the scripts



# Technical details

- ◆ Serious work started in the last two weeks
  - have started porting use case scripts into theUserCode area  
<http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/UserCode/kalanand/StatTools/>  
with detailed description and comments for readability
  - continue refining the updating the code as we go along
  - after we validate the existing results with new tool and with growth of user base we will move to more permanent place
- ◆ Using current scripts one can, e.g.,
  - fit QCD spectra and signal shape for dijet mass bump
  - compute likelihood ratio, p-value, and signal significance for dijet resonance
  - generate and fit pseudo data
- ◆ See next slide for illustration

# Example: generate QCD spectrum toy MC & fit



<http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/UserCode/kalanand/StatTools/GenerateAndFit.C>

Generate the following QCD line shape

$$\frac{d\sigma}{dm} = p_0(1 - x + p_3 x^2)^{p_1}/m^{p_2}, \quad x = m/\sqrt{s}$$

and fit using the same function (or variation of it) using unbinned maximum likelihood fit.

With 1M events, it took about 15 minutes to perform the unbinned ML fit. However, in real data, with efficiency and resolution unfolding, expect a few times longer. Still not a big deal.

Script “SignificanceCalculatorBinned” can generate dijet resonances of various mass using few template points from MC and template morphing technique. Then it performs likelihood fit to compute CL(S) and CL(S+B) to extract signal significance as a function of production cross-section for a given resonance mass.



# Summary

- ◆ Working on streamlined common tools to compute CL or set limit
  - make use of the RooFit data modeling/ fitting machinery
  - RooStat functionality for computing significance or setting limit
    - following recommendations of the CMS statistics group
  - Focus on simple and configurable scripts
    - so that everyone knows where the code is, what it does
    - easy to cross-check and combine results from different channels
    - relatively straightforward to deal with correlated systematics
- ◆ Porting our code/scripts to CVS/UserCode area
  - have already few use case scripts available
  - working on validation using pre-existing results
  - current work focused on understanding QCD background shape and fit systematics for the dijet mass analysis
  - we will report periodically as we go along