

LZStats

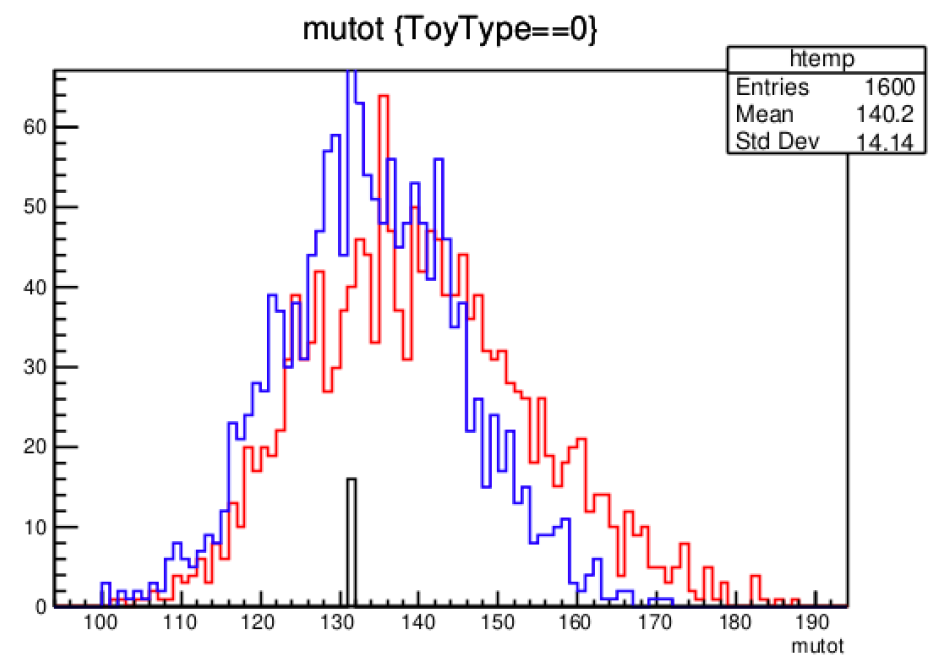
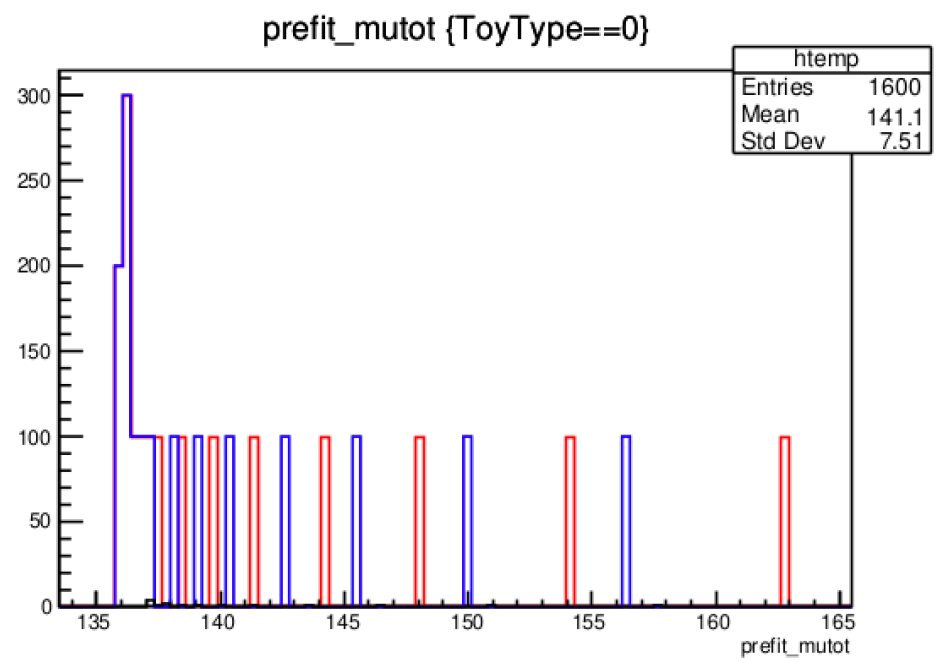
Yitong Liu

plr_analyser

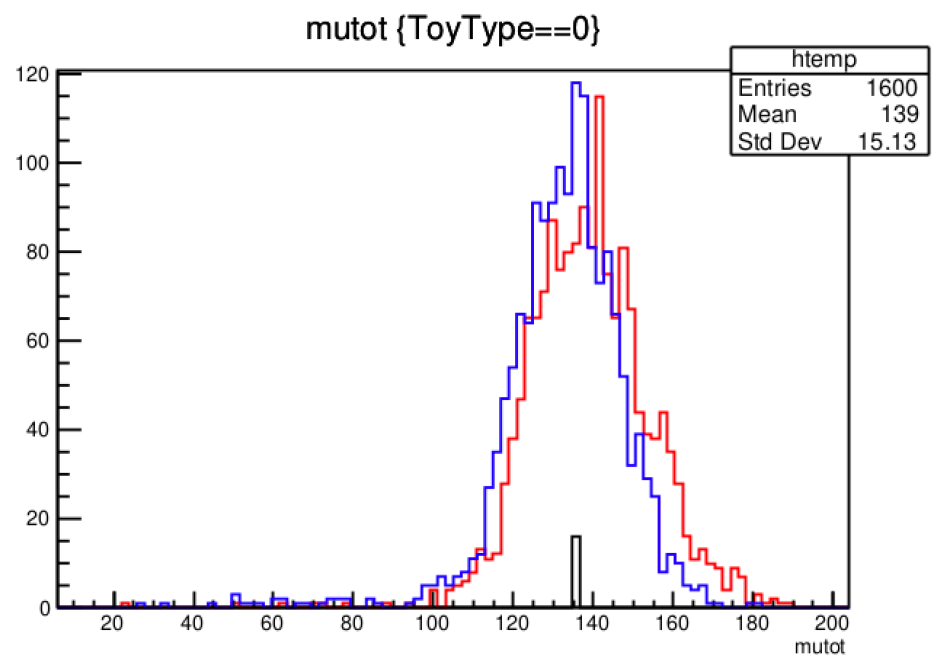
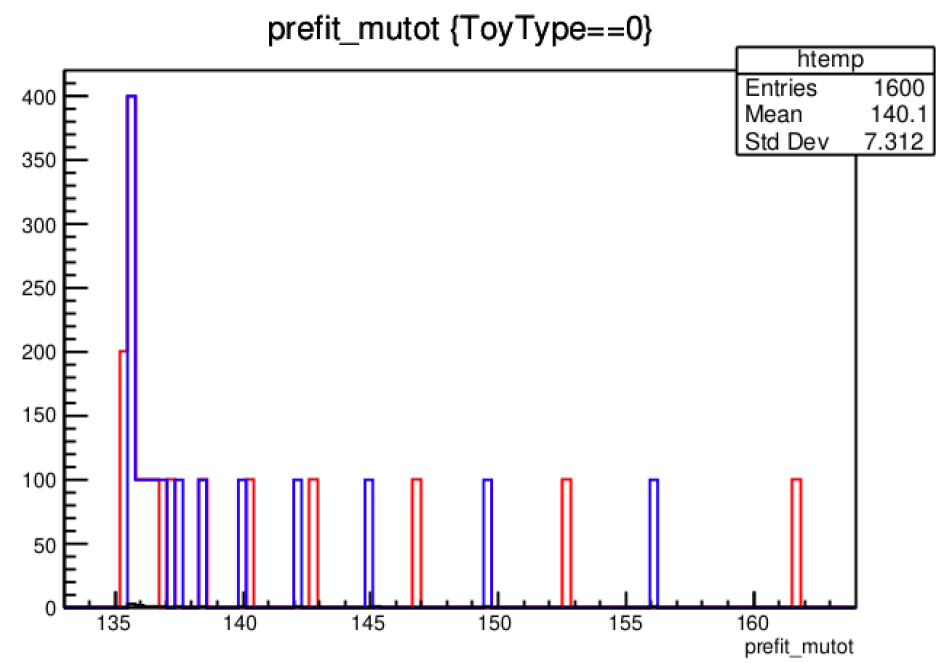
$$\lambda(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta}, \hat{\boldsymbol{\nu}})}{L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\nu}})}$$

- For each run, a FitResults.root is generated
 - **Conditional & unconditional fit** of mu_sig and backgrounds mu_(pp, vvBB, beta, lownr, highnr, acc)
 - **Global observables** a_(pp, vvBB, beta, lownr, highnr, acc)
 - Conditional fit & unconditional fit for **negative log likelihood**
 - **Test statistic**
 - **Toy type**
 - **prefit_mutot**: the “expected events” with the prefit values of nuisance parameters
 - **mutot**: : the “expected events” with the best fit values of nuisance parameters
 - **n0**: the true number of events inside the toy dataset
- Run PlotFitResults.C and compare different mu_sig range

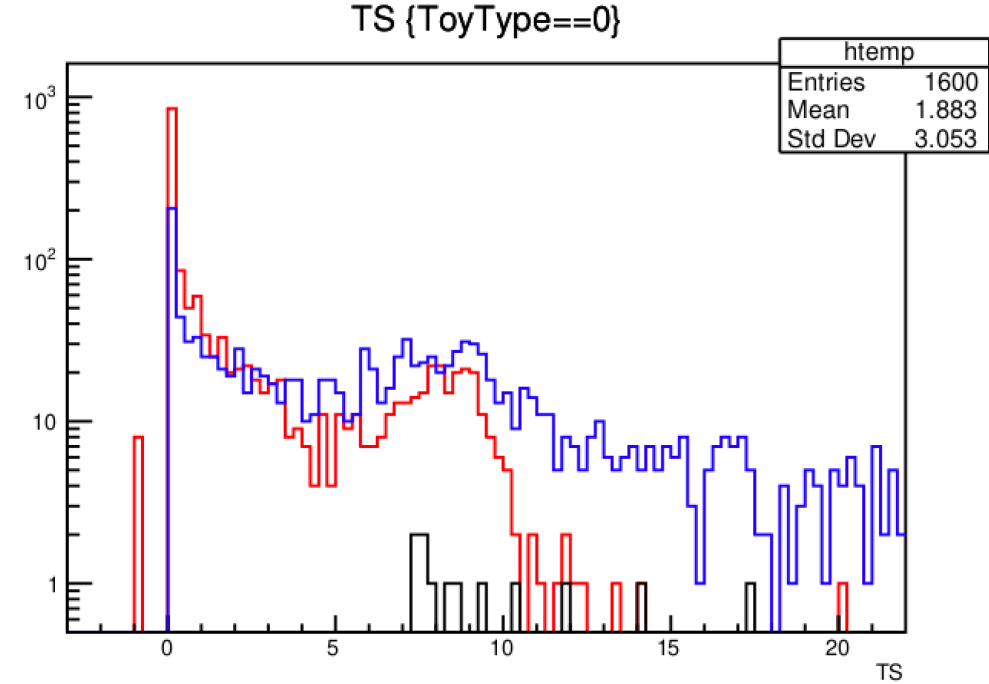
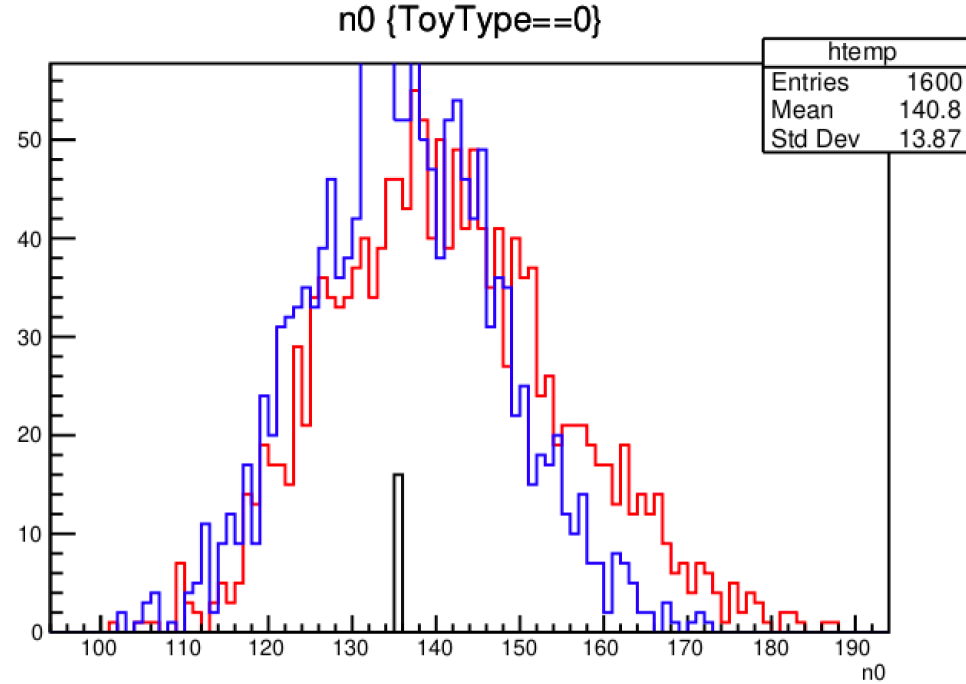
mu_sig (-5, 100)



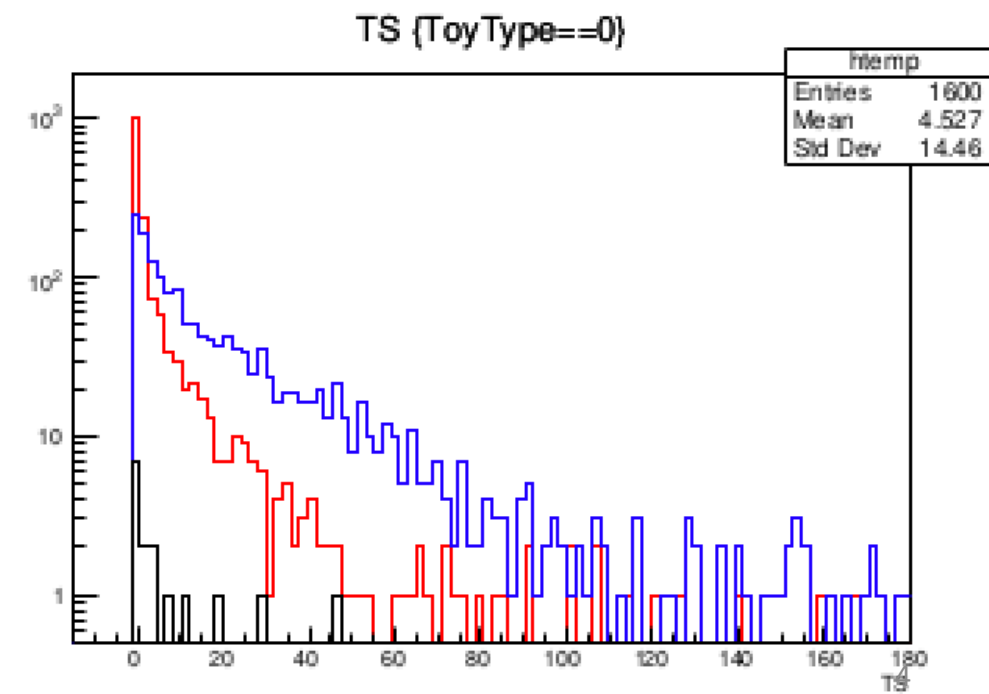
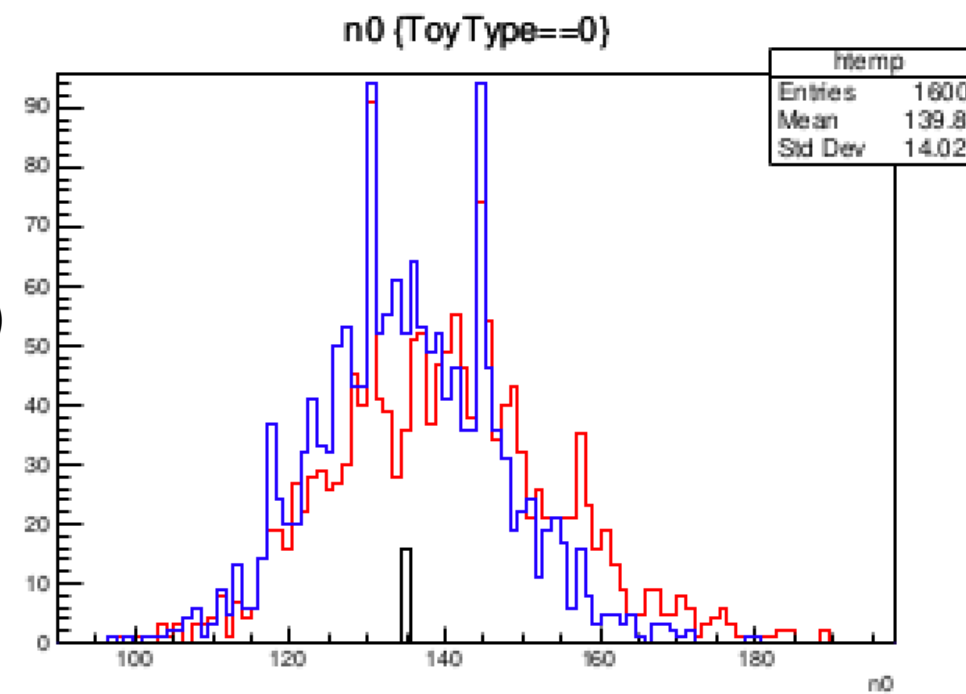
mu_sig (-100, 100)



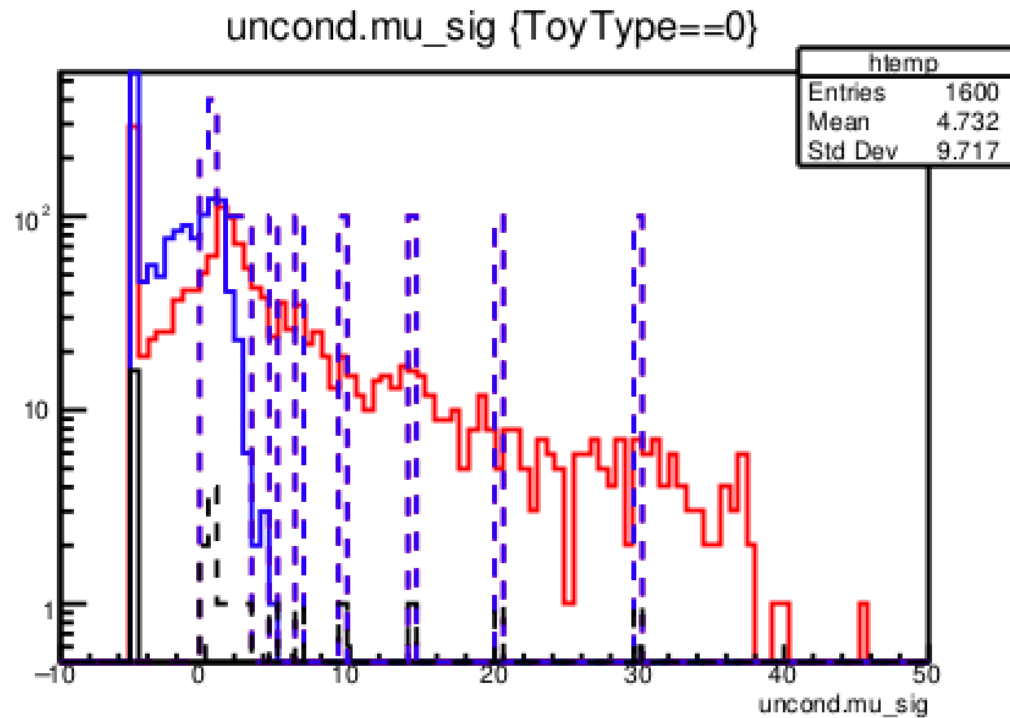
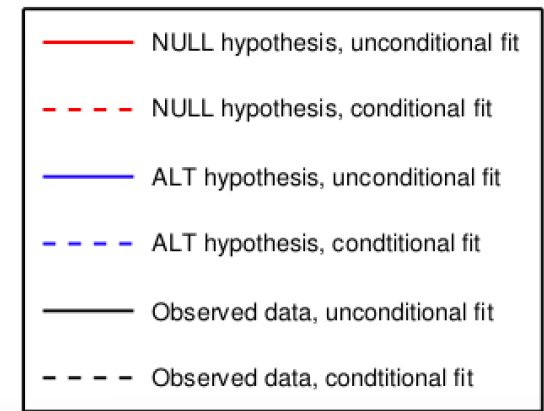
$\mu_{\text{sig}} (-5, 100)$



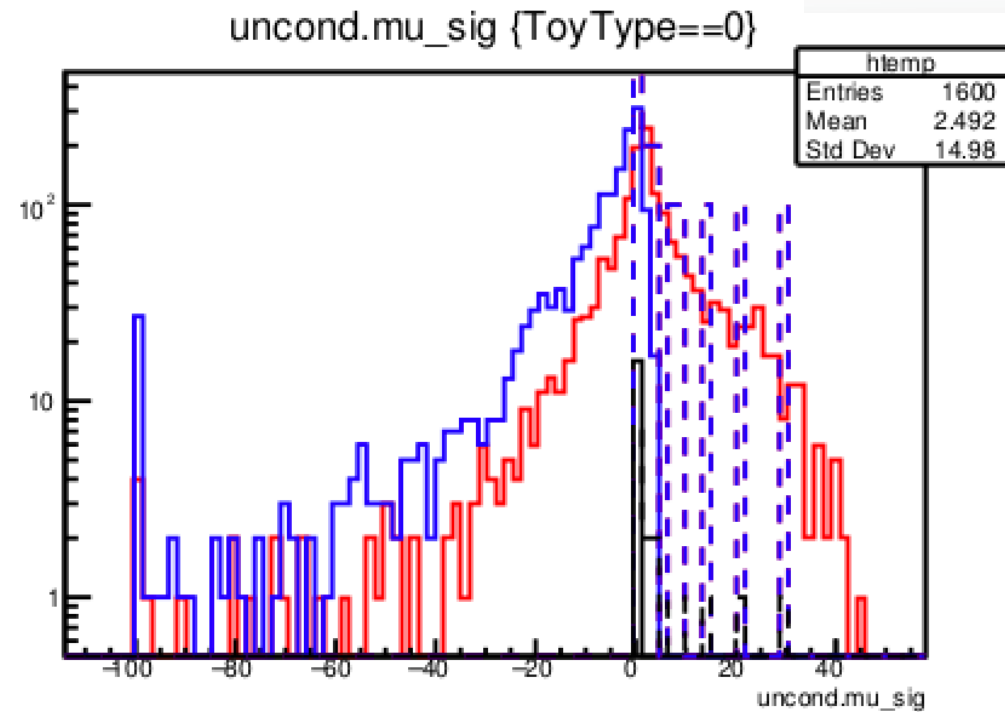
$\mu_{\text{sig}} (-100, 100)$



uncond.mu_sig (40 GeV)

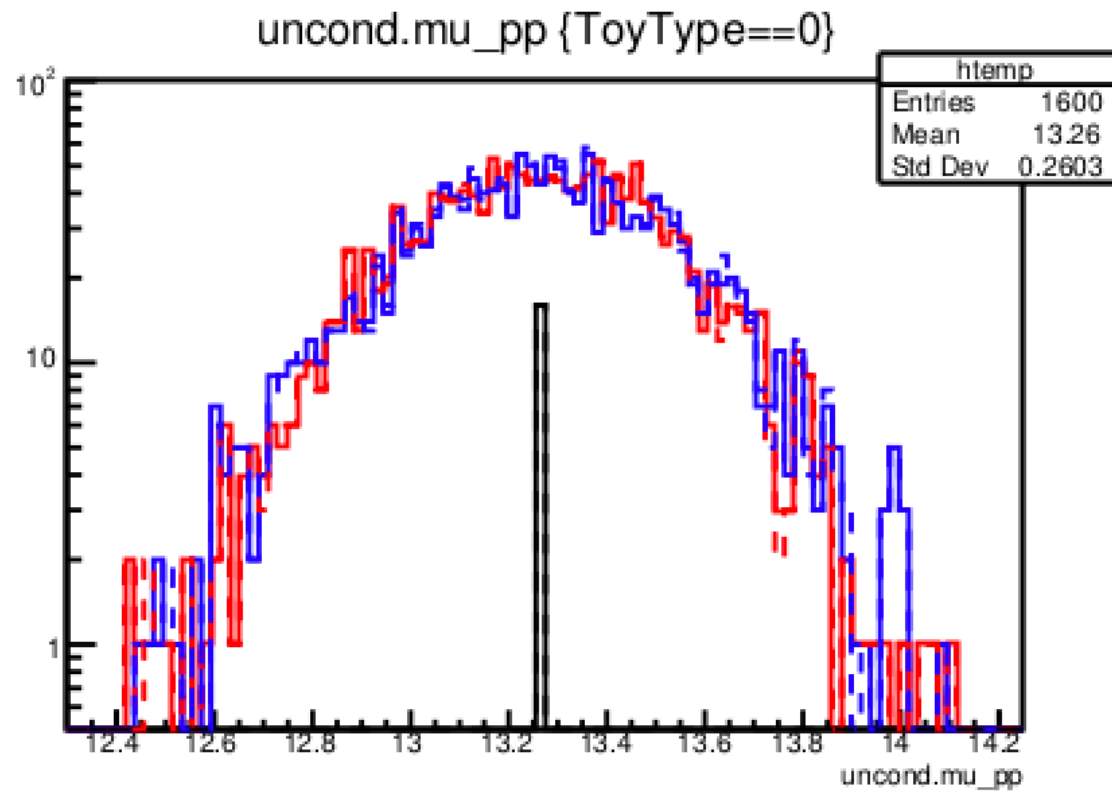


mu_sig (-5, 100)

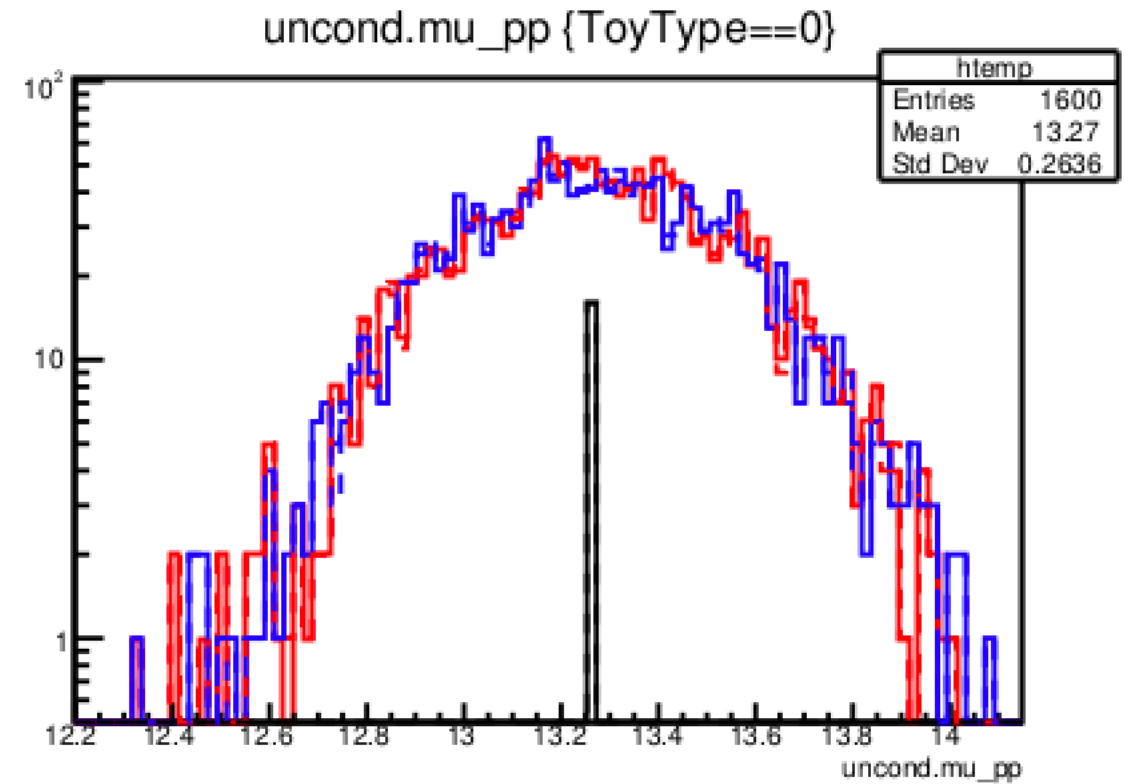


mu_sig (-100, 100)

uncond.mu_pp

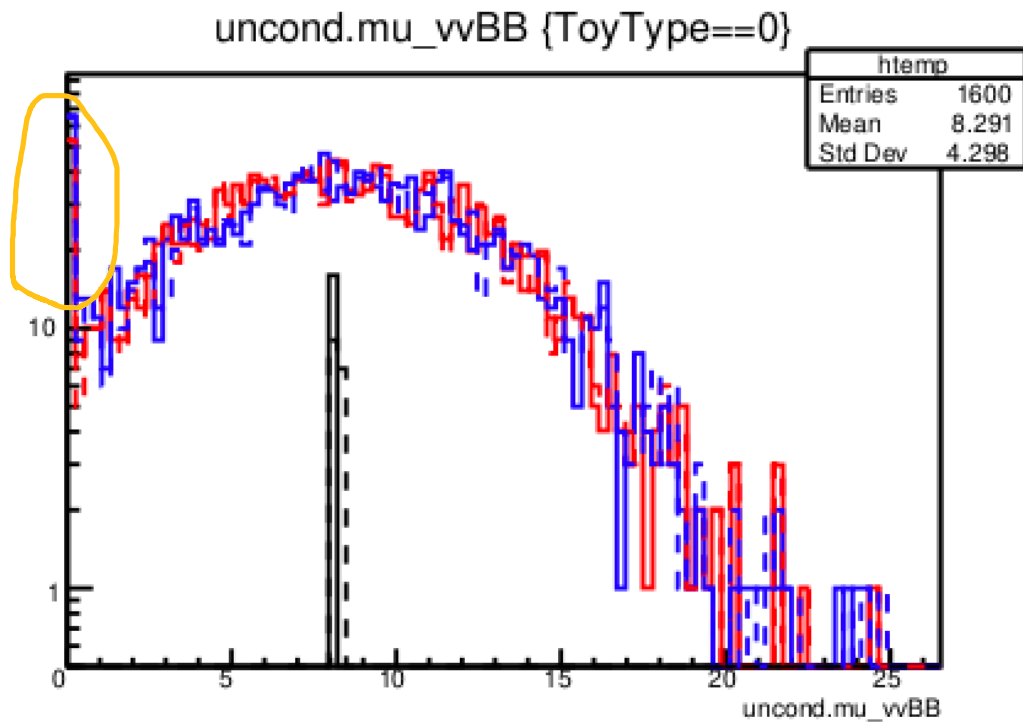


mu_sig (-5, 100)

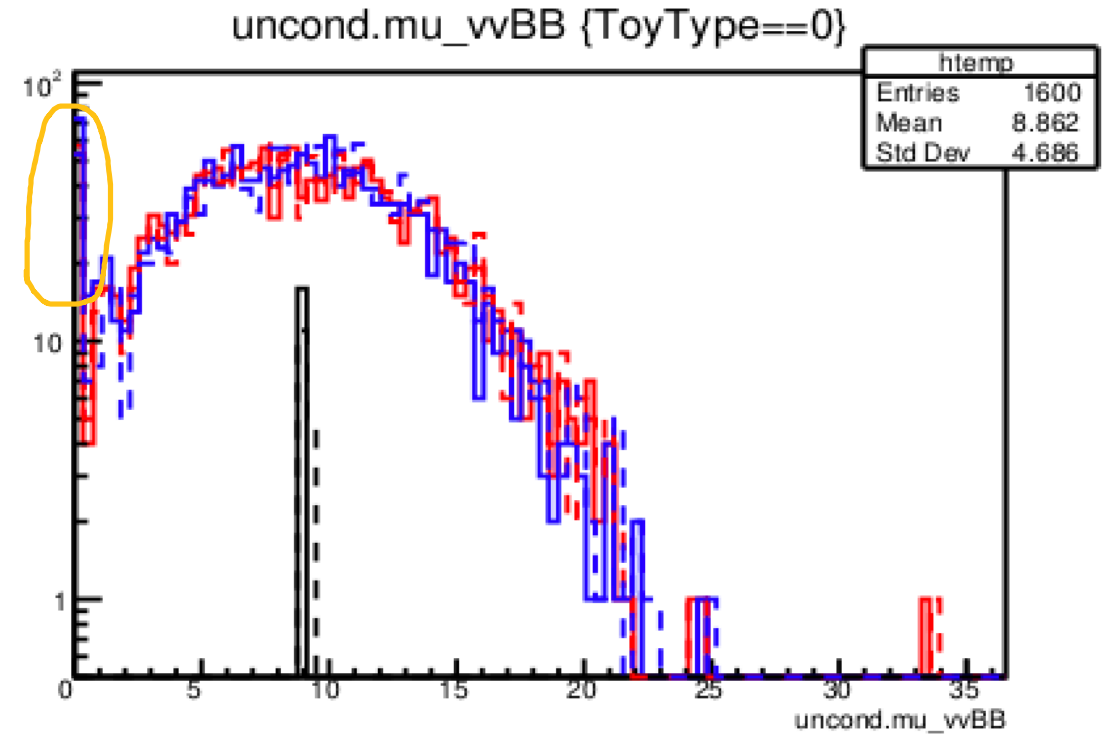


mu_sig (-100, 100)

uncond.mu_vvBB

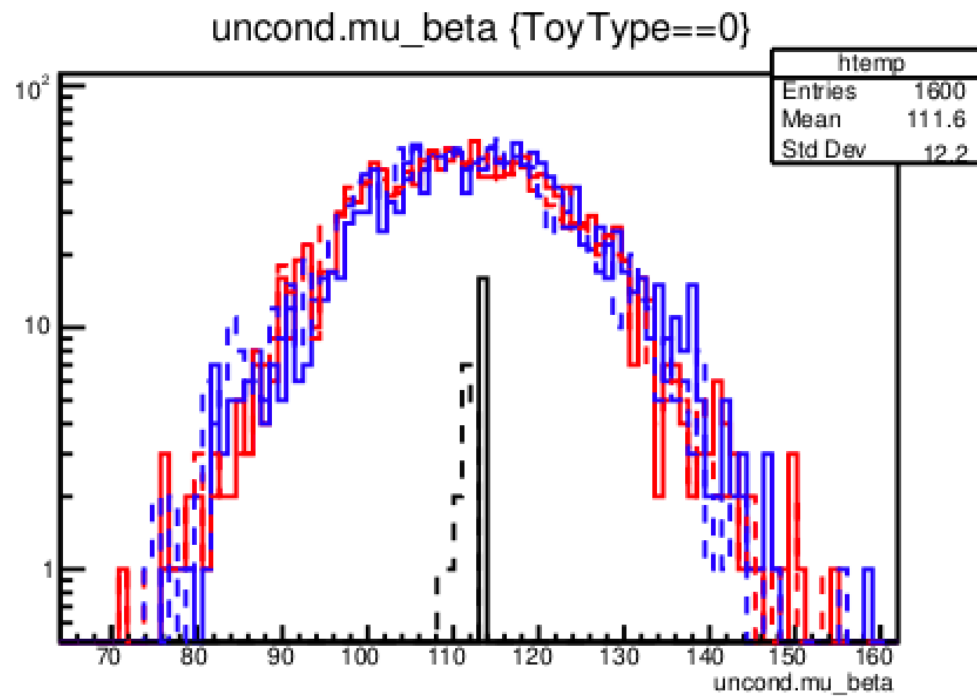


mu_sig (-5, 100)

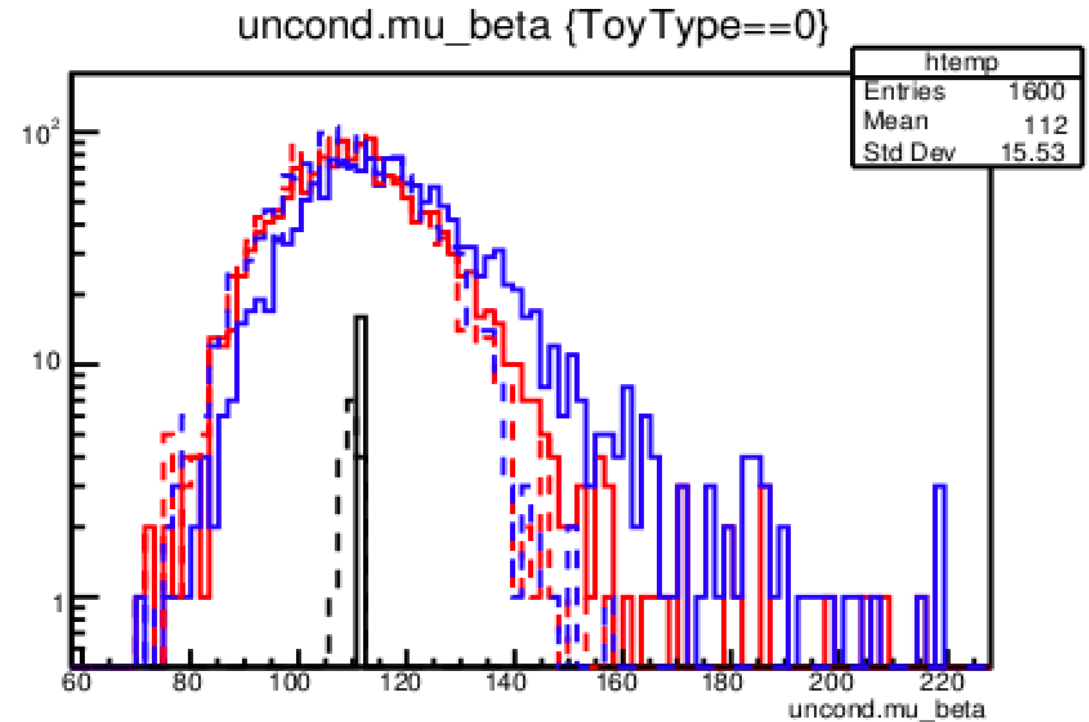


mu_sig (-100, 100)

uncond.mu_beta

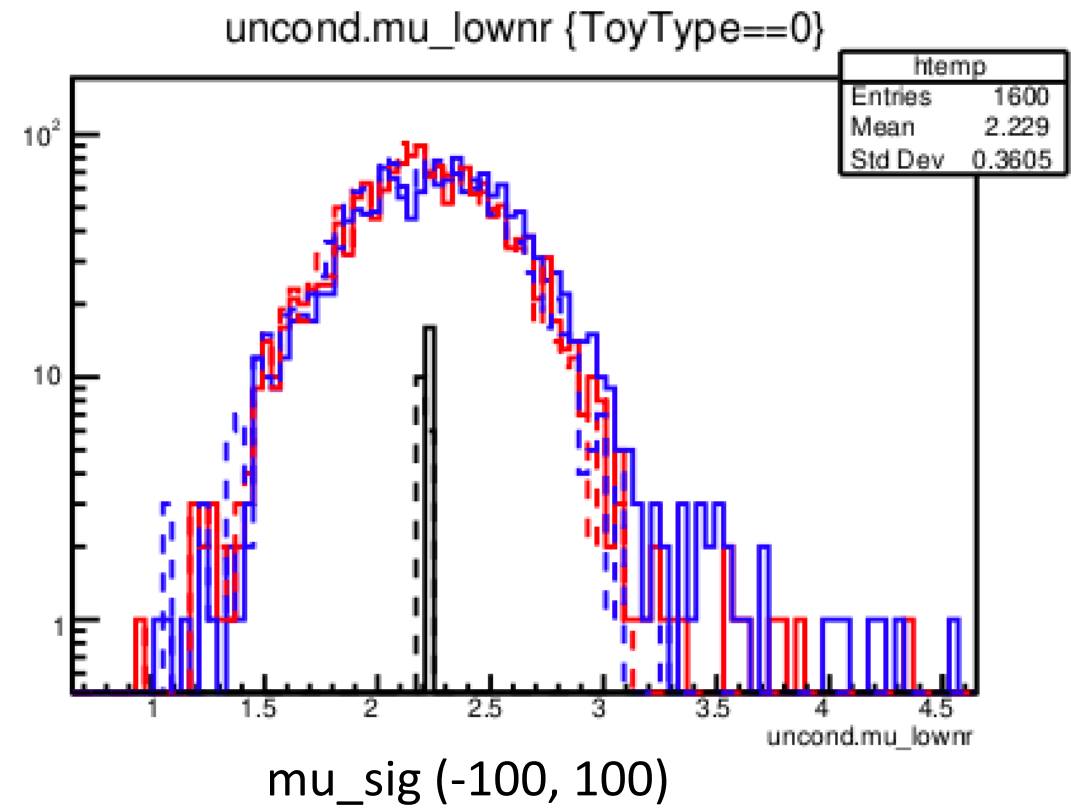
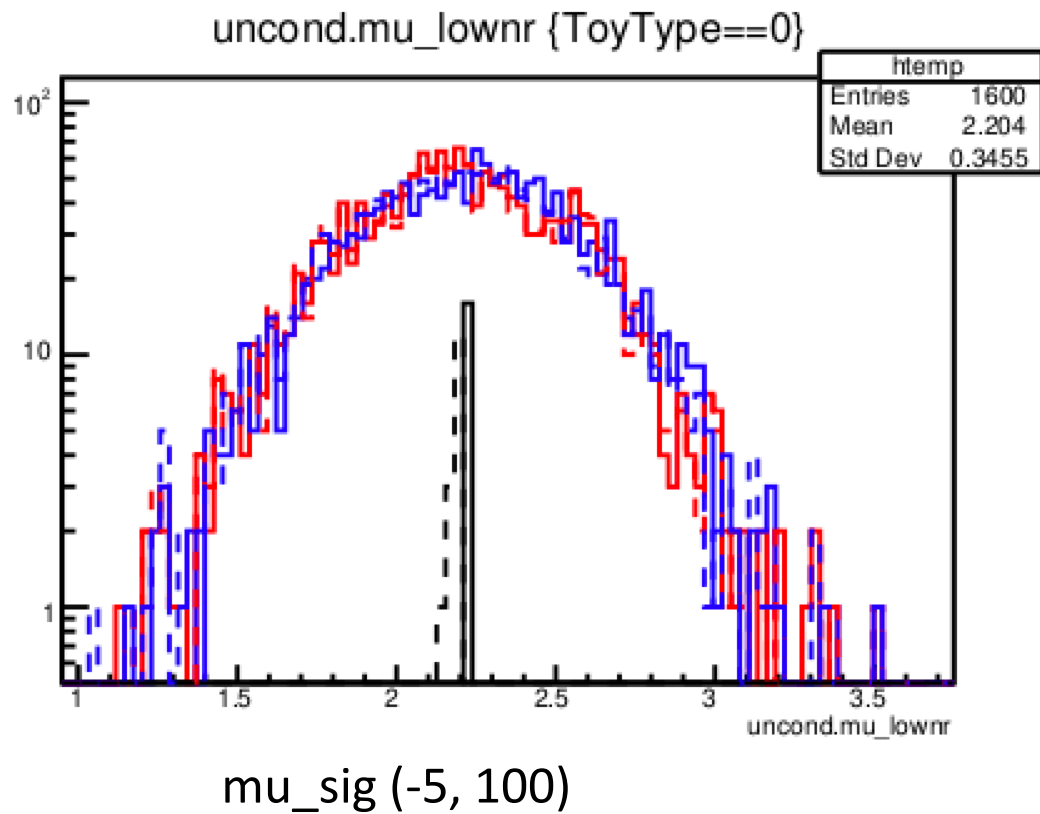


mu_sig (-5, 100)

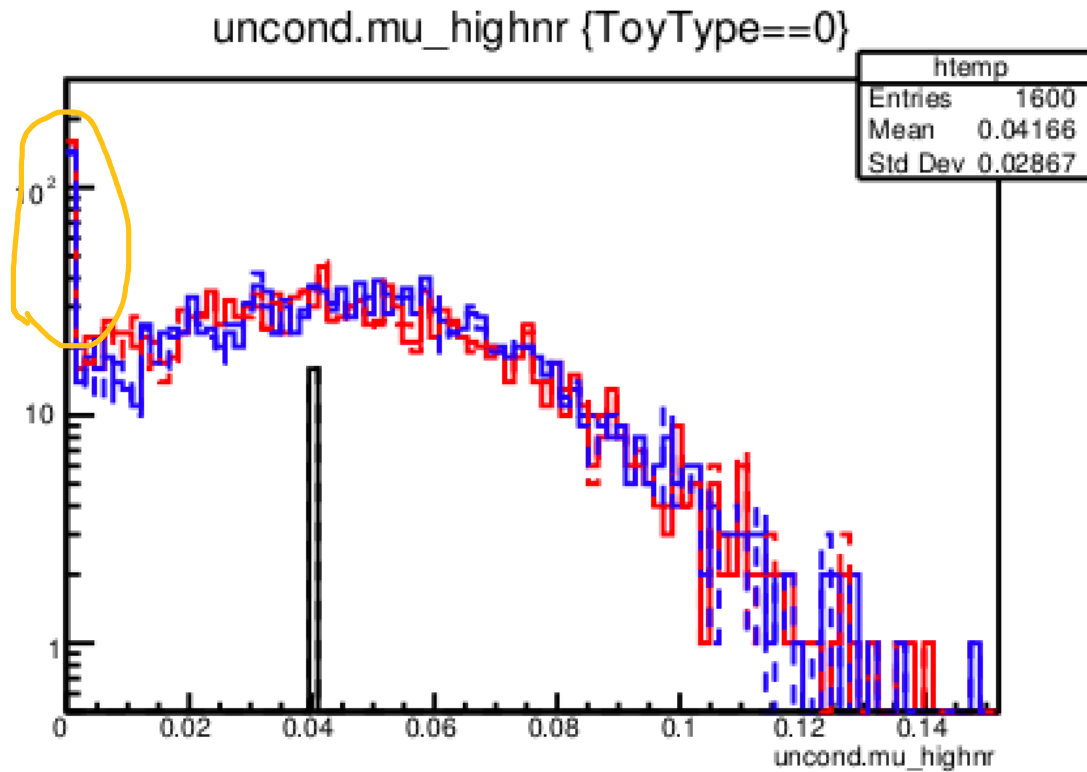


mu_sig (-100, 100)

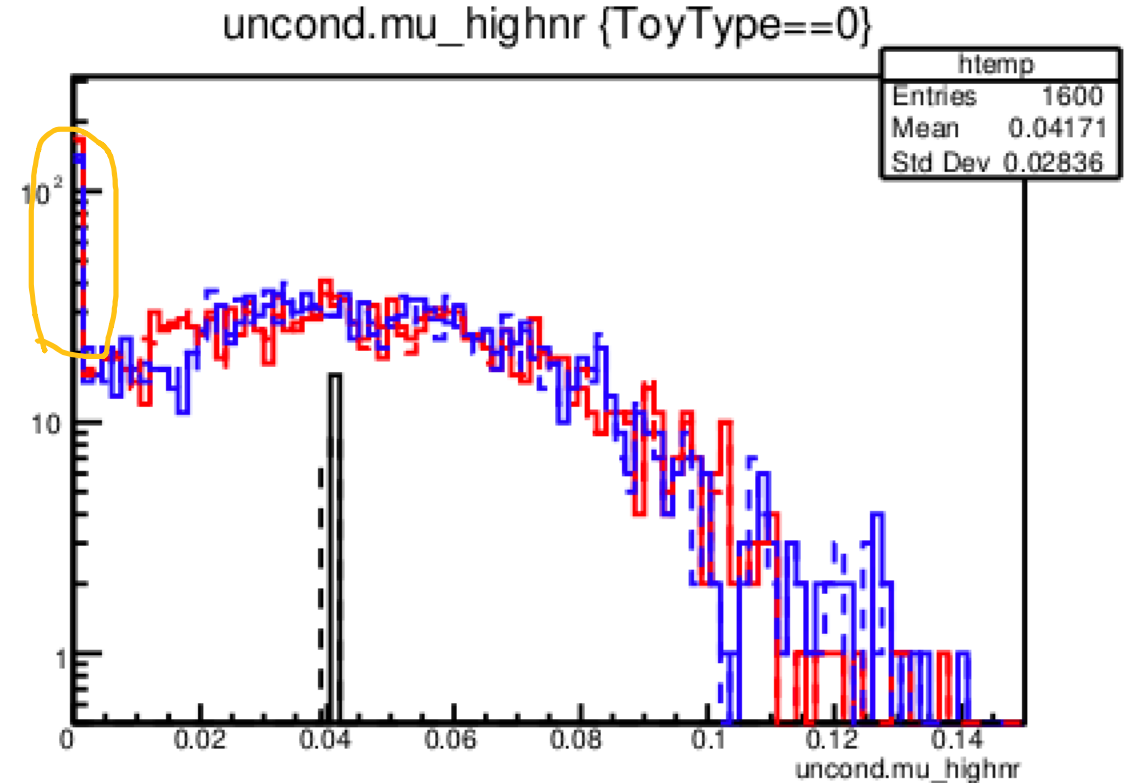
uncond.mu_lownr



uncond.mu_highnr

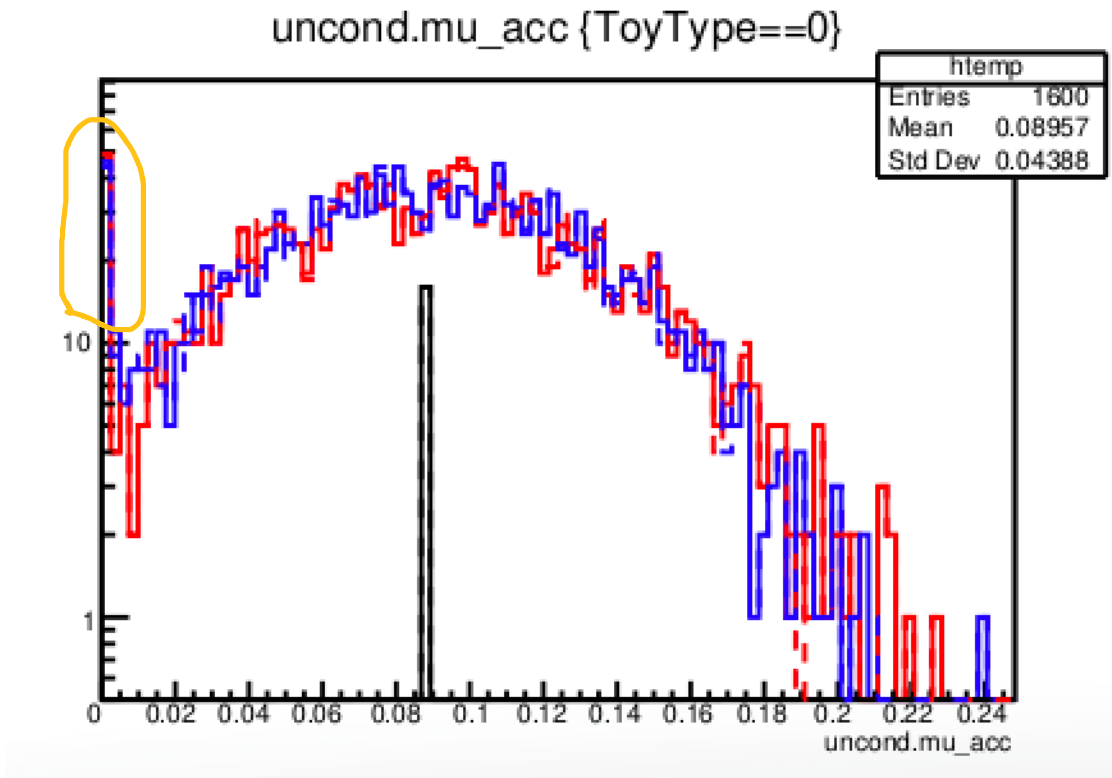


mu_sig (-5, 100)

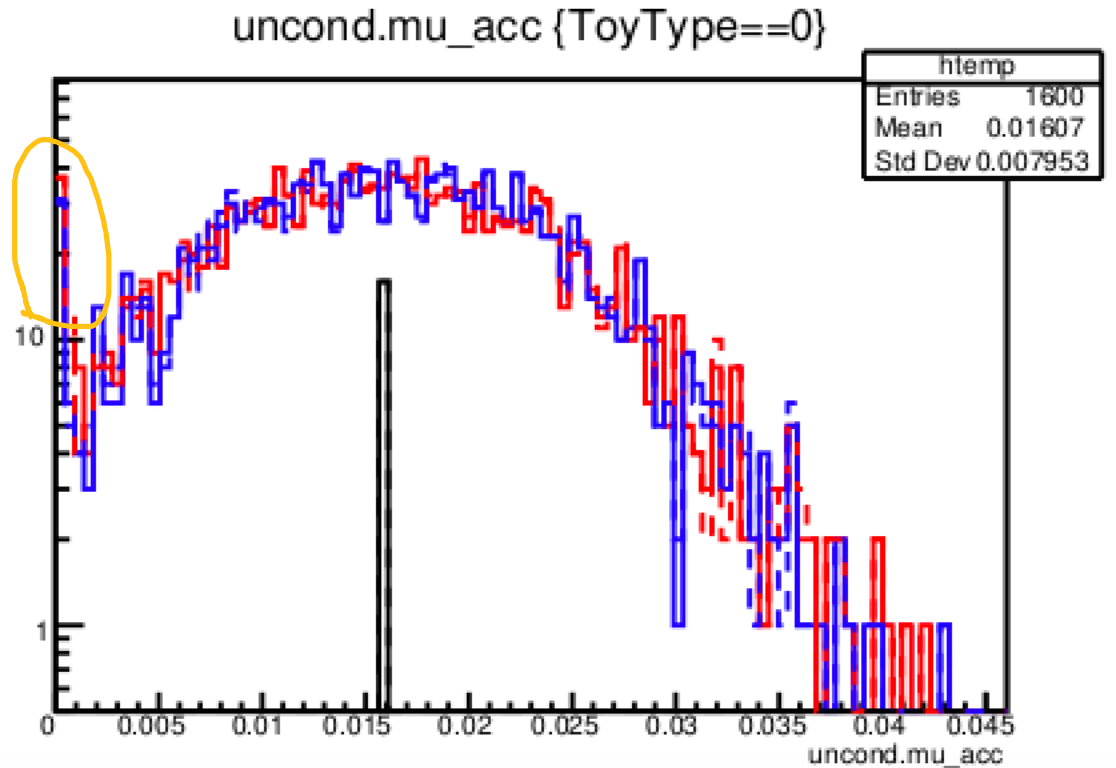


mu_sig (-100, 100)

uncond.mu_acc



mu_sig (-5, 100)

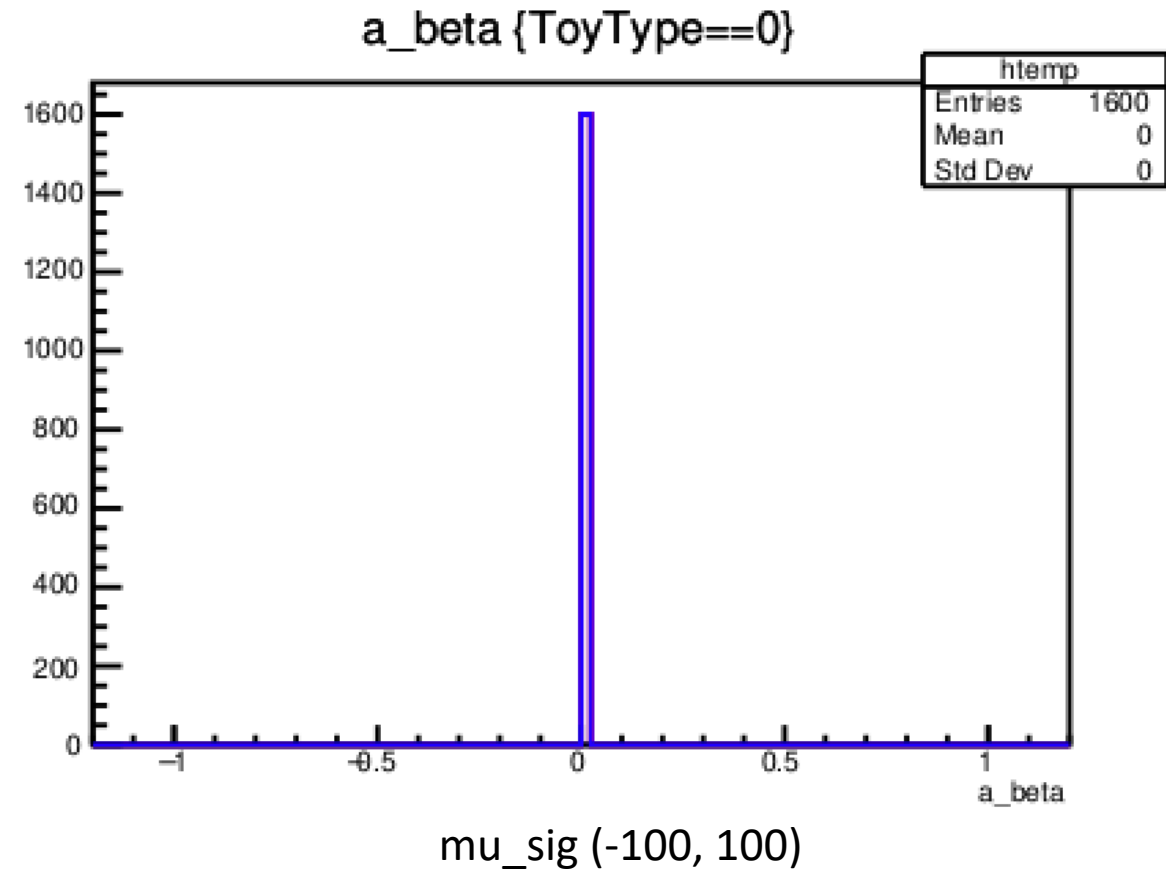


mu_sig (-100, 100)

Error message

```
Error in <TTreeFormula::Compile>: Bad numerical expression : "a_sig"
Info in <TSelectorDraw::AbortProcess>: Variable compilation failed: {a_sig,ToyType==0}
Error in <TTreeFormula::Compile>: Bad numerical expression : "a_sig"
Info in <TSelectorDraw::AbortProcess>: Variable compilation failed: {a_sig,ToyType==1}
a_pp      a_pp
a_vvBB    a_vvBB
Error in <TTreeFormula::Compile>: Bad numerical expression : "a_beta"
Info in <TSelectorDraw::AbortProcess>: Variable compilation failed: {a_beta,ToyType==0}
Error in <TTreeFormula::Compile>: Bad numerical expression : "a_beta"
Info in <TSelectorDraw::AbortProcess>: Variable compilation failed: {a_beta,ToyType==1}
a_lownr a_lownr
a_highnr      a_highnr
a_acc  a_acc
Info in <TCanvas::Print>: pdf file fitResults_glob_lzstats-20200423_workspace_sumbkg_freebetas_wimp_m40.0_SI_lowPOIbound_5-Freq-one_sided-NP_on-run1.pdf has been created
Info in <TCanvas::Print>: pdf file fitResults_counts_lzstats-20200423_workspace_sumbkg_freebetas_wimp_m40.0_SI_lowPOIbound_5-Freq-one_sided-NP_on-run1.pdf has been created
```

Global observable (Beta)



New update

- New workspace:
 - WIMP: 40GeV
 - Fiducial mass (kg): 5600
 - Livetime (days): 24
 - Min POI = 0.1
 - Max POI = 30
 - Toys/point: 100
- Generation errors with the workspace μ_{sig} (-100, 100), no errors with the workspace μ_{sig} (-5, 100)

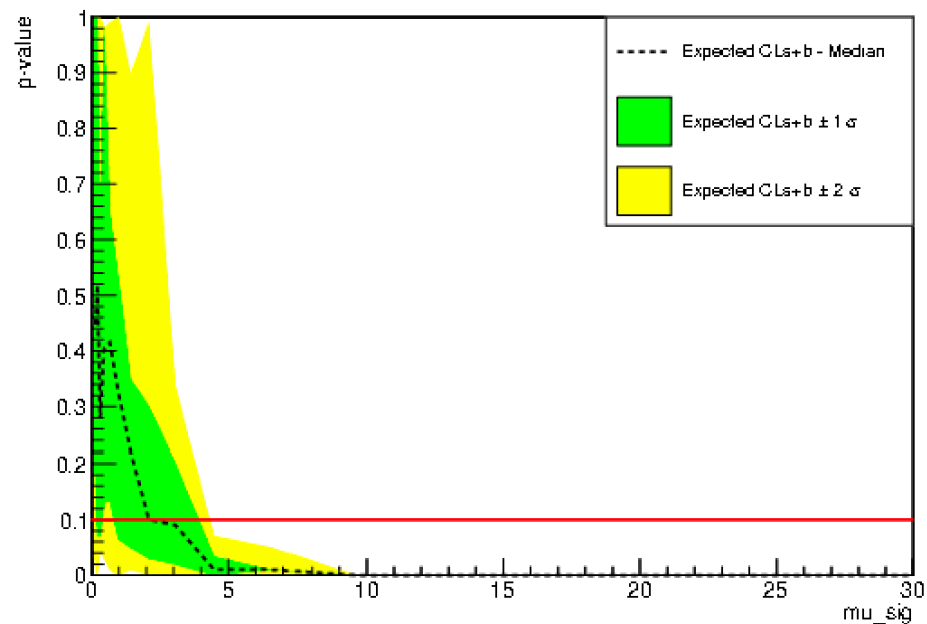
Observed data

- Pass MDC3_unsalted_wimp_data as the observed data instead of “sensitivity”.
- Errors:

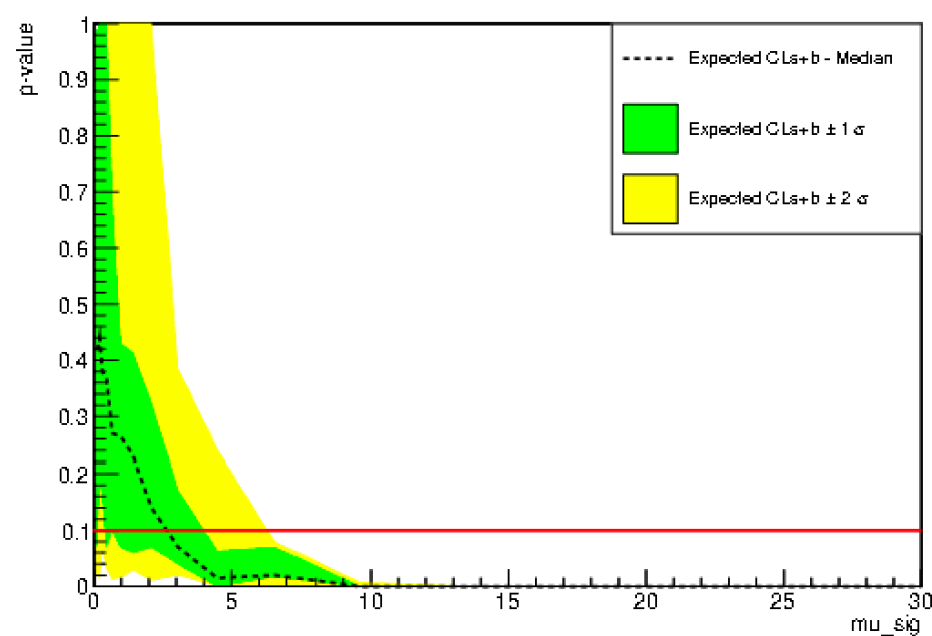
```
[#0] WARNING:Minization -- RooMinimizerFcn: Minimized function has error status.
```

```
p.d.f value is less than zero (-0.000170), forcing value to zero @ !
```

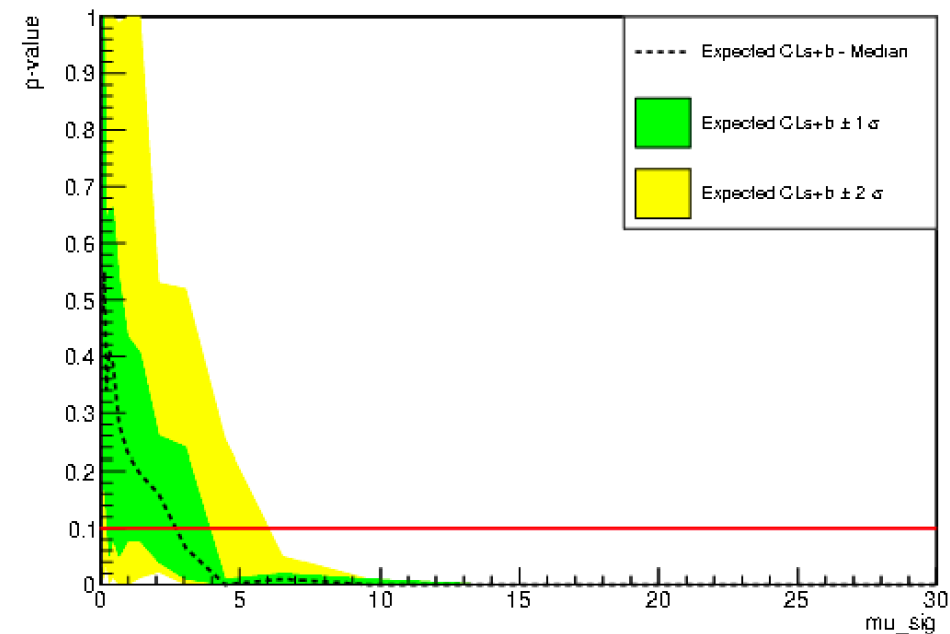
```
[#0] WARNING:Eval -- HypoTestInverterResult::CalculateEstimatedError - cannot estimate the lower limit error
```



Mu_sig: (-5, 100)
 Mu_sig: 2.0947
 Cross section: 2.15e-47



Mu_sig: (-100, 100)
 Mu_sig: 2.9341
 Cross section: 3e-47



Mu_sig: (-5, 100)
 Mu_sig: 3.0639
 Cross section: 3.14e-47
 Unsalted_wimp_data

Error (limit setting)

```
EvalPLR: POI_test = 30, POI_hat = -4.71328, unconstrained ML = -50.3983, constrained ML = -16.9961, TS = 66.8045, time (create/fit1/2) 0, 2.54, 0.79, status (U/C) 0, 0  
[#0] ERROR:Generation -- a_beta::RooRealVar::randomize: fails with unbounded fit range  
----> Trying with Migradimproved  
----> Trying with Scan
```

Each toy run has this error, 3 GeV and 40 GeV



$\mu_{\text{sig}} = (-100, 100)$, negative bound setting

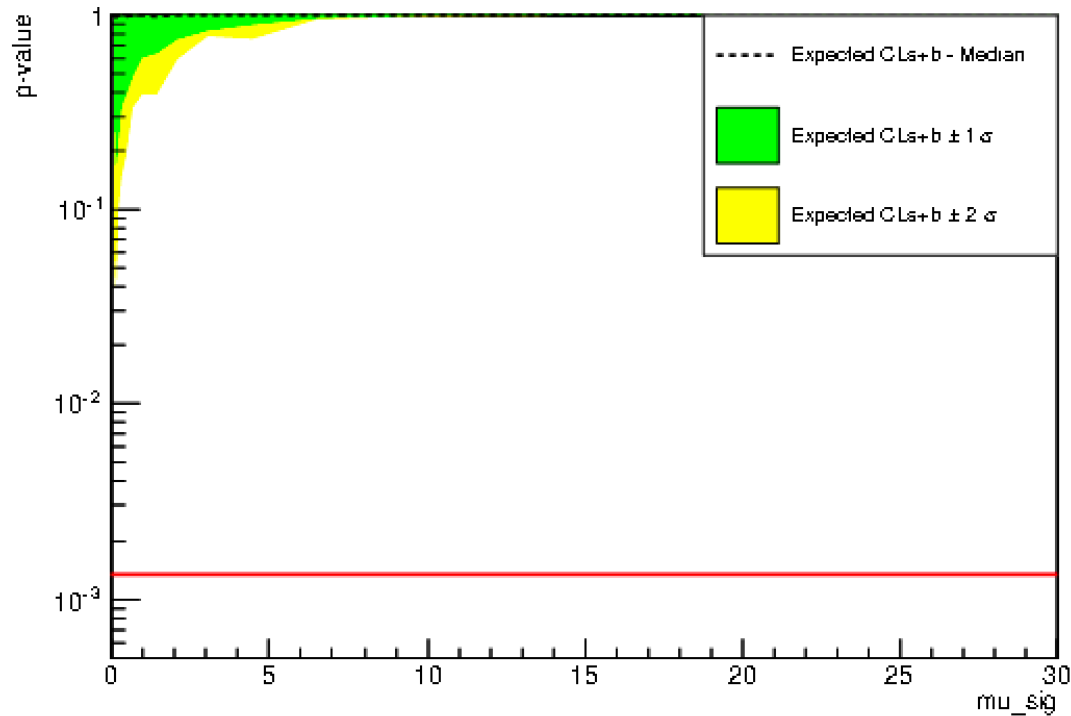


46 GeV w, $\mu_{\text{sig}} (0,100)$



No more this error

Error (Discover)

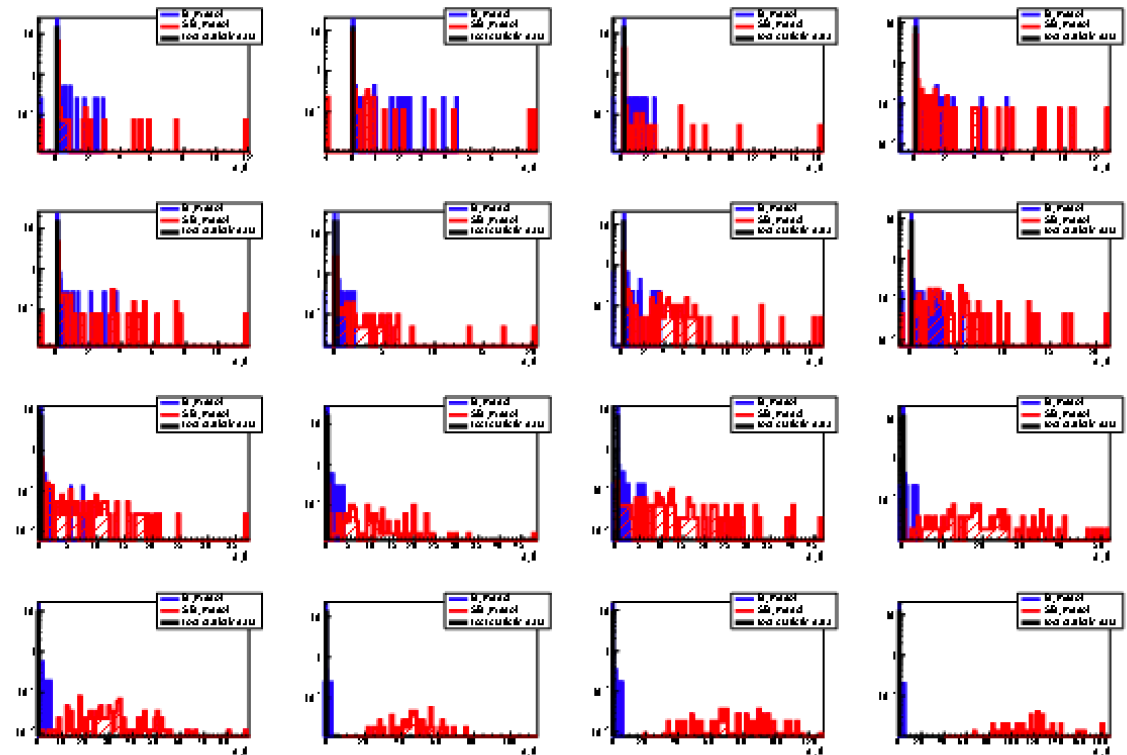


40 GeV, μ_{sig} : (-100, 100)

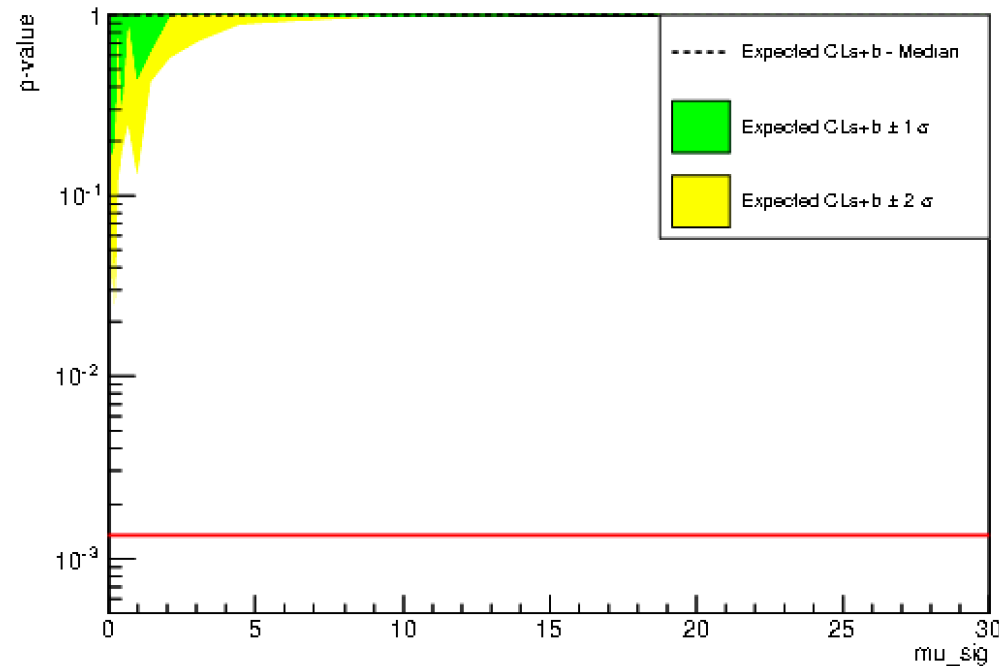
Error: generation

H0: only background, $\mu_{\text{signal}} = 0$

H1: both background and signal



New Error (Discovery)



46 GeV, $\mu_{\text{sig}} (0, 100)$

Fraction: 1839 errors out of 3200 total toy run

```
-> test statistic is negative (fit failed) - setting it to -1  
EvalPLR: POI_test = 0, POI_hat = 8.32396e-06, uncond_ML = -2.8915, cond_ML = -2.  
89173, TS = -1, time (create/fit1/2) 0, 0.3, 0.17, status (U/C) 0, 0  
-> test statistic is negative (fit failed) - setting it to -1
```

To do

- Dig into the cause of errors:
 - Start with 3 GeV, change the μ_{sig} to (0, 100), limiting μ_{sig} to positive value.

Update+To do

- Fitting errors at the low mass (3GeV) for the new workspace → write a code to count the total errors → figure out what causes the error
- Also errors for the 40 GeV.
 - Screen, can only see the parts starting from POI=12.5, no errors after that
 - Run 40 GeV again, errors at lower POI
- Have problem with the master branch on LZStats

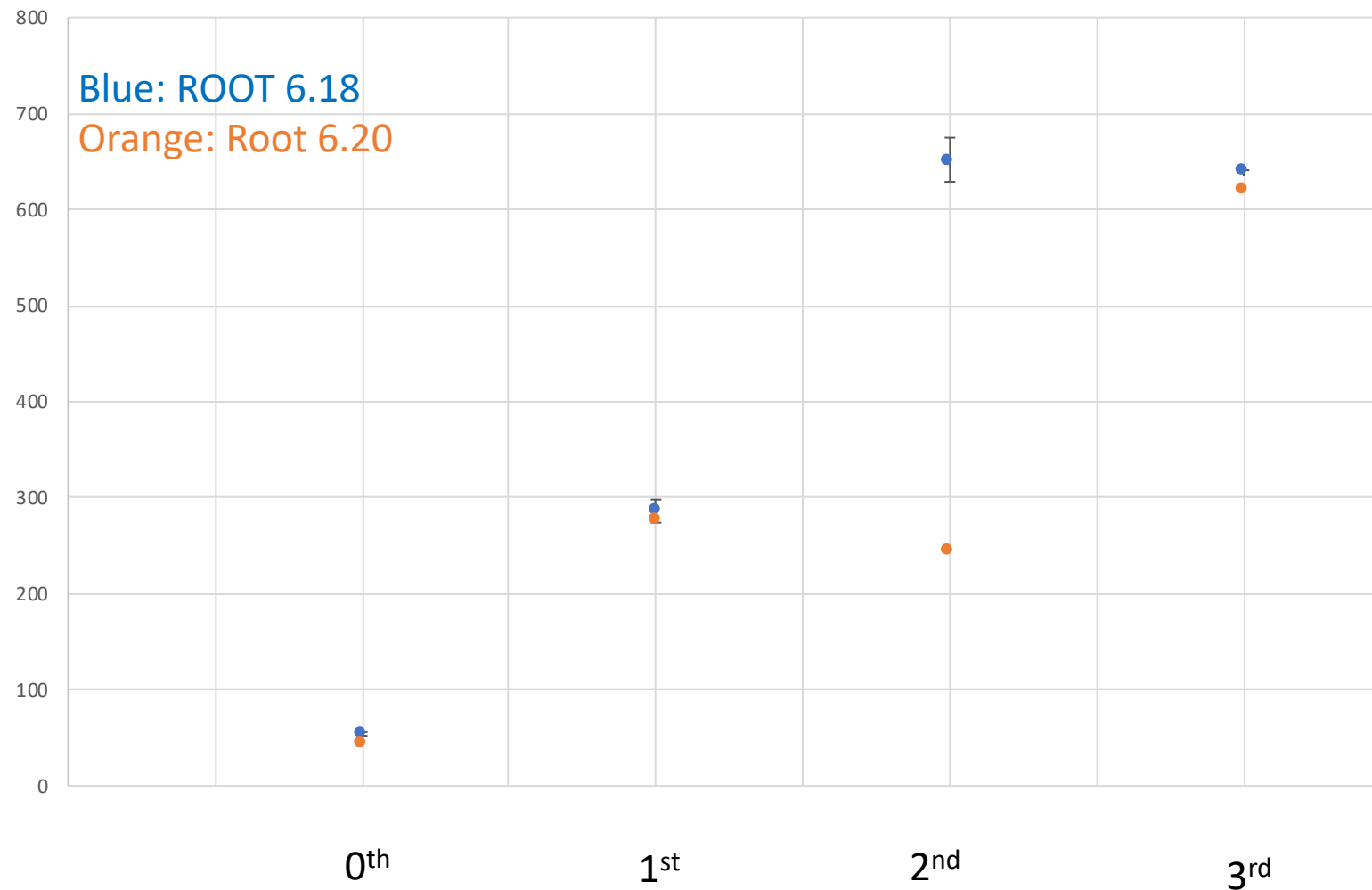
New update

- PdfMaker: add YAML configuration to WIMP & Background PDF
 - Do not need to change parameters in LZNESTUtils and source setup & build everytime changes a parameter
 - Tested n_sample parameter in YAML config, it works as expected!
- LZStats:
 - When source setup.sh, it changes ROOT to version 6.16
 - Works on Jonathan's git branch (jn_ROOT6.20.00) to test how ROOT 6.20 affect the run time of different interpolation order

ROOT 6.18 & 6.20

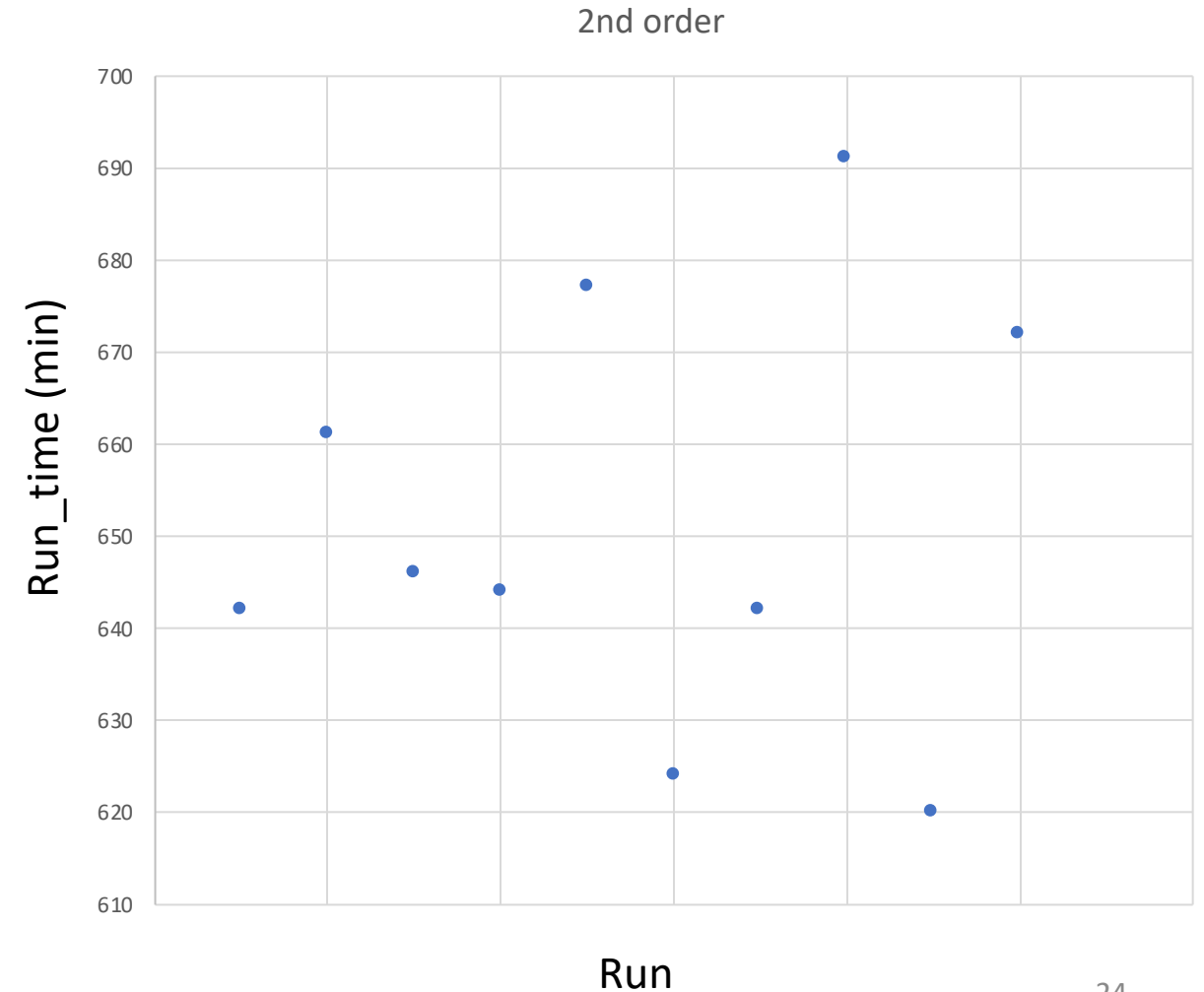
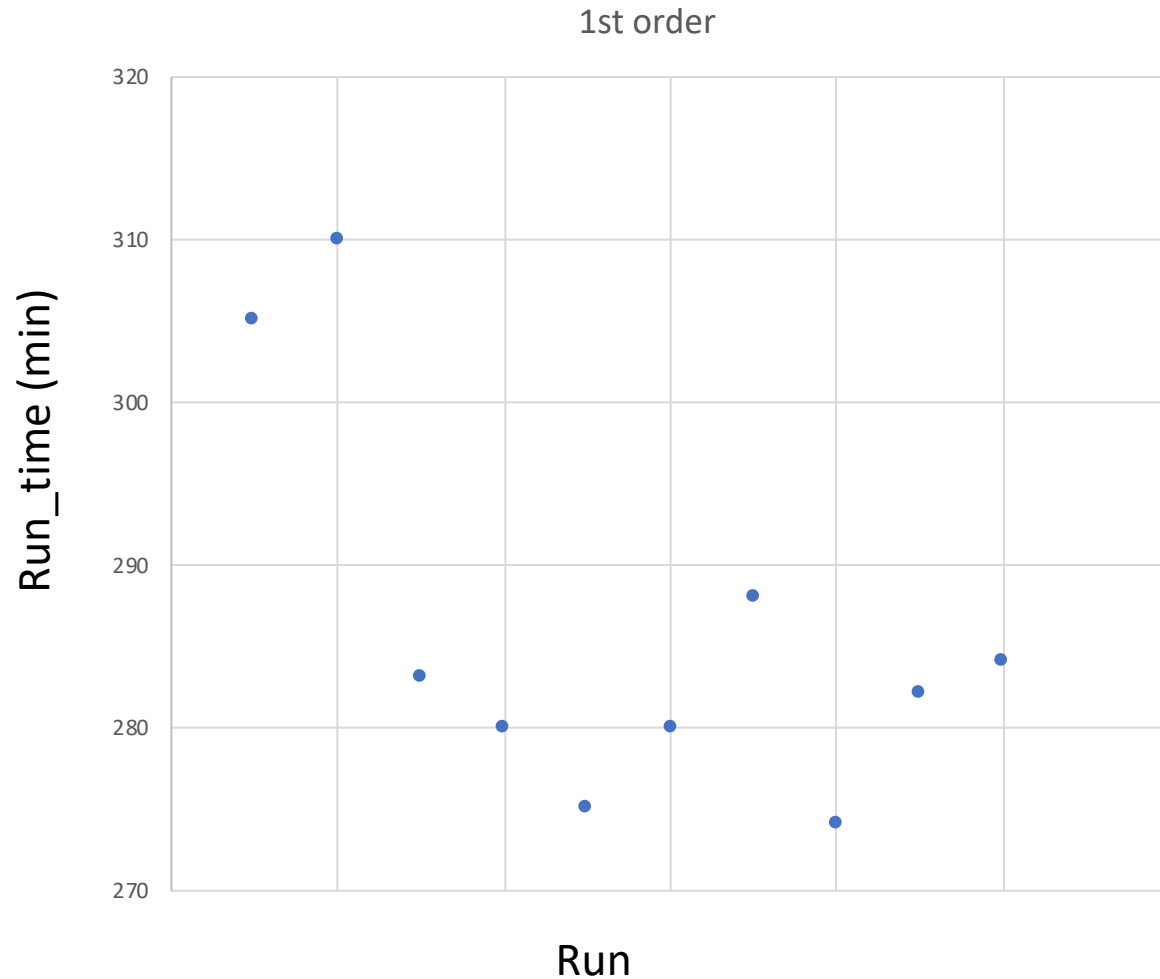
	0th	1st	2nd	3rd
run_time(min)	53.9 (10 runs)	286.1 (10 runs)	651.9 (10 runs)	641
STD	1.9	12.1	22.9	
run_time(min)	44	278.5 (2 runs)	245?	621

run_time vs. order



1st & 2nd orders

Submit 10 runs at the same time, some runs are not using 100% CPU at the beginning -> higher run time for these specific runs -> higher STD

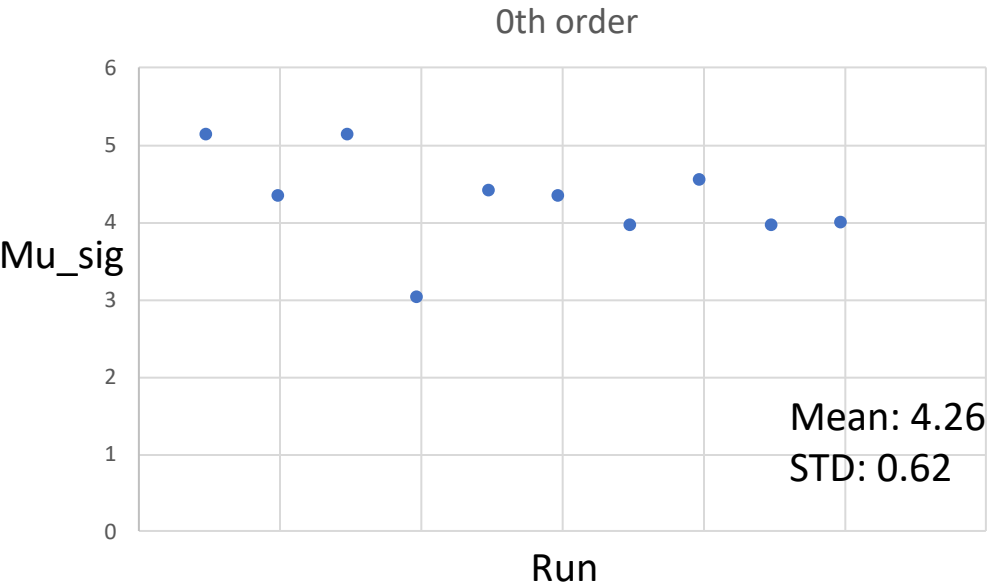


Previous Slides

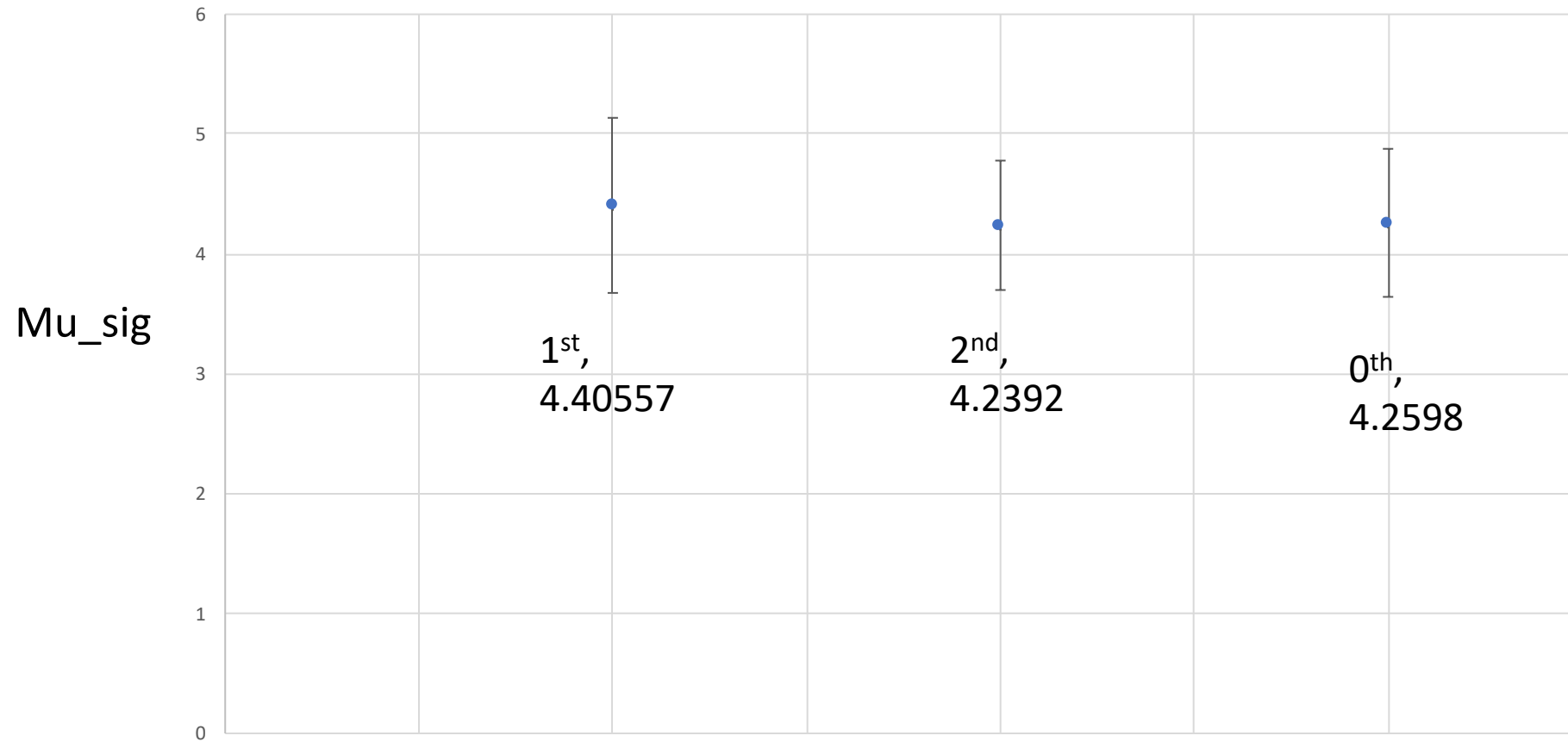
Interpolation

WIMP mass: 10GeV

N_sample: 10^6



mu_sig vs. order



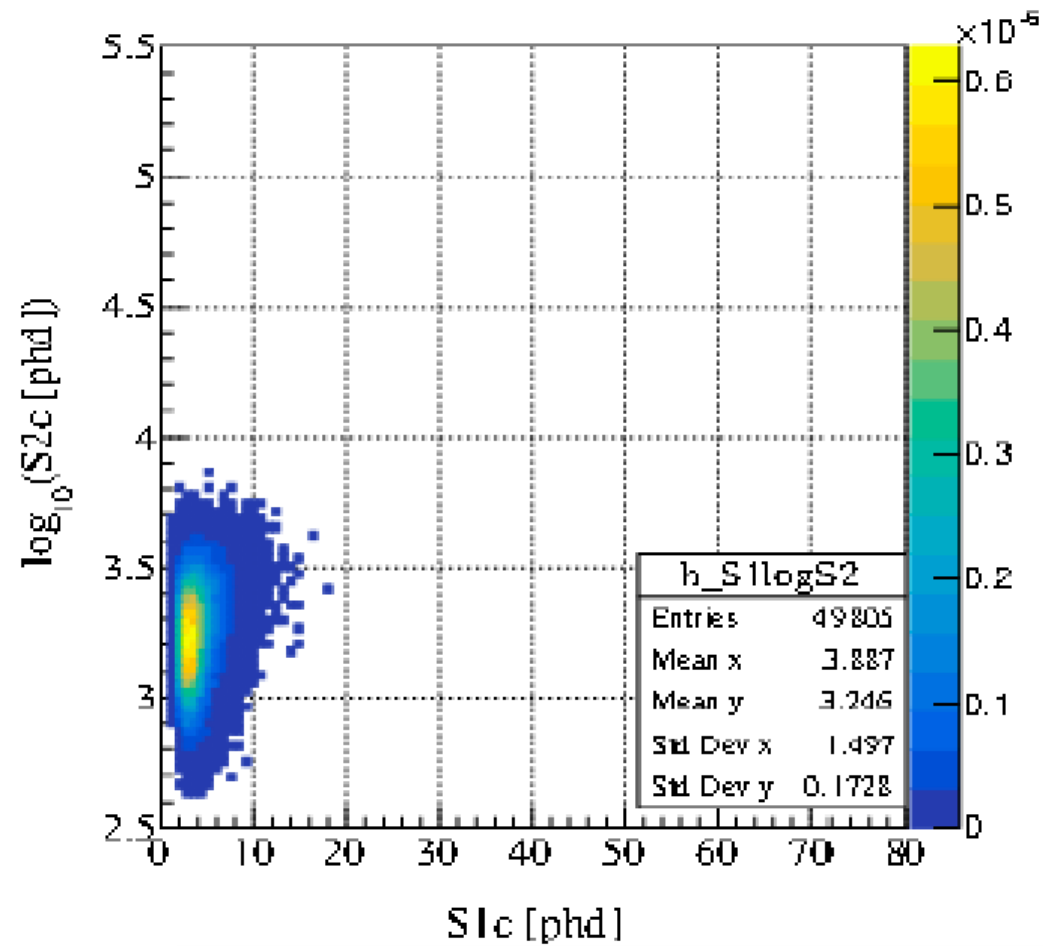
What's new?

- New update
- Try different interpolation order 0~3
- Organize the output

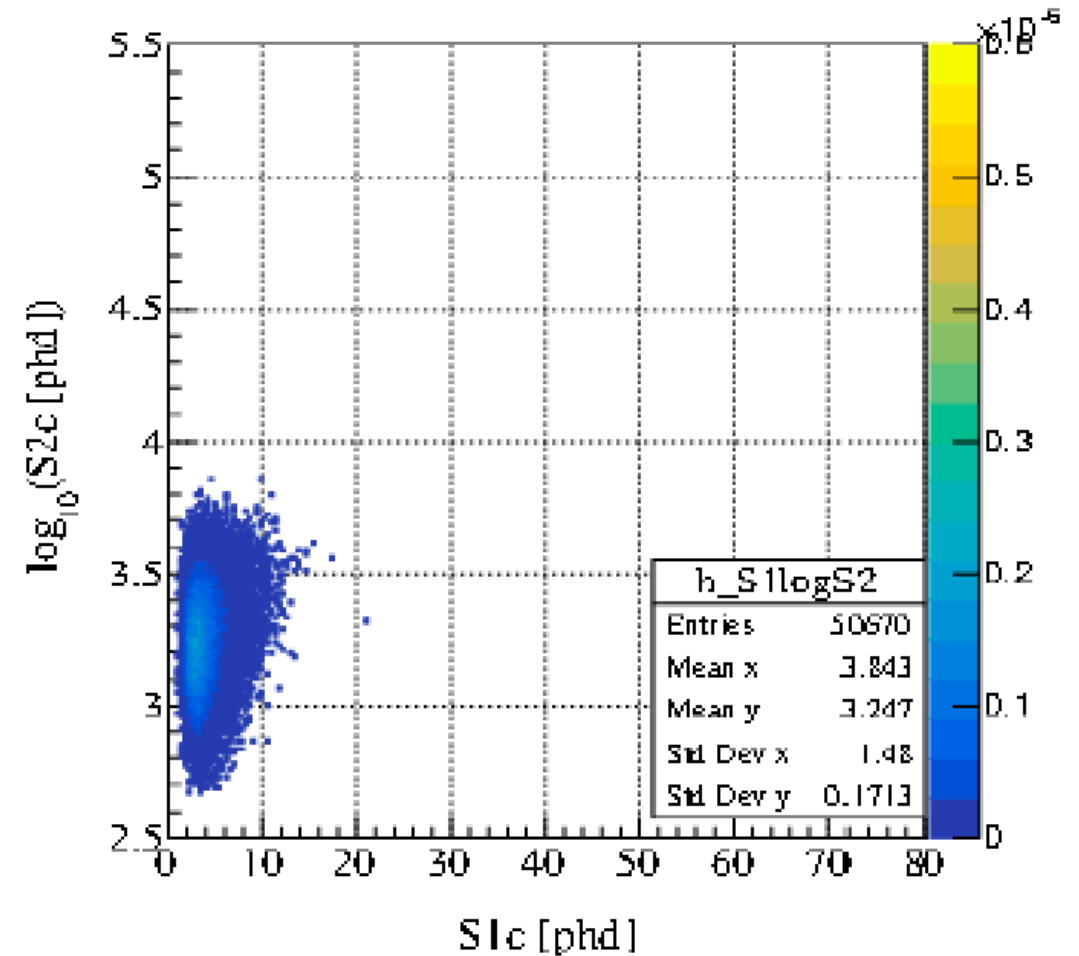
New update (Detector)

Parameter	Before	After
g1	0.118735	0.150
g1_gas	0.1018	0.1186
eLife_us	850	944.56
r_max_fiducial	688 cm	688 mm
z_max_fiducial	1330 cm	1330 mm

WIMP signal PDF



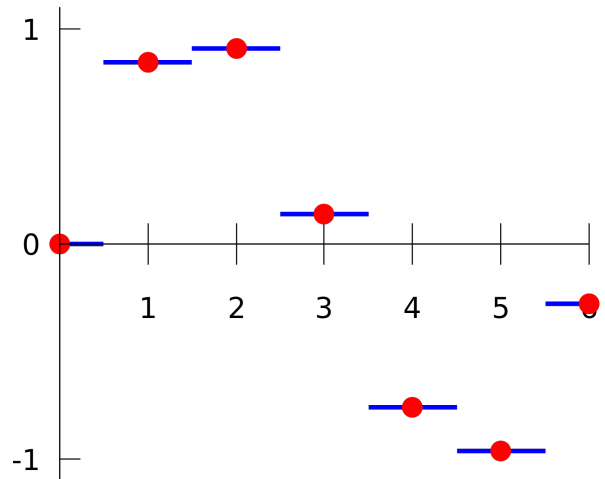
Before



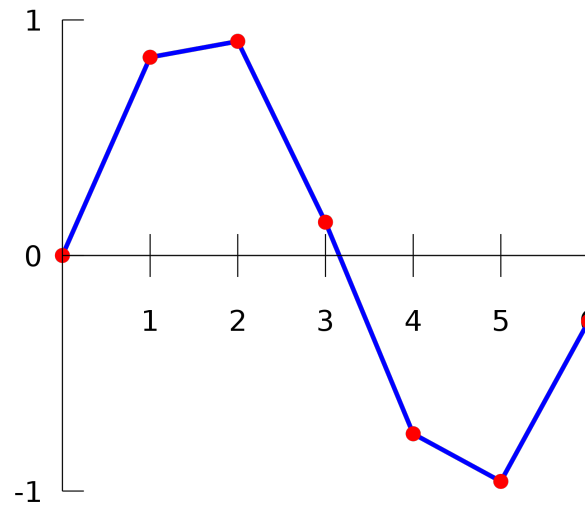
After

Interpolation Method

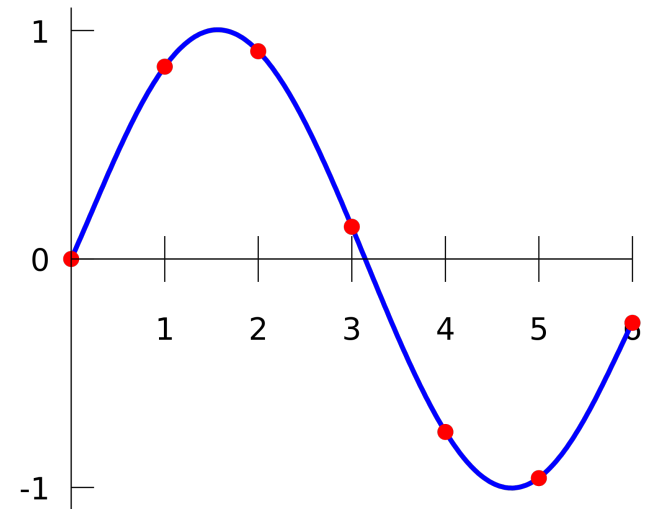
- Construct new data points within the range of know data points (discrete set) → bring up the bins surrounded by 0 events
- Change interpolation order 0~3(in the code to create workspace)



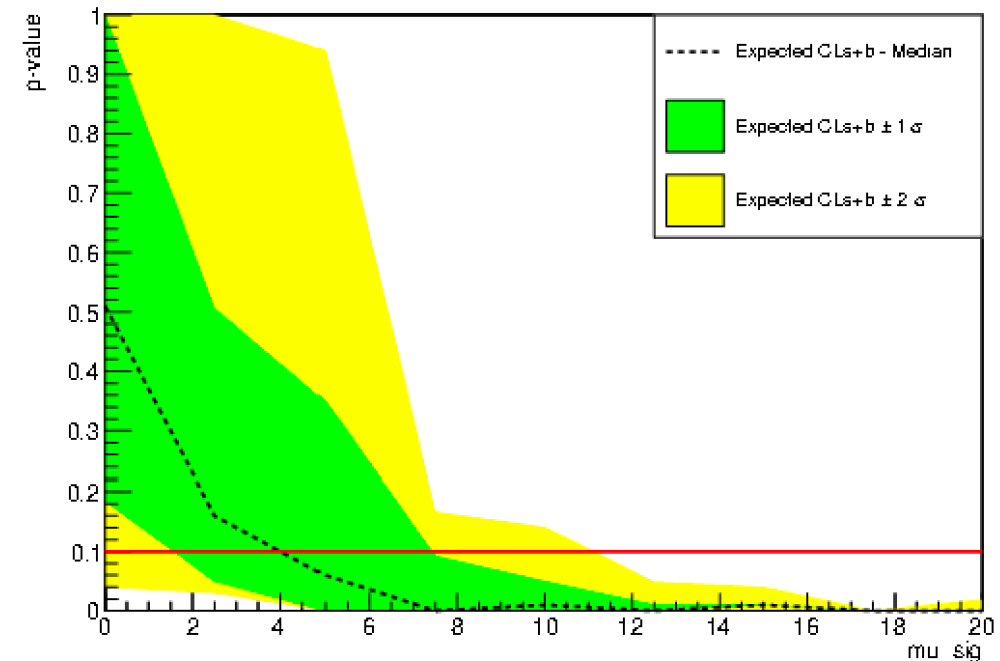
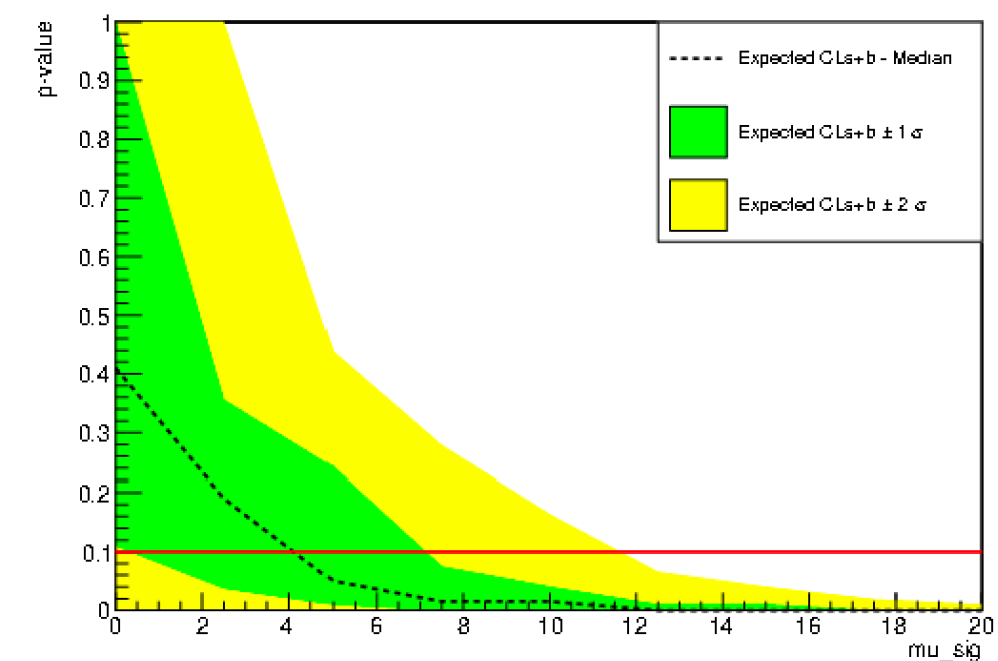
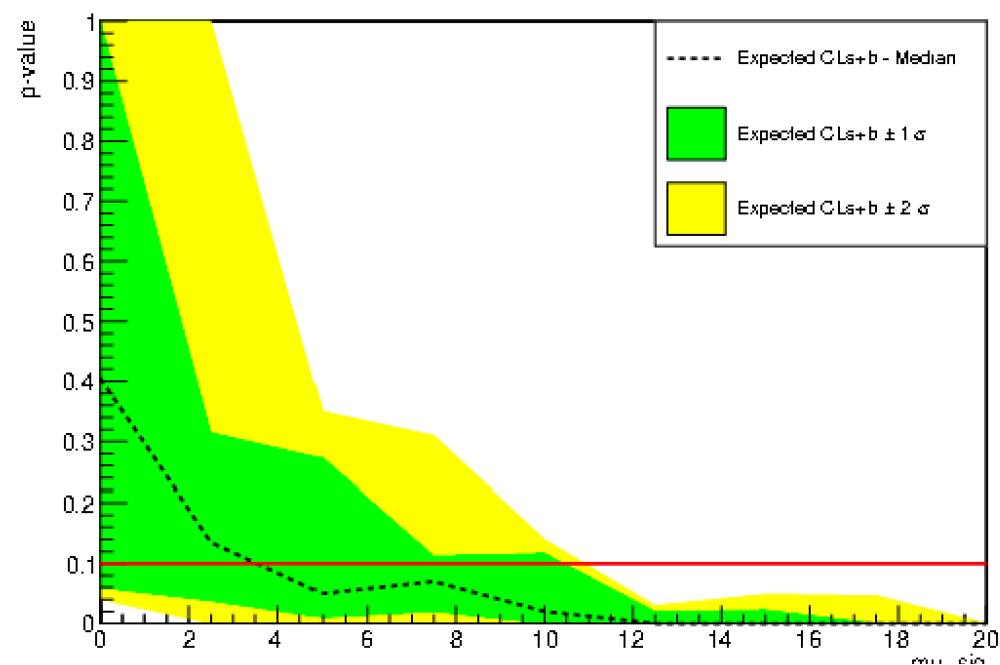
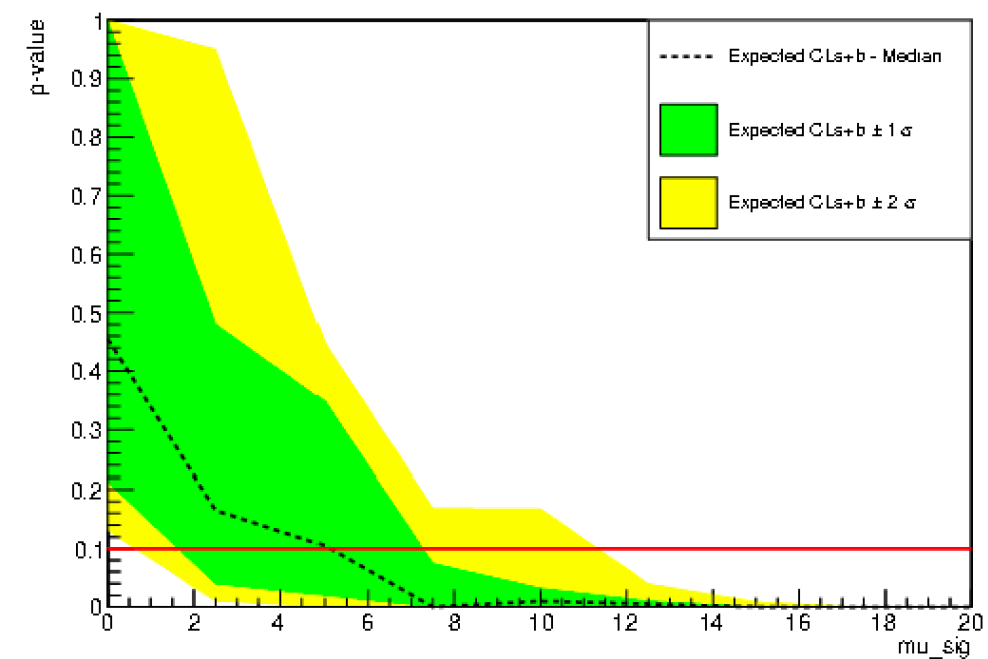
0th

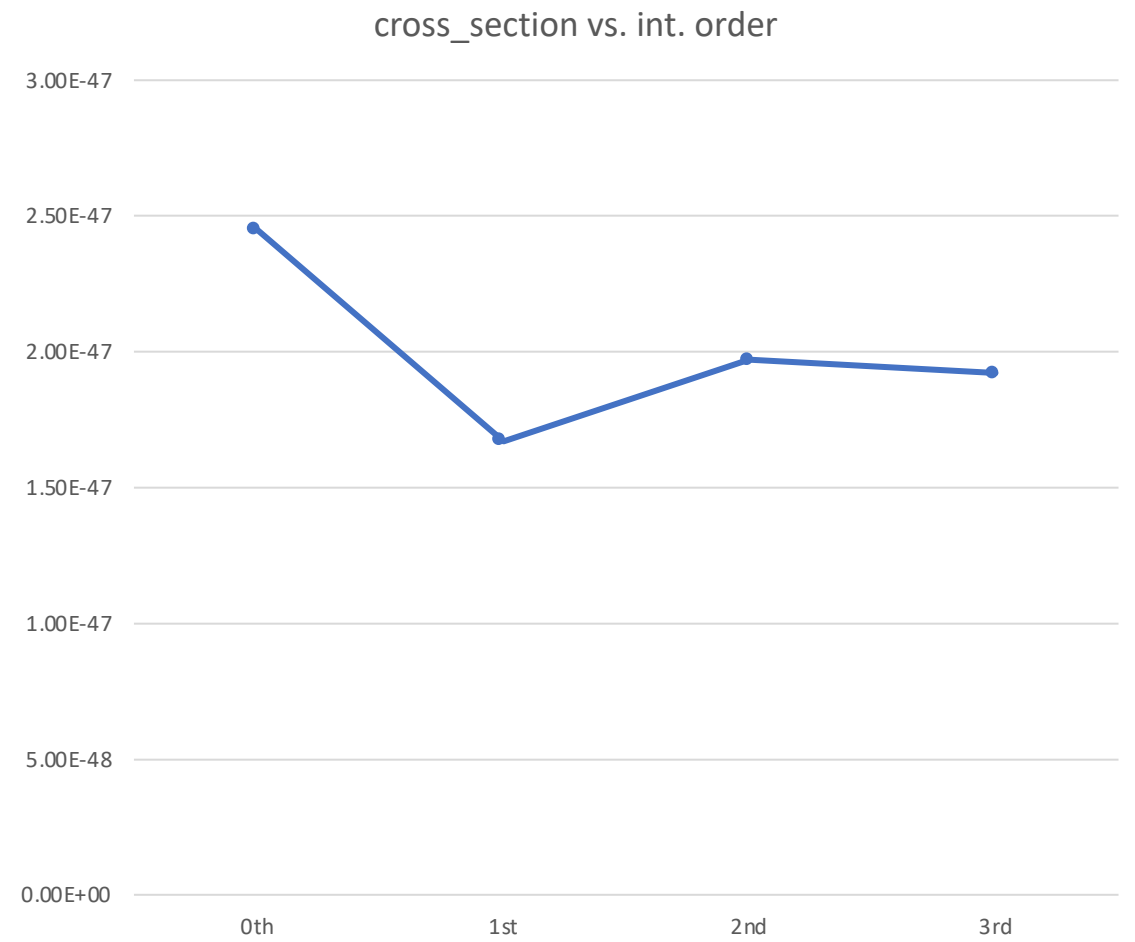
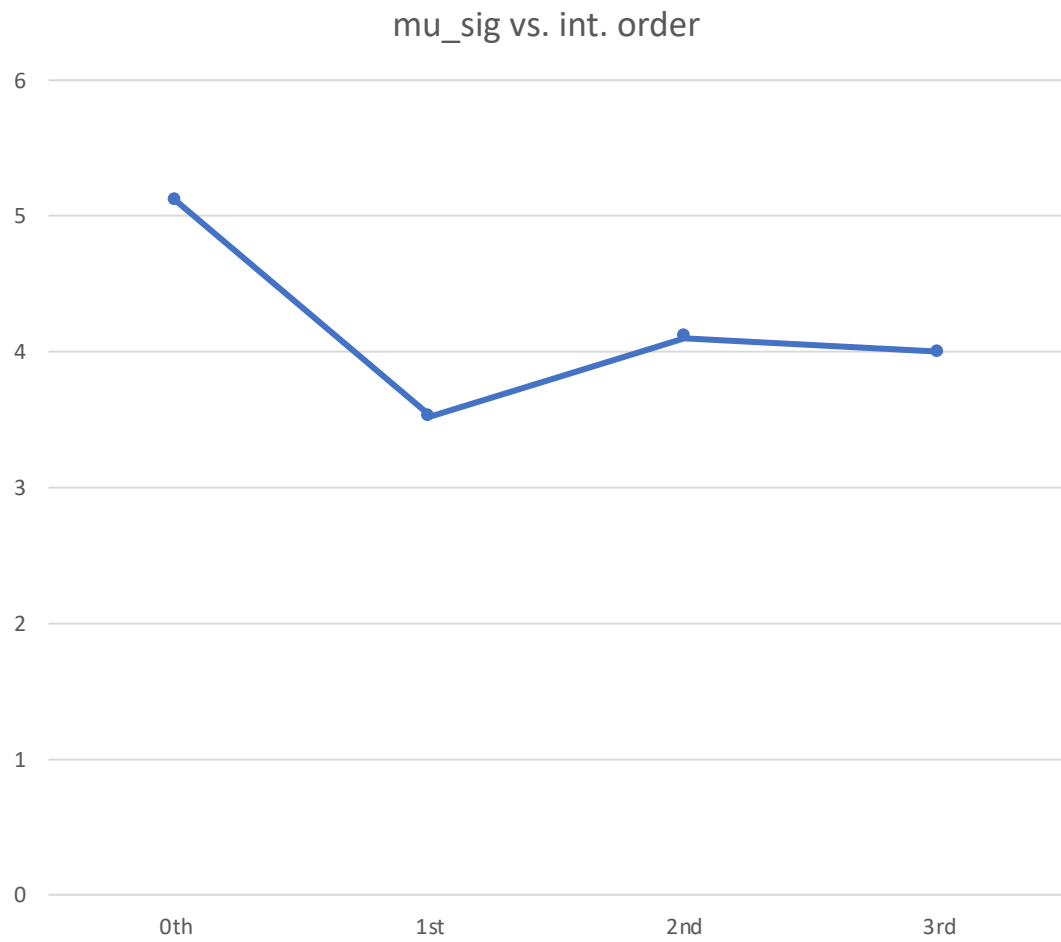


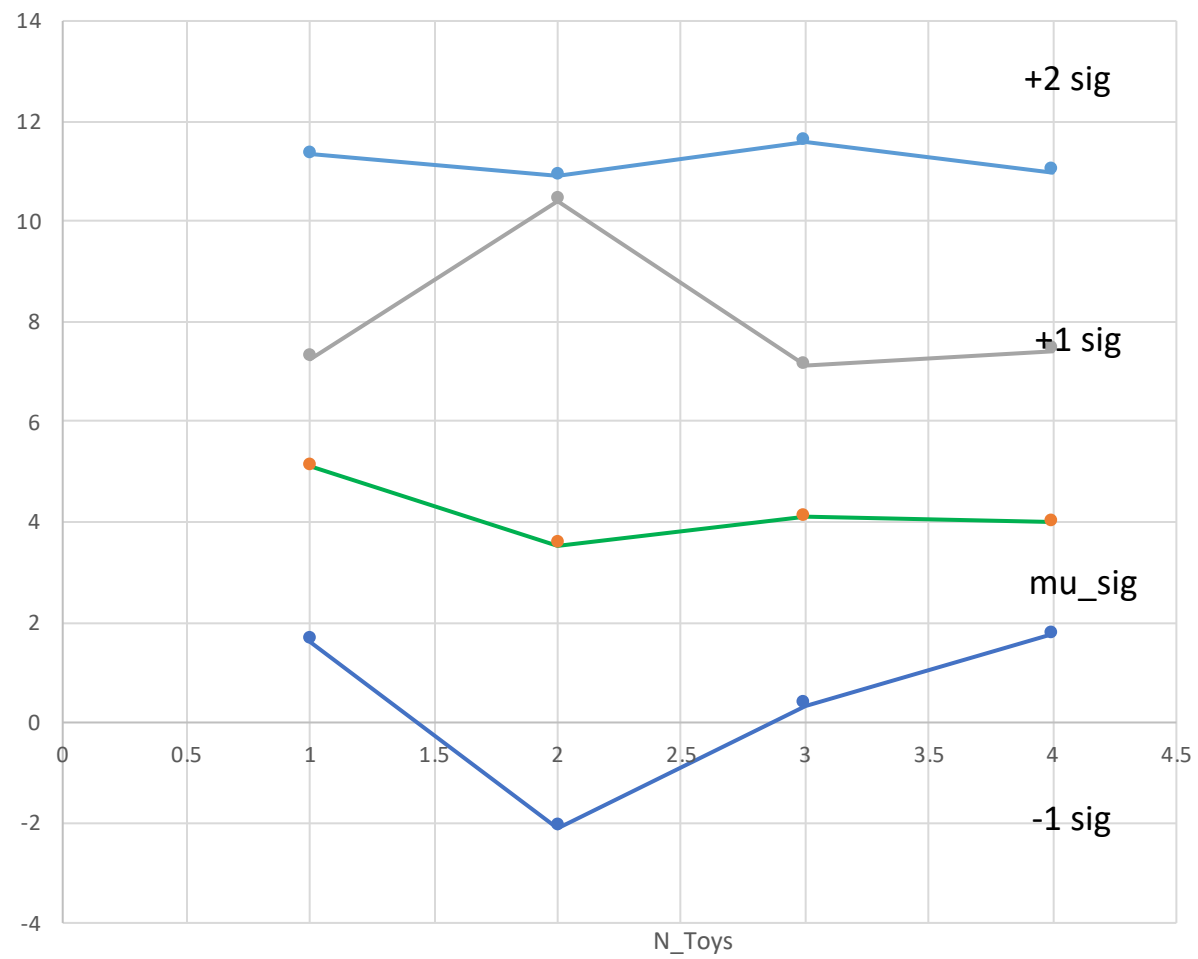
1st



2nd?



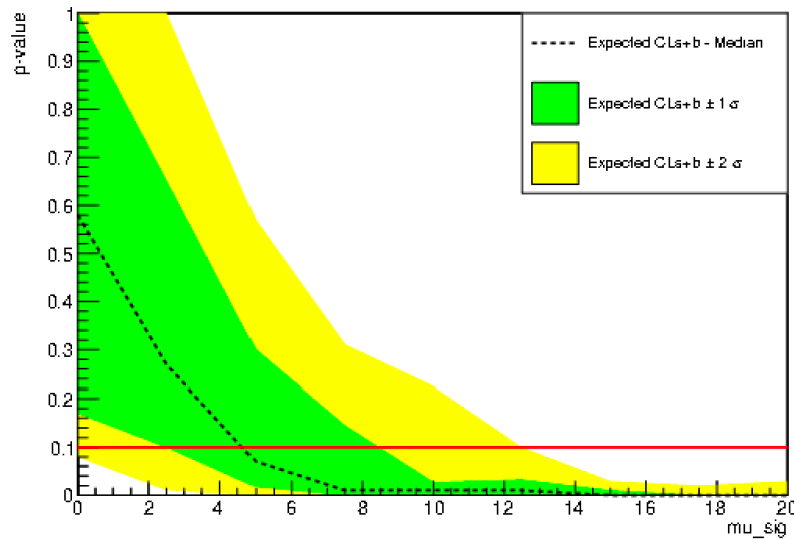




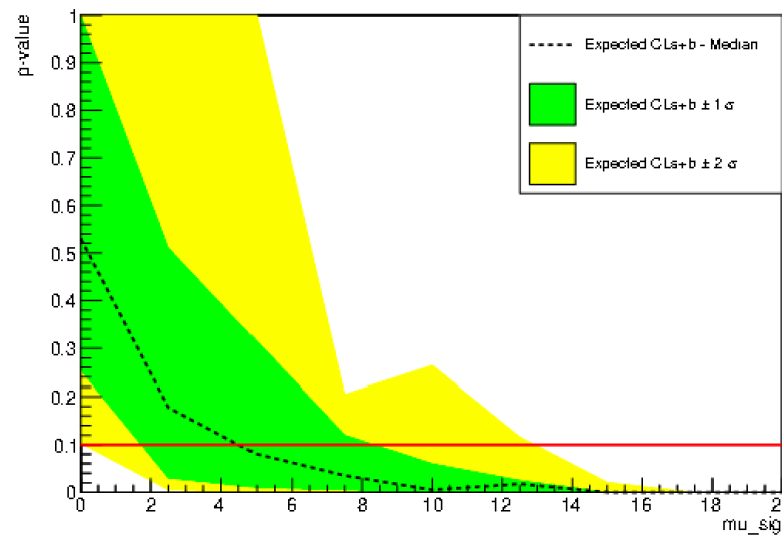
What's new?

- Try different # of Toys.
- Plot μ_{sig} , -1 sig, +1 sig, -2 sig, +2sig, cross section vs. NToys

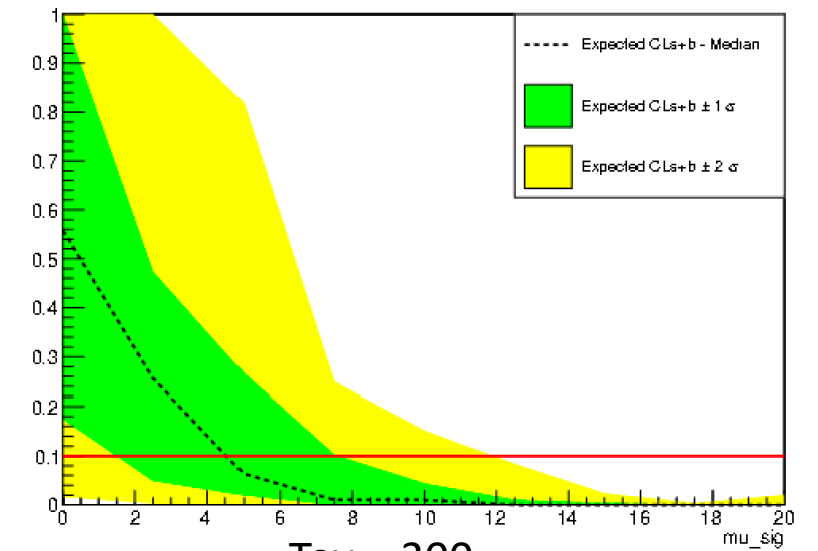
$g1 = 0.118$, 10^7 sample size



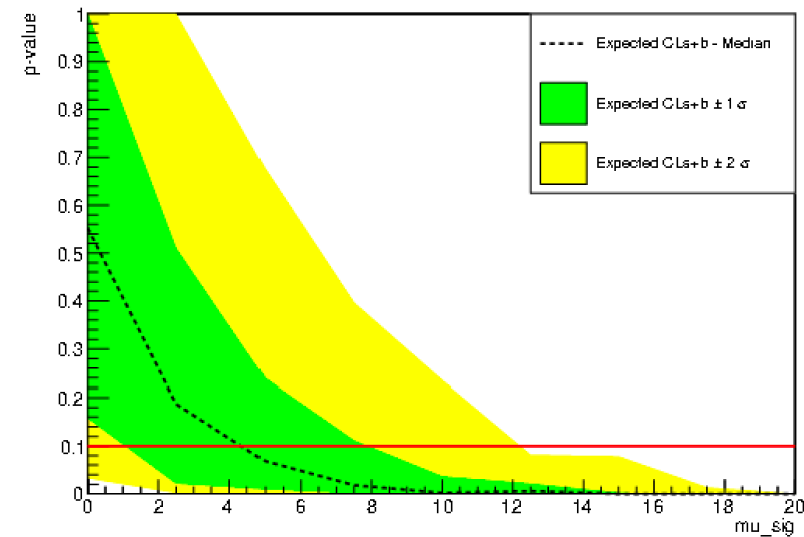
Toy = 100



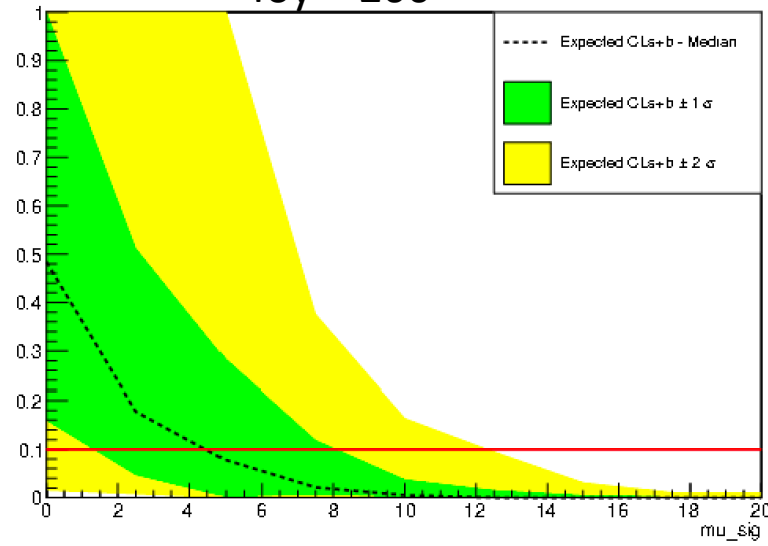
Toy = 200



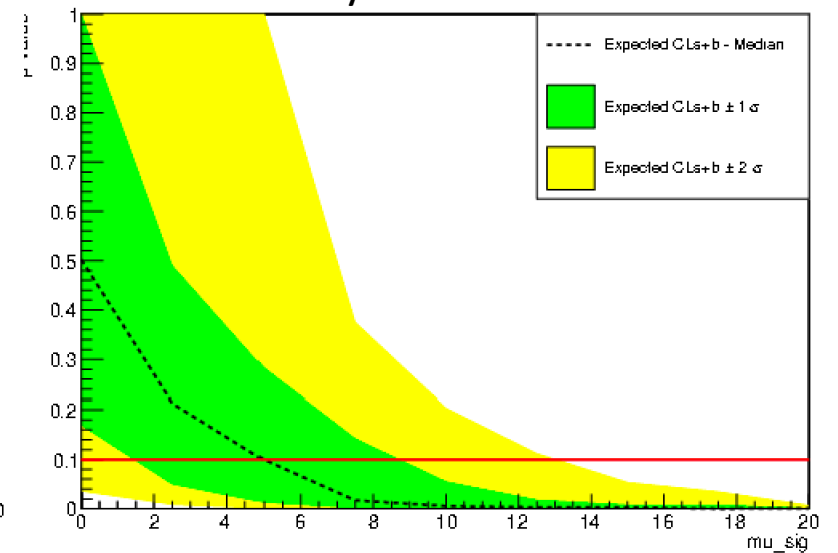
Toy = 300



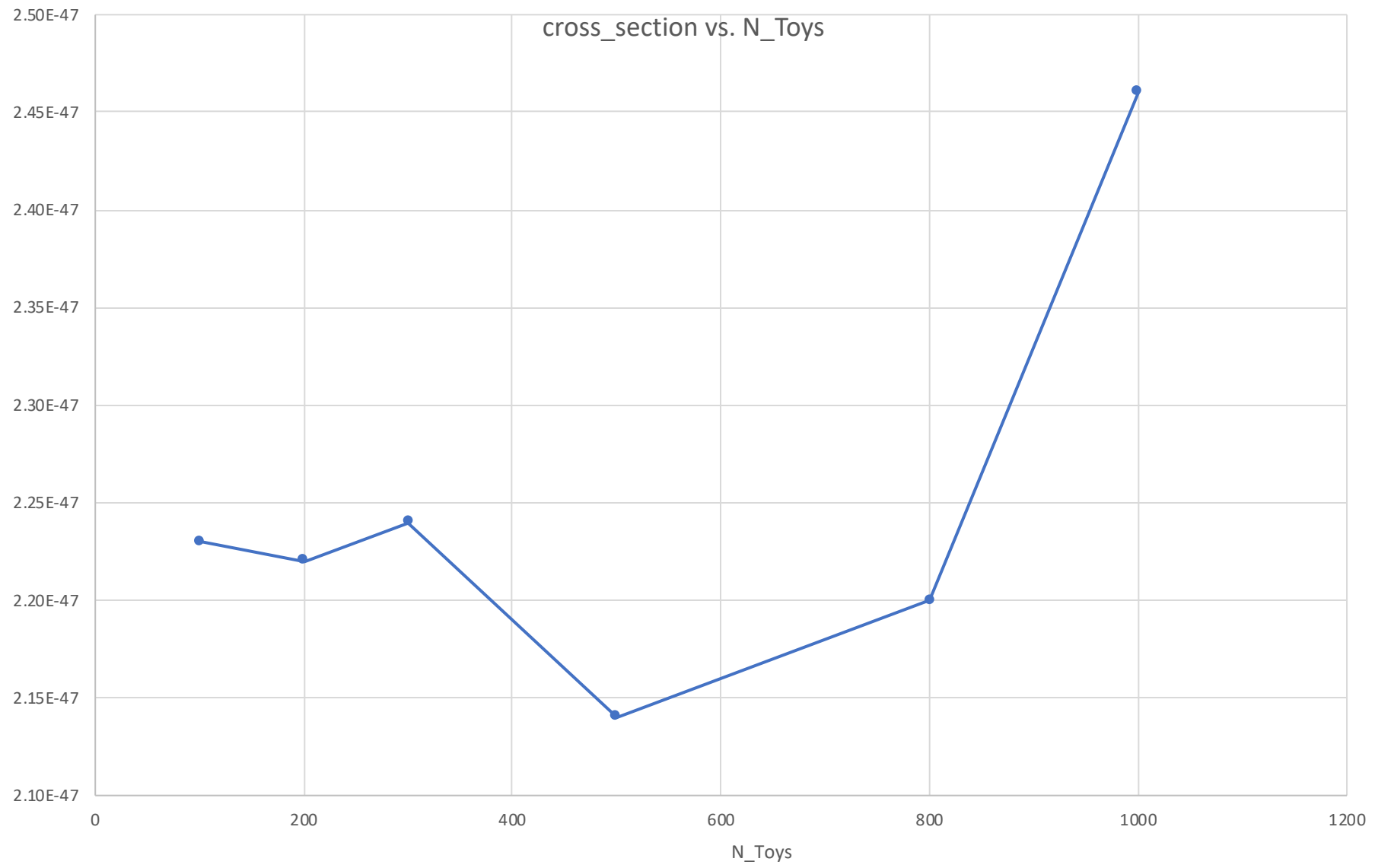
Toy = 500

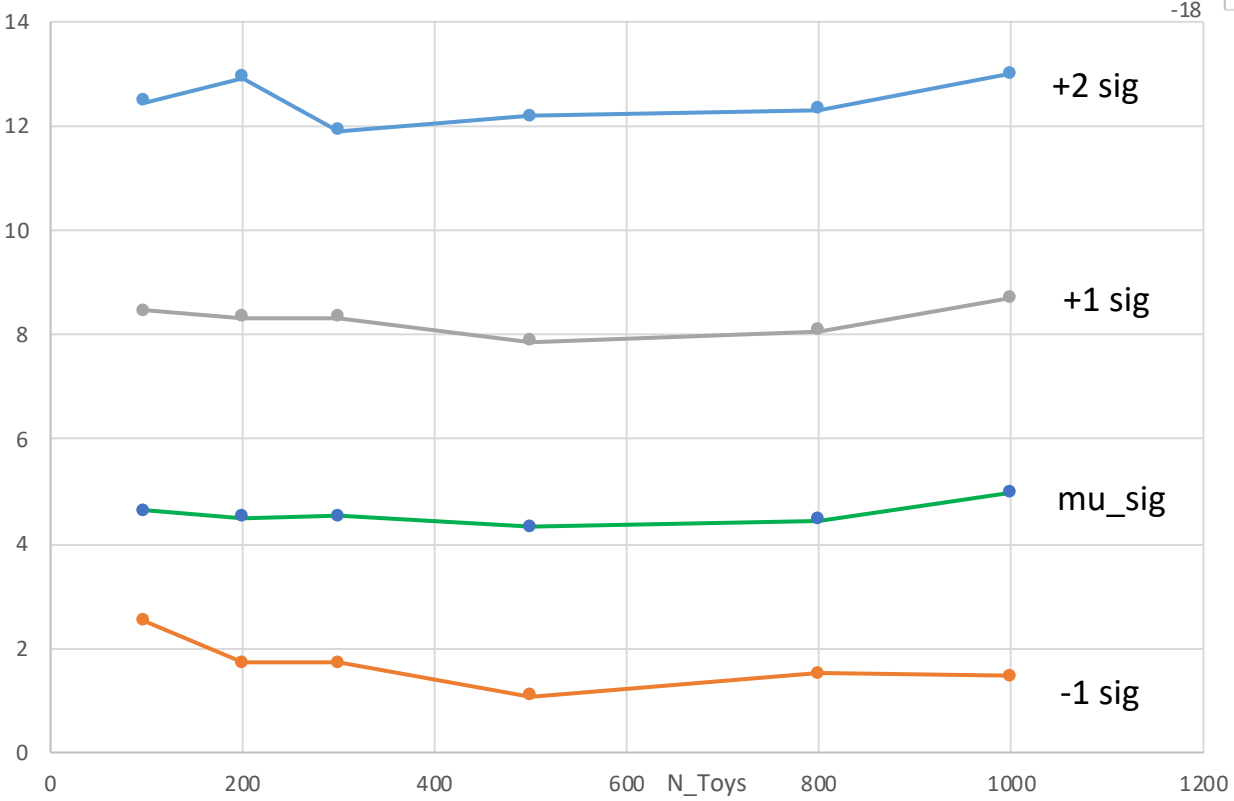
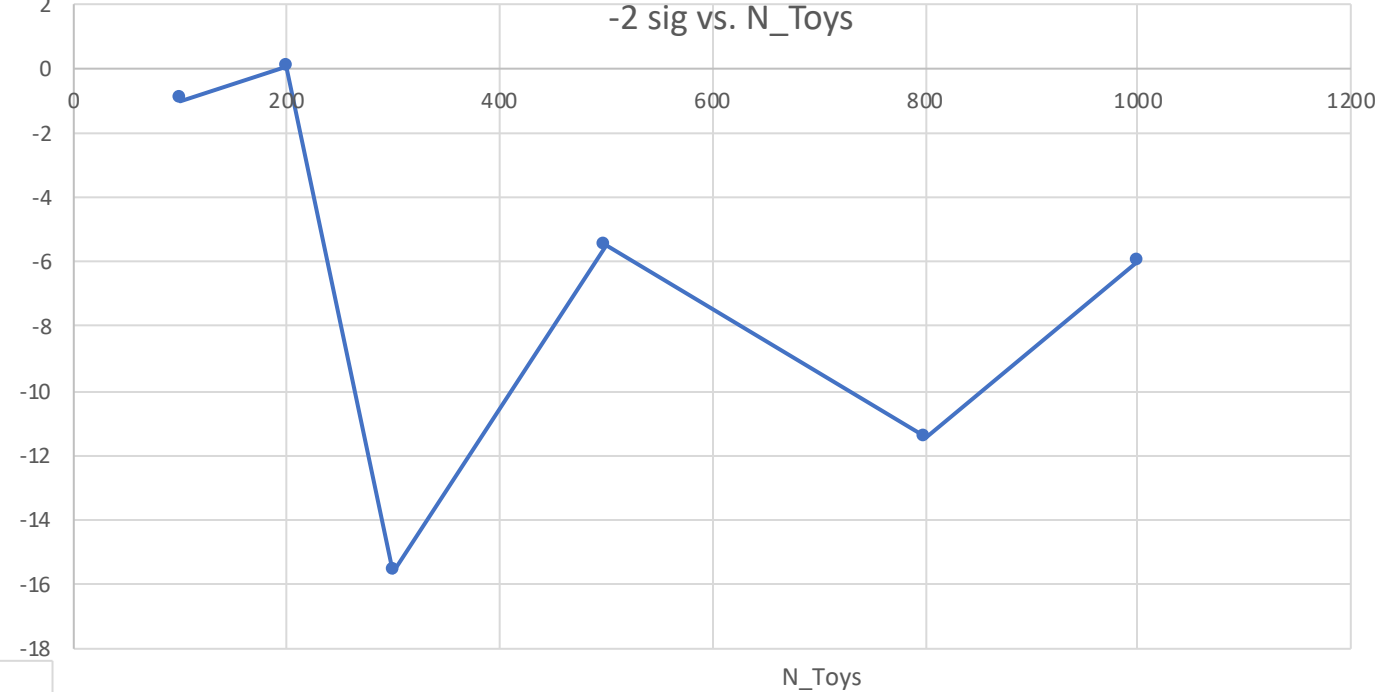


Toy = 800



Toy = 1000



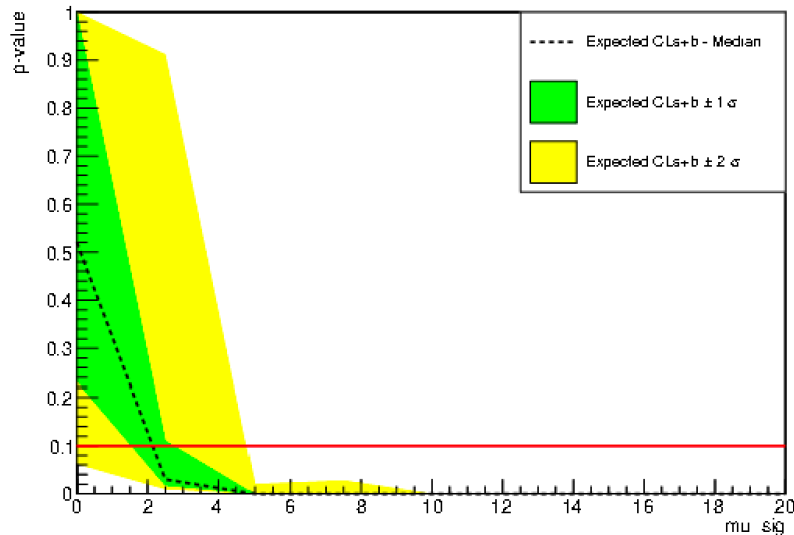


What's new?

- Run 10 GeV WIMP mass with changing g_1 values (0.05, 0.118, 0.13) and changing n_{sample} values (10^6 & 10^7)

10 GeV

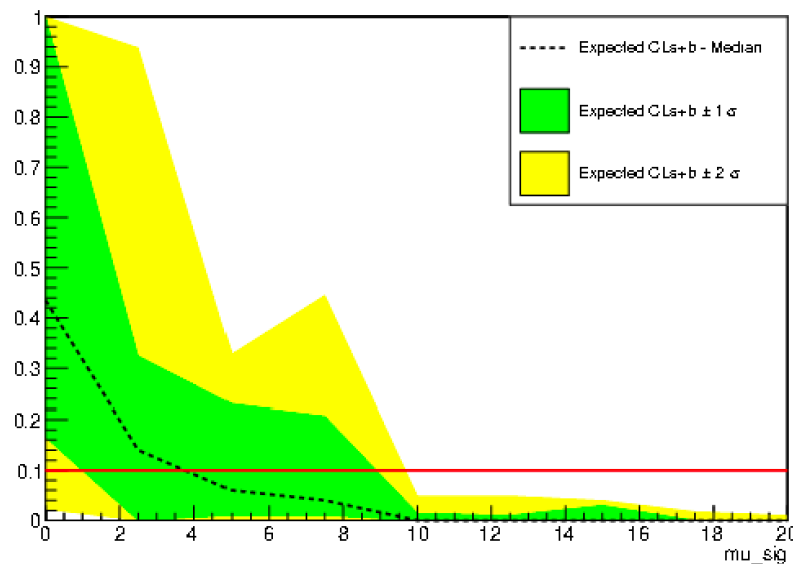
$n_{\text{sample}} = 10^6$



$g_1 = 0.05$

$\mu_{\text{sig}} = 2.1429$

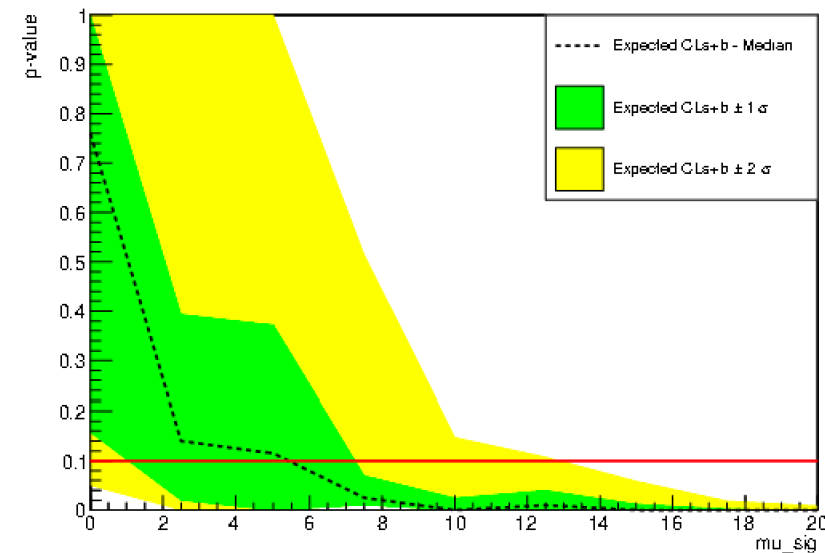
$\text{cross_section} = 6.8e-47$



$g_1 = 0.118$

$\mu_{\text{sig}} = 3.7500$

$\text{cross_section} = 1.84e-47$

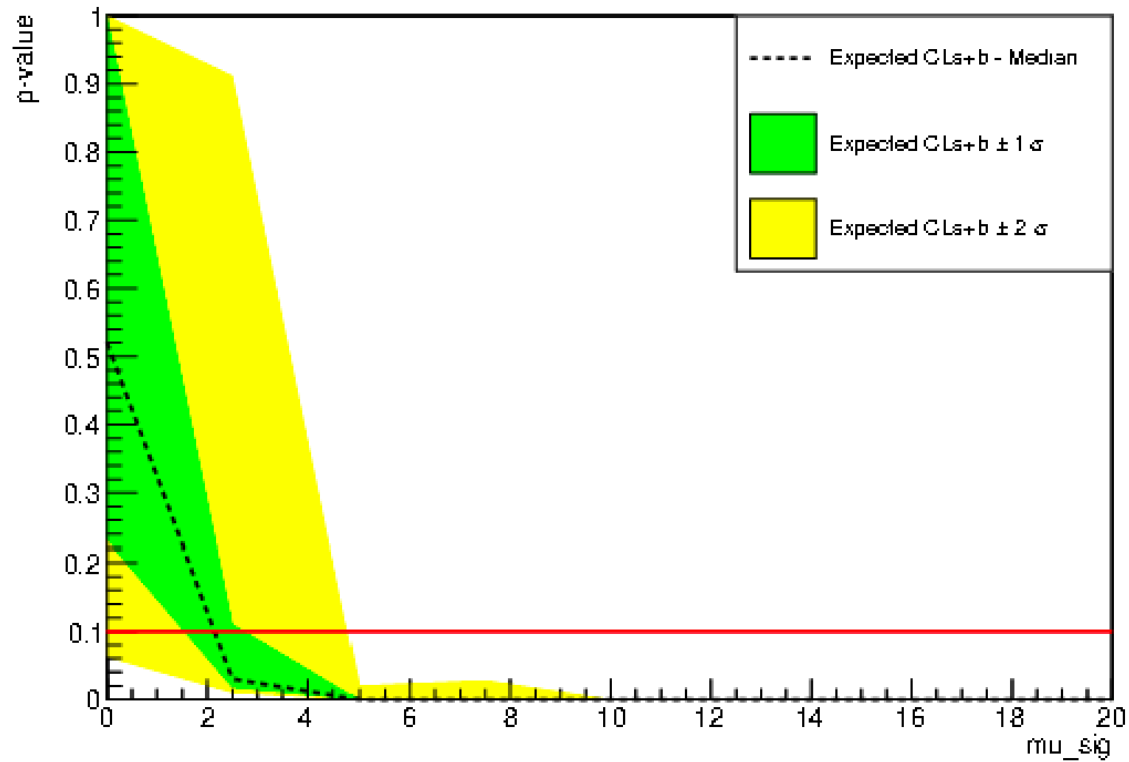


$g_1 = 0.13$

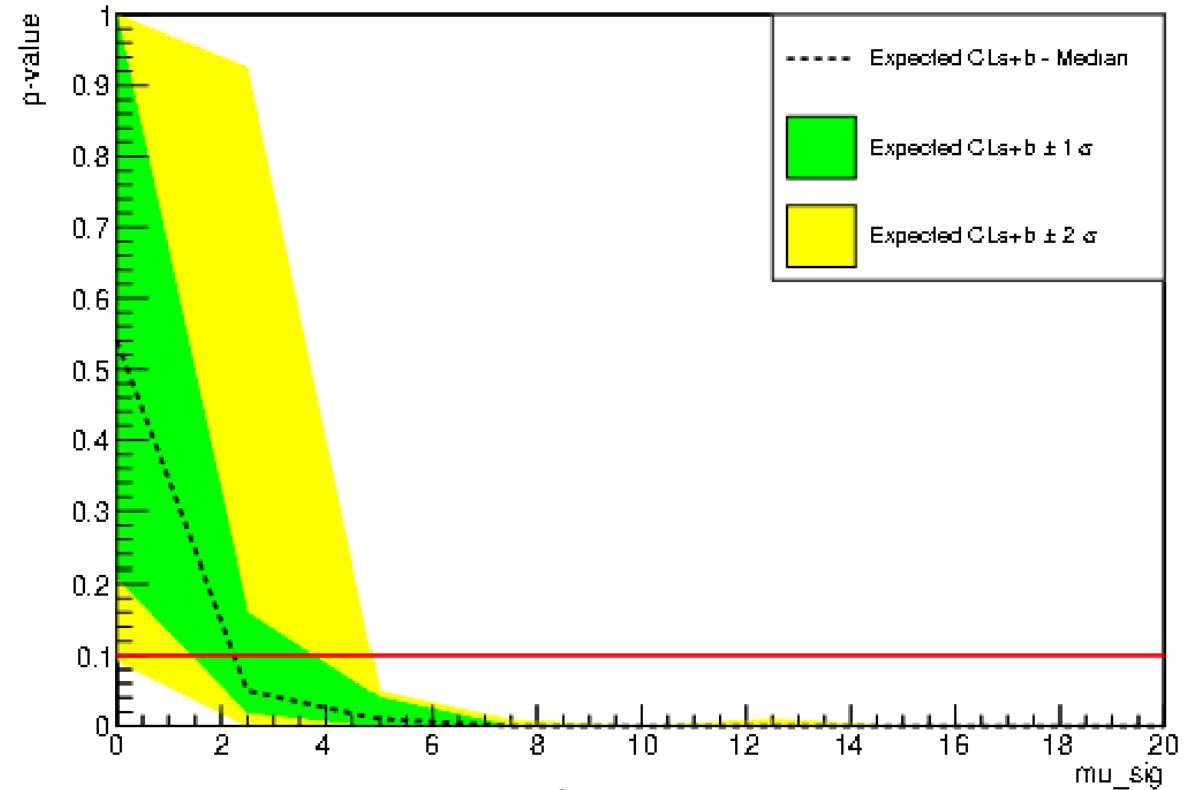
$\mu_{\text{sig}} = 5.4167$

$\text{cross_section} = 2.3e^{-47}$

10 GeV, $g_1 = 0.05$

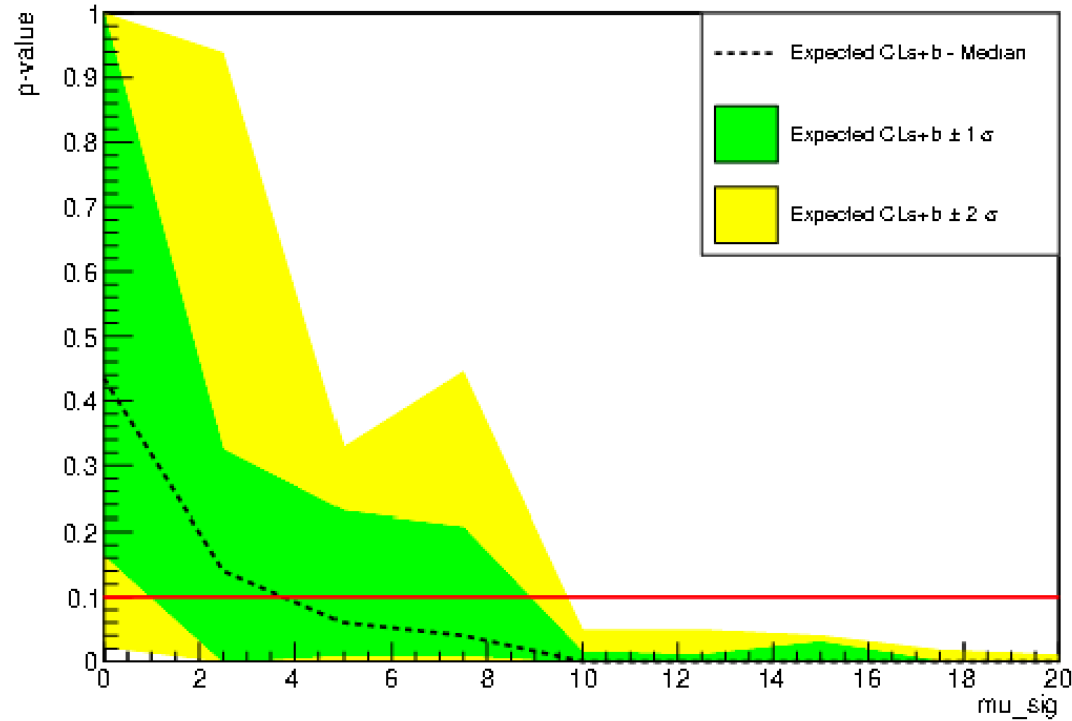


$N_{\text{sample}} = 10^6$
 $\mu_{\text{sig}} = 2.1429$
 $\text{cross_section} = 6.8\text{e-}47$

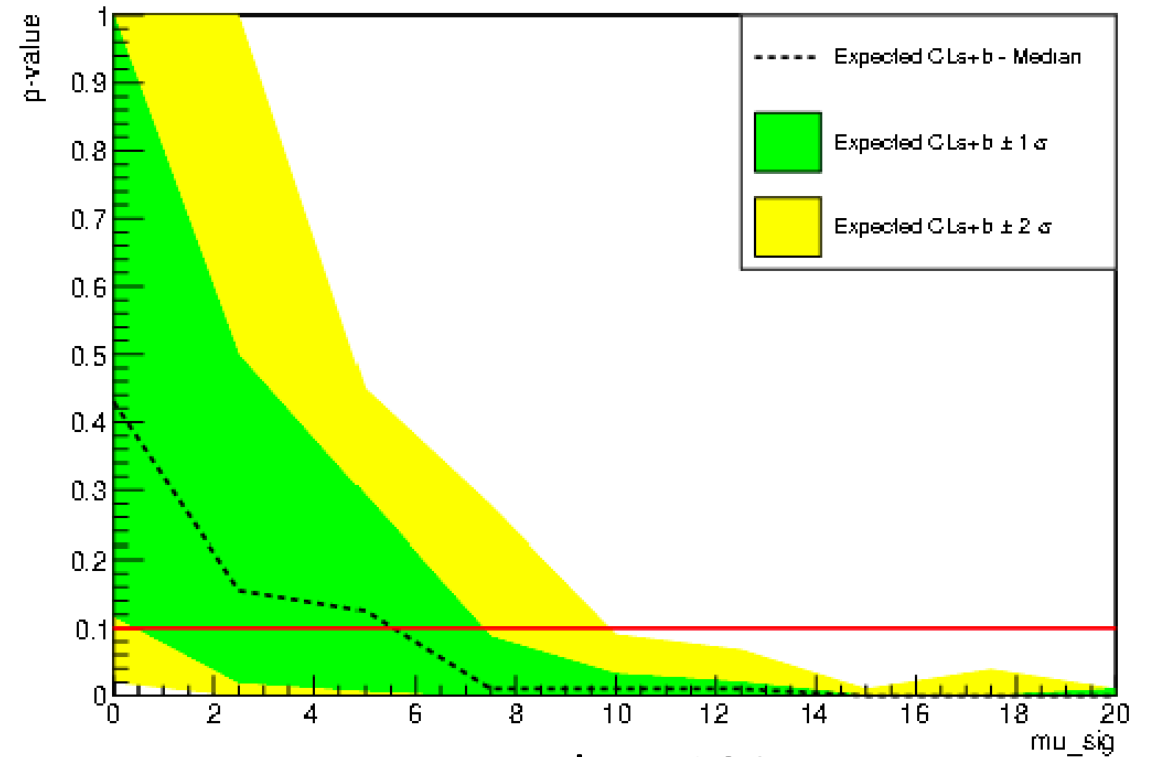


$N_{\text{sample}} = 10^7$
 $\mu_{\text{sig}} = 2.2449$
 $\text{cross_section} = 7.2\text{e-}47$

10 GeV, $g_1 = 0.118$

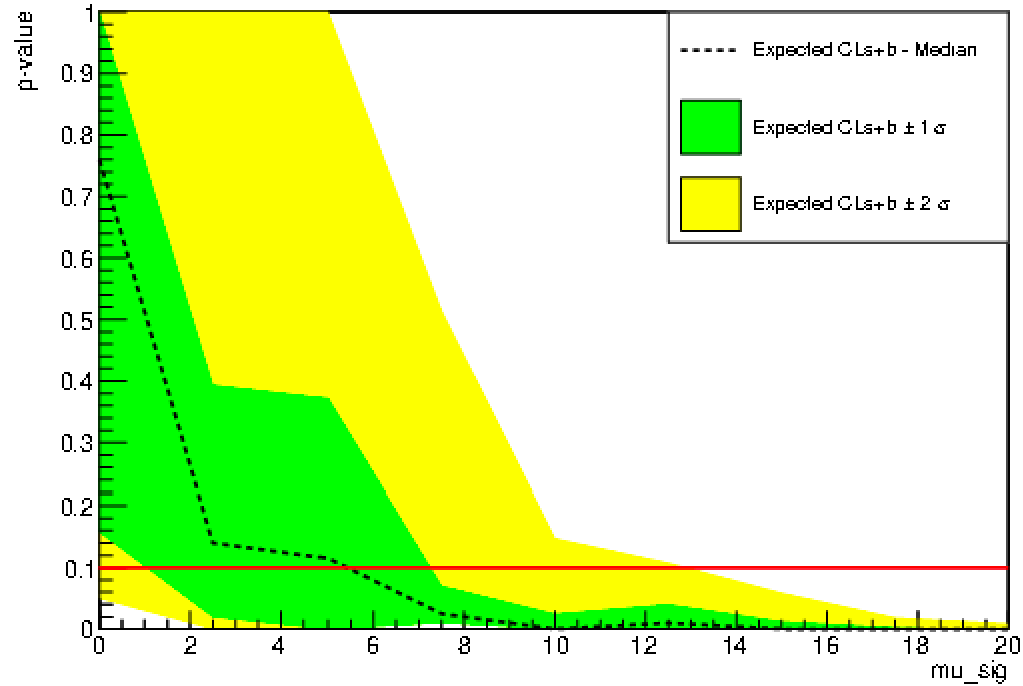


$N_{\text{sample}} = 10^6$
 $\mu_{\text{sig}} = 3.7500$
 $\text{cross_section} = 1.84\text{e-}47$

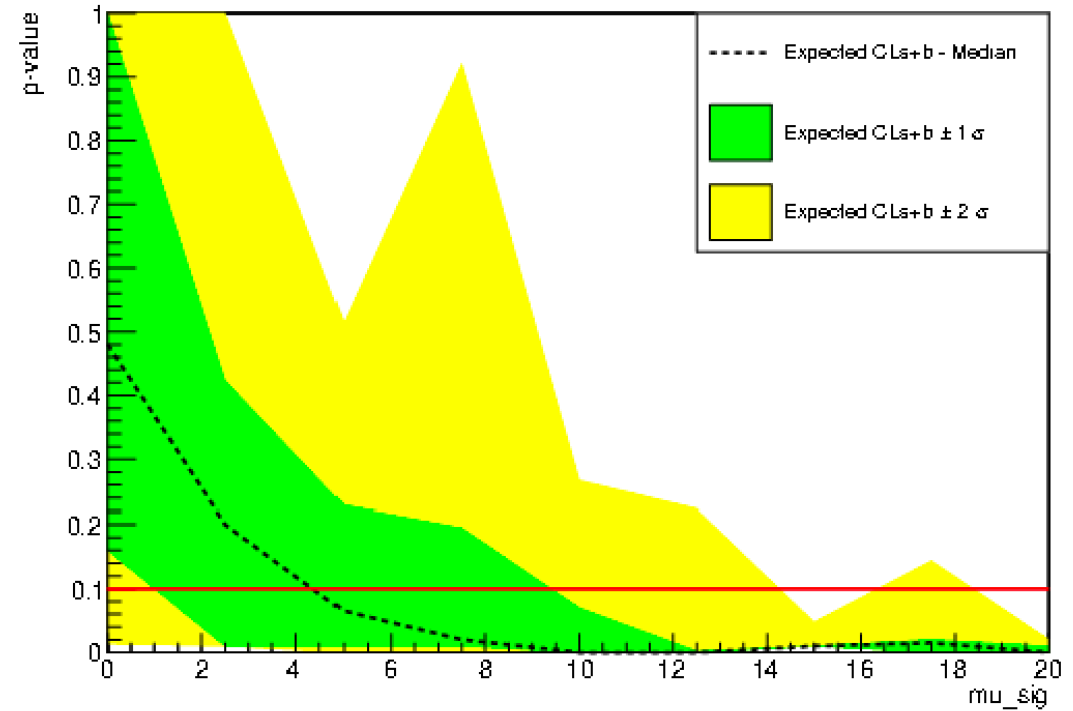


$N_{\text{sample}} = 10^7$
 $\mu_{\text{sig}} = 5.54$
 $\text{cross_section} = 2.74\text{e-}47$

10 GeV, $g_1 = 0.13$

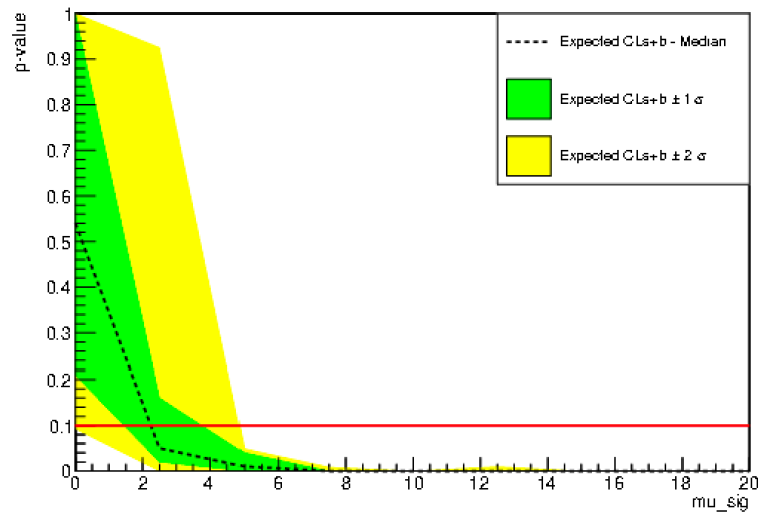


$N_{\text{sample}} = 10^6$
 $\mu_{\text{sig}} = 5.4167$
cross_section = 2.3×10^{-47}

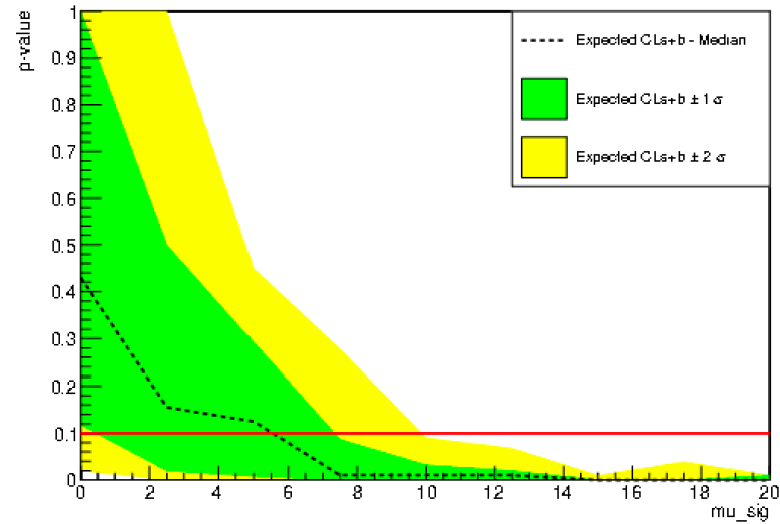


$N_{\text{sample}} = 10^7$
 $\mu_{\text{sig}} = 4.3544$
cross_section = 1.8×10^{-47}

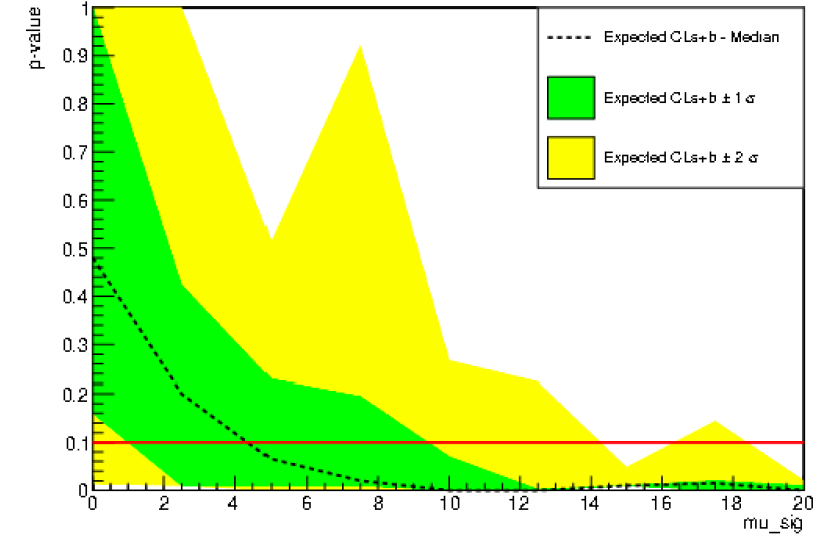
10 GeV
 $n_{\text{sample}} = 10^7$



$g_1 = 0.05$
 $\mu_{\text{sig}} = 2.2449$
 $\text{cross_section} = 7.2\text{e-}47$



$g_1 = 0.118$
 $\mu_{\text{sig}} = 5.54$
 $\text{cross_section} = 2.74\text{e-}47$



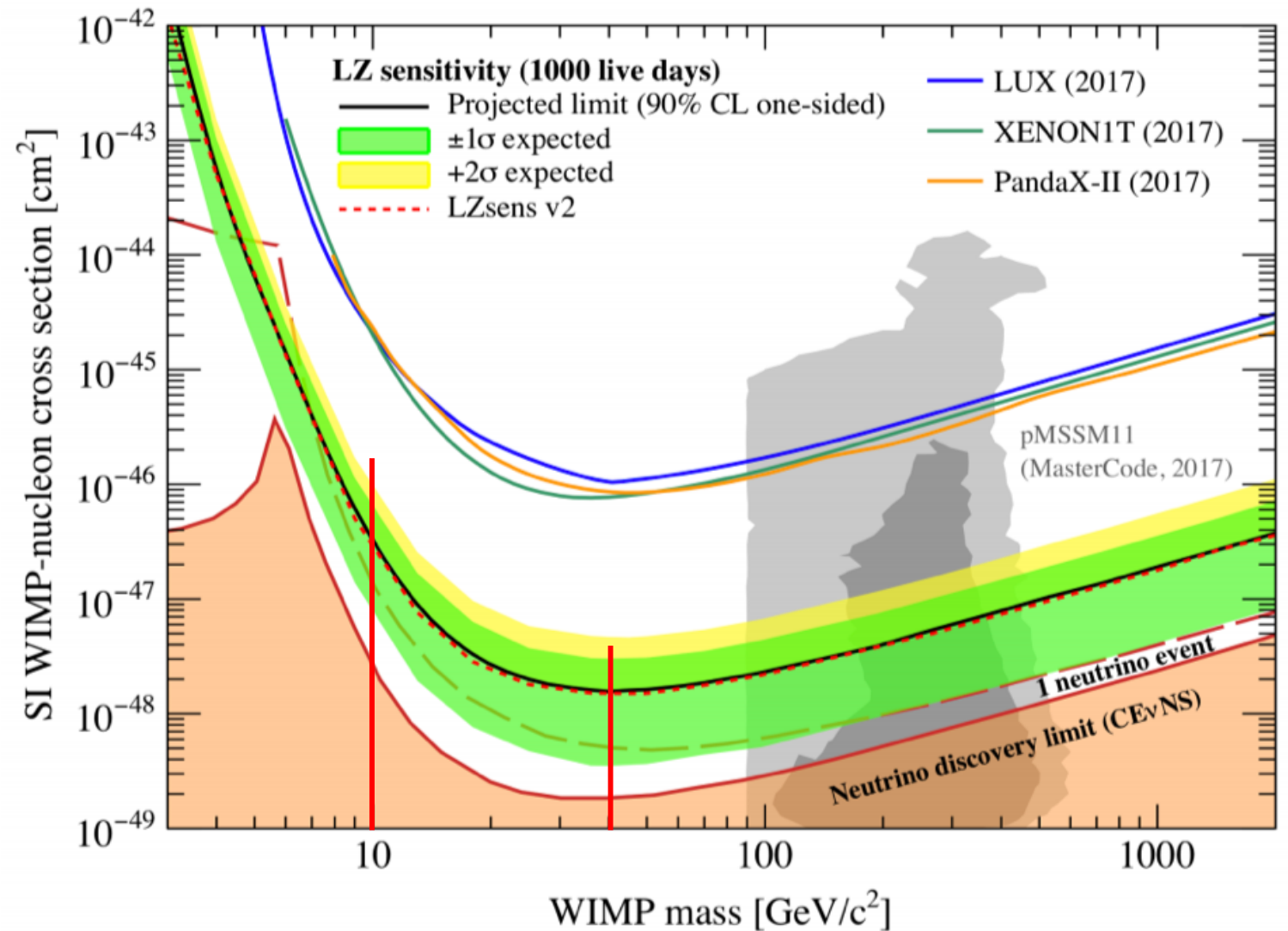
$g_1 = 0.13$
 $\mu_{\text{sig}} = 4.3544$
 $\text{cross_section} = 1.8\text{e-}47$

Comparison (10 GeV vs 40 GeV)

- Change g_1 from 0.05 $\rightarrow$ 0.13
 - Cross section for 10 GeV changes 75%
 - Cross section for 40 GeV changes 27%
- Why is there a more dramatic change in lower WIMP mass when changing g_1 ?

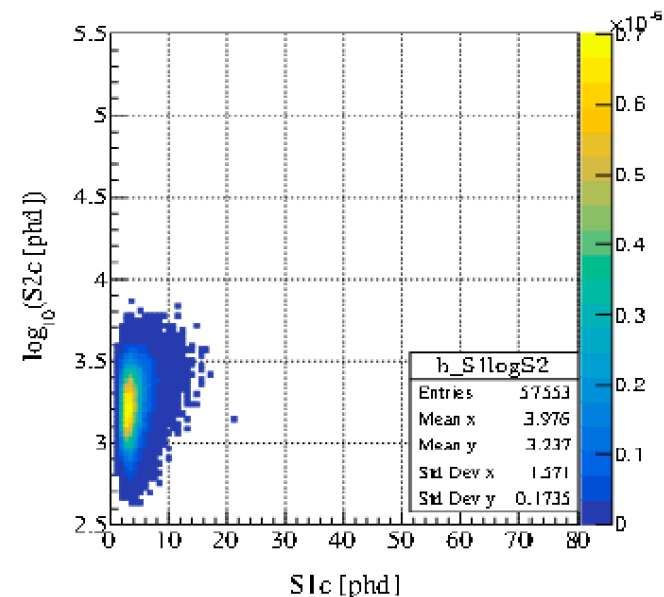
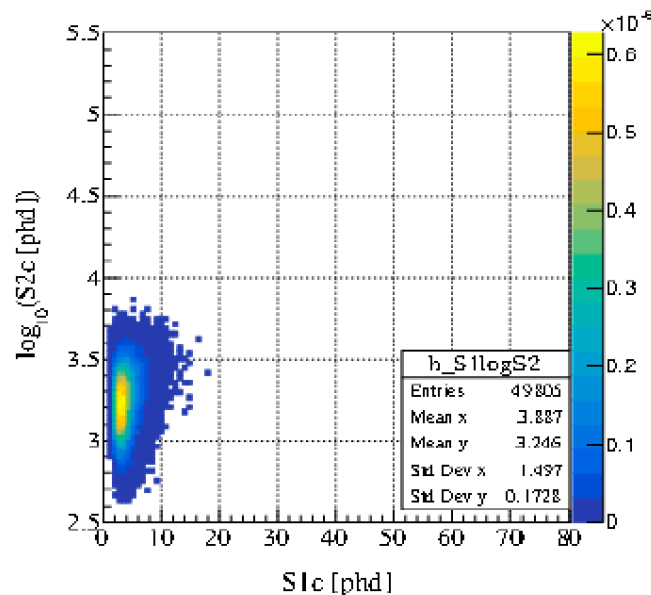
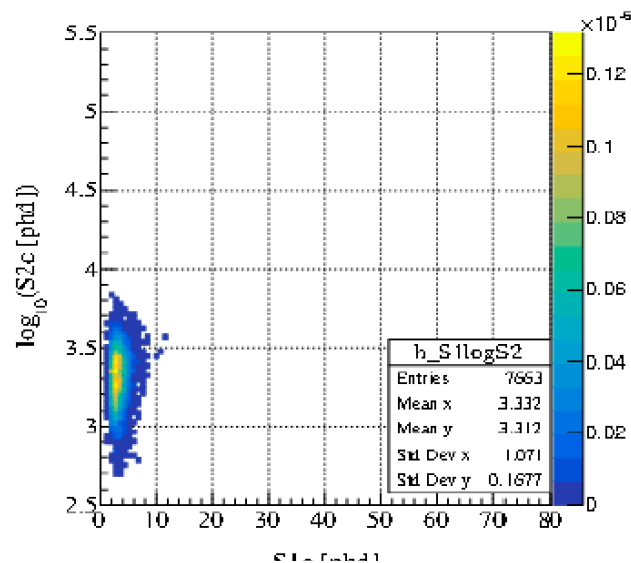
Limit plot

$10 \text{ GeV} \sim 3e^{-47}$

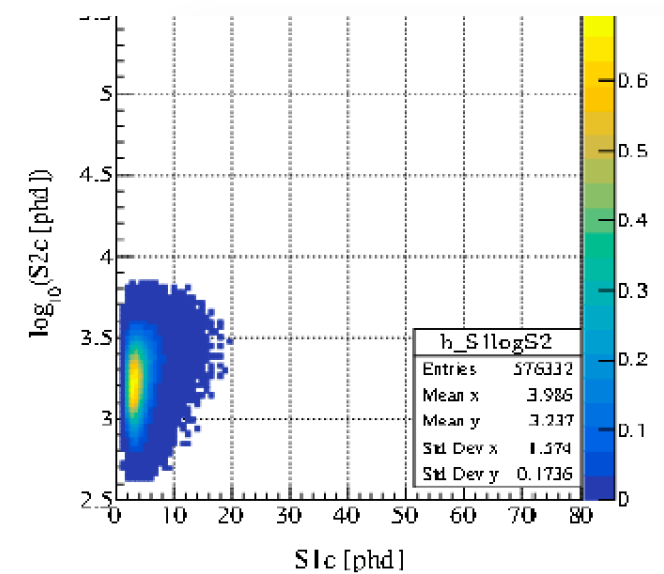
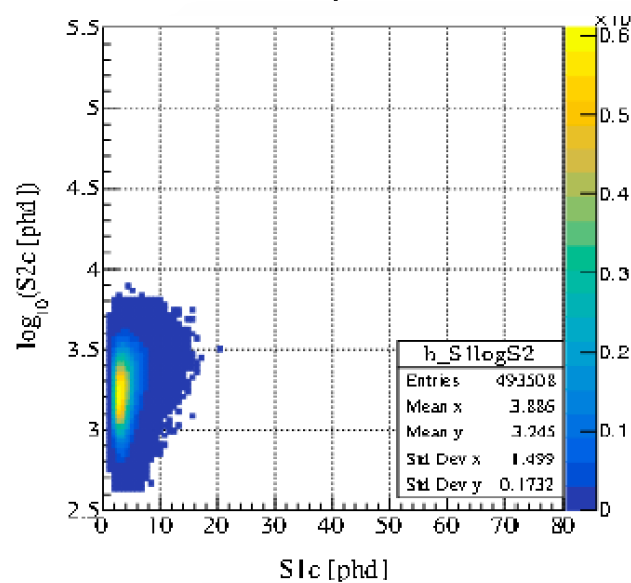
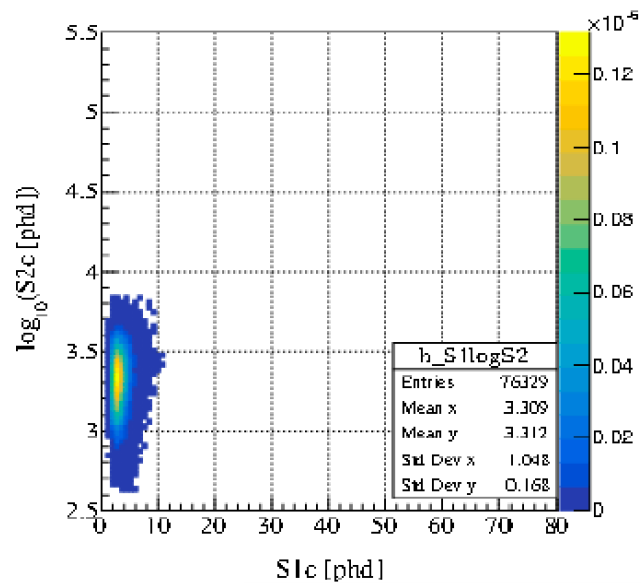


WIMP 10GeV

$N_{\text{sample}} = 10^6$



$N_{\text{sample}} = 10^7$



$g1 = 0.05$

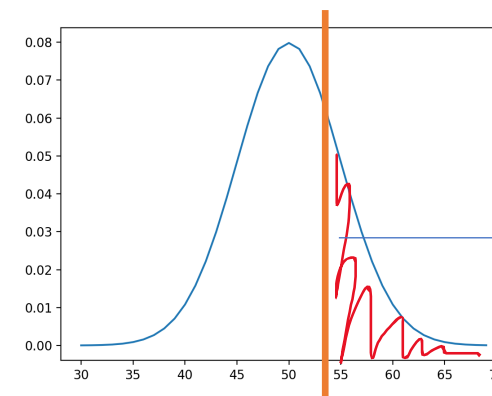
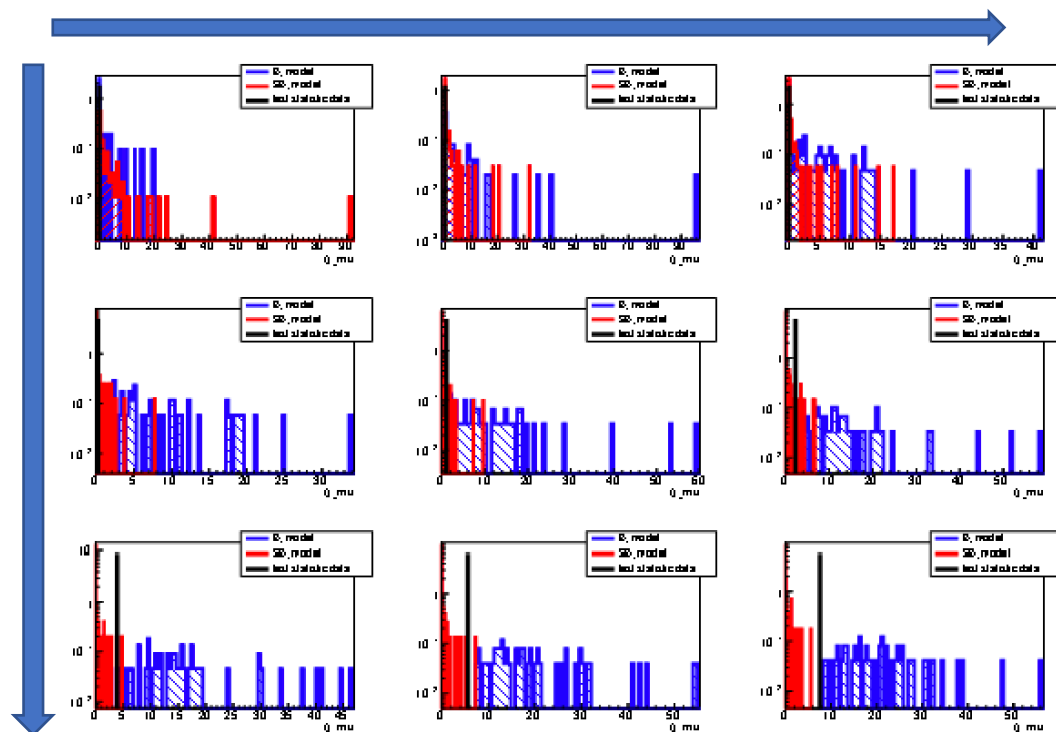
$g1 = 0.118$

$g1 = 0.13$

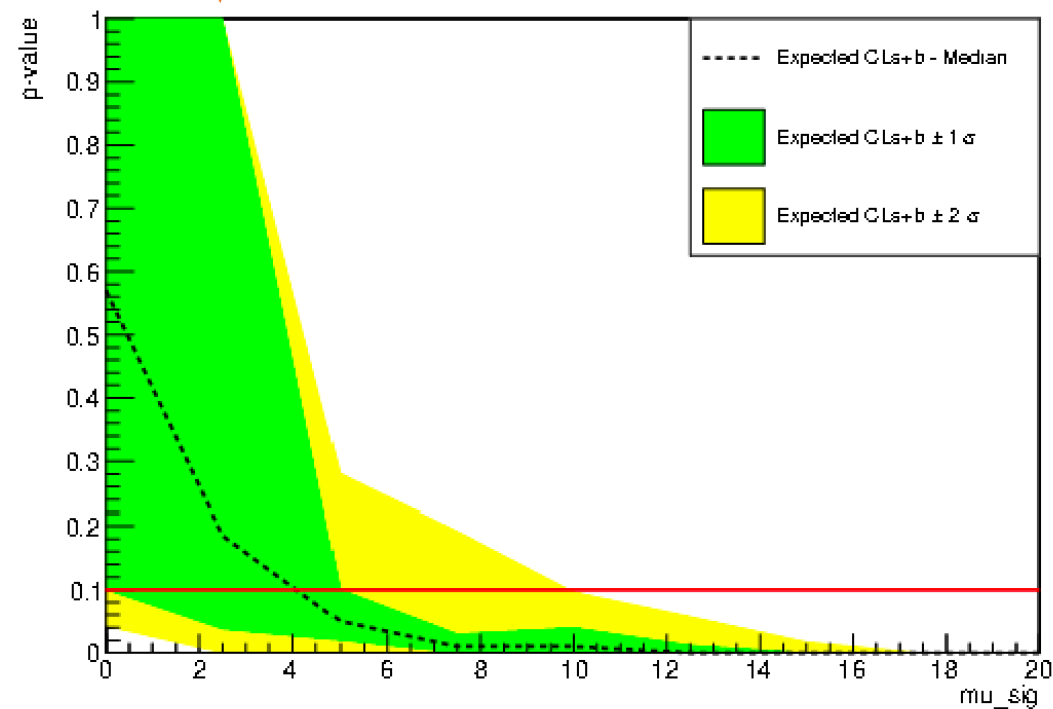
What's new?

- Analyze the graph
- Change g_1 value and compare the limits

POI = 0 $\sim$ 20



P-value



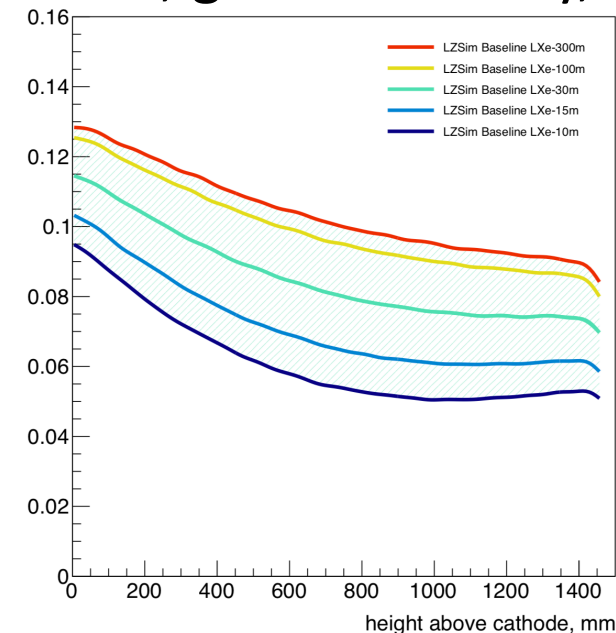
g1

- g1
 - Photon detection efficiency: the average fraction of S1 light that is detected by PMTs.
 - Detected photoelectrons/emitted scintillation photon
 - Property of the combination of PMTs and detector, grid reflectivity, Xe absorption of light...
- LZsim baseline in Liquid Xe

$g1_{\text{min}} \sim 0.05$

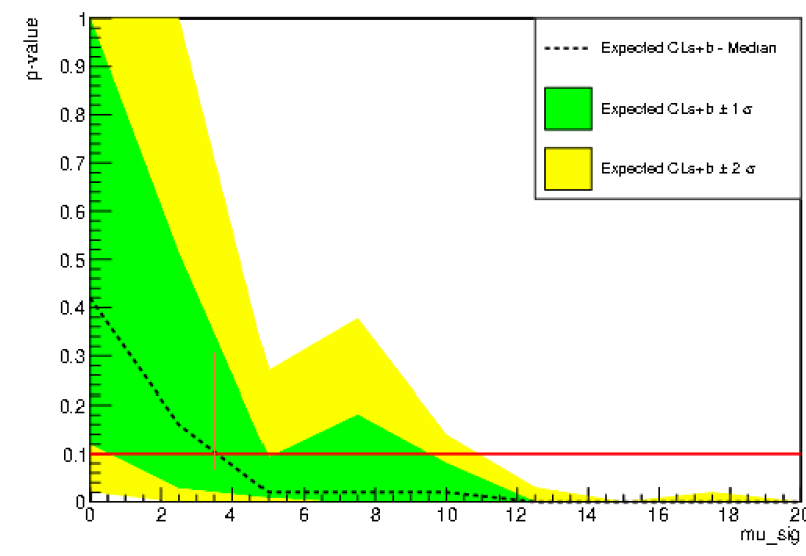
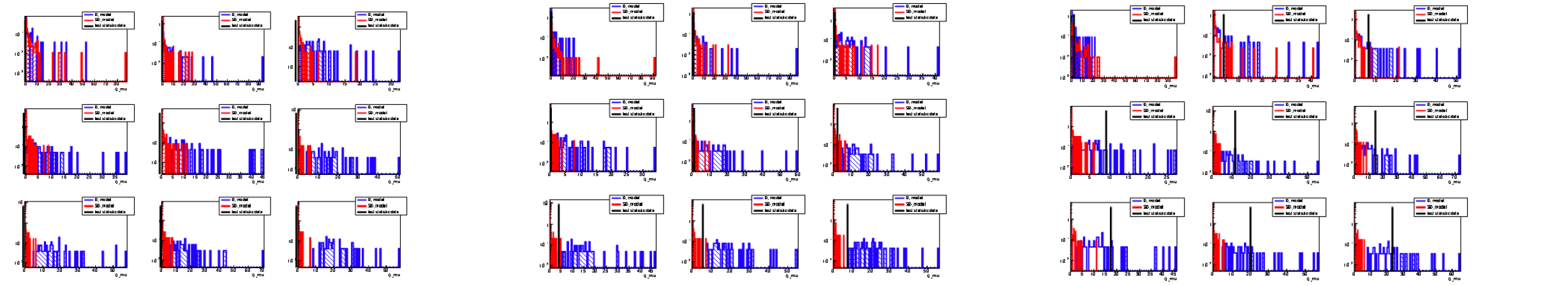
$g1_{\text{default}} \sim 0.118$

$g1_{\text{max}} \sim 0.13$

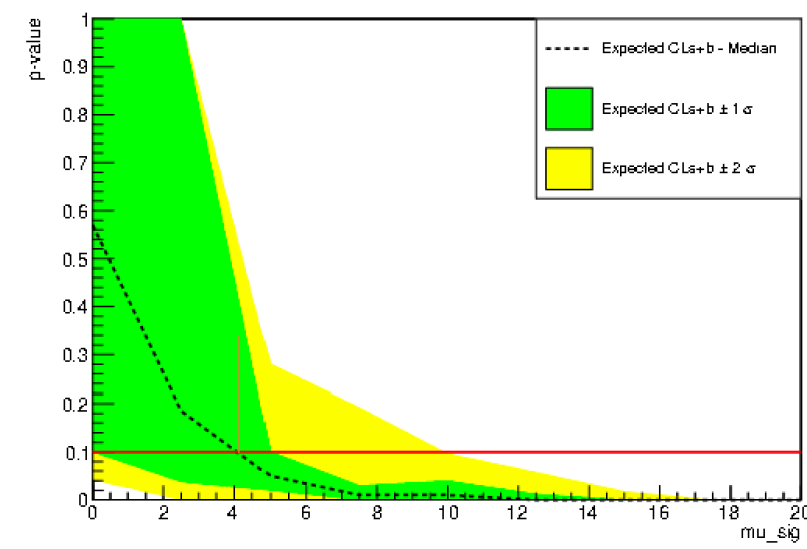


Code

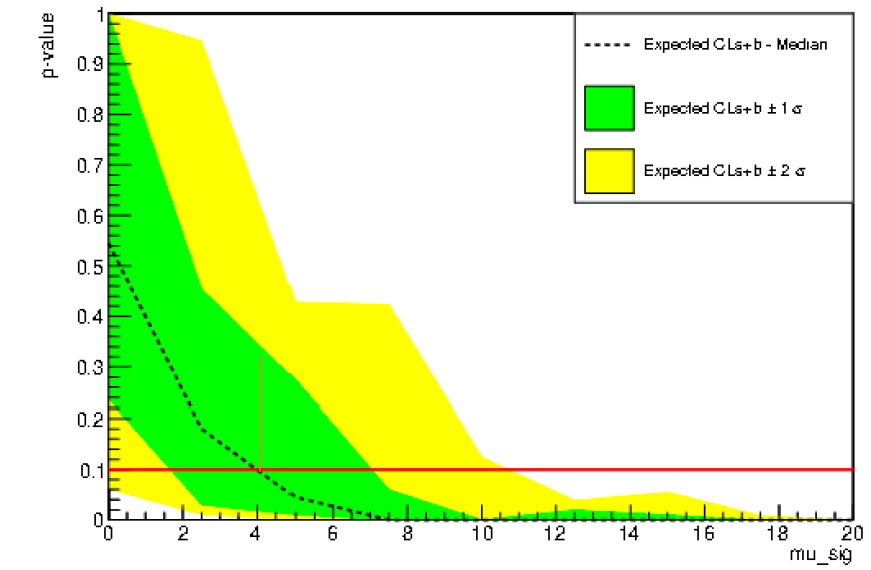
- LZNESTUtils
 - Change Detector parameters
 - Source setup.sh
 - Source .ci/build.sh
 - Get new NESTInterference
- PdfMaker
 - Recompile signal and background codes with the new NESTInterference



$g1 = 0.05$
 $\mu_{sig} = 3.5027$
 $cross_section = 1.44e^{-48}$



$g1 = 0.118$
 $\mu_{sig} = 4.0625$
 $cross_section = 1.12e^{-48}$



$g1 = 0.13$
 $\mu_{sig} = 3.9835$
 $cross_section = 1.05e^{-48}$

To do

- Try lower WIMP mass (10GeV) and see how the sensitivity changes.

What's new:

- Figure out what's causing the error.
 - Constant pp value → in the code(for HypoTestInversion), remove any constant parameters → number of parameters are not consistent.
- Update the Low_NR_Groups, re-generate the workspace.
 - No more problem running the LZStats.
- Compare the two versions of code and workspaces

Code (pp)

Before	After
<pre>a0_tmp = bckg_rate * exposure = bckg_rate*livedays*fiducial_mass = 542.16 a_pp = a0_tmp mu_pp = Range (507, 574) sigma_pp = systematic_uncertainties * a0_tmp = 2% * 542.16 ~10</pre>	<pre>a0_tmp = bckg_rate * exposure = bckg_rate*livedays*fiducial_mass = 542.16 a_pp = a0_tmp mu_pp = a0_tmp sigma_pp = systematic_uncertainties * a0_tmp = 2% * 542.16 ~10</pre>
<pre>w->factory("Gaussian::constraint_pp(mu_pp,a_pp,sigma_pp)");</pre>	

Workspace (comparison)

- Information printed out from workspace:
 - Constraints changed:

Background	Before	After
B8	0.011109	1
DSN	0.011109	1
DetER	0.011109	1
DetNR	0.011109	1
Kr85	0.011109	1
Rn220	0.011109	1
Rn222	0.572416	1
atm	0.011109	1
hep	0.011109	1
pp	0.000521151	1
vvBB	0.825786	1

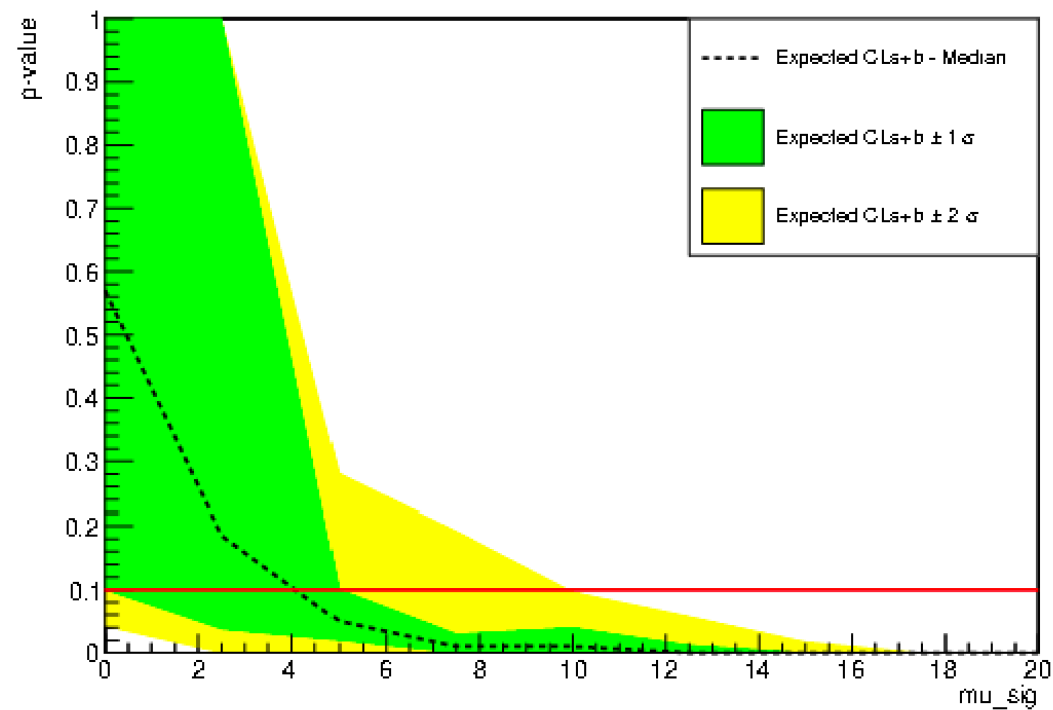
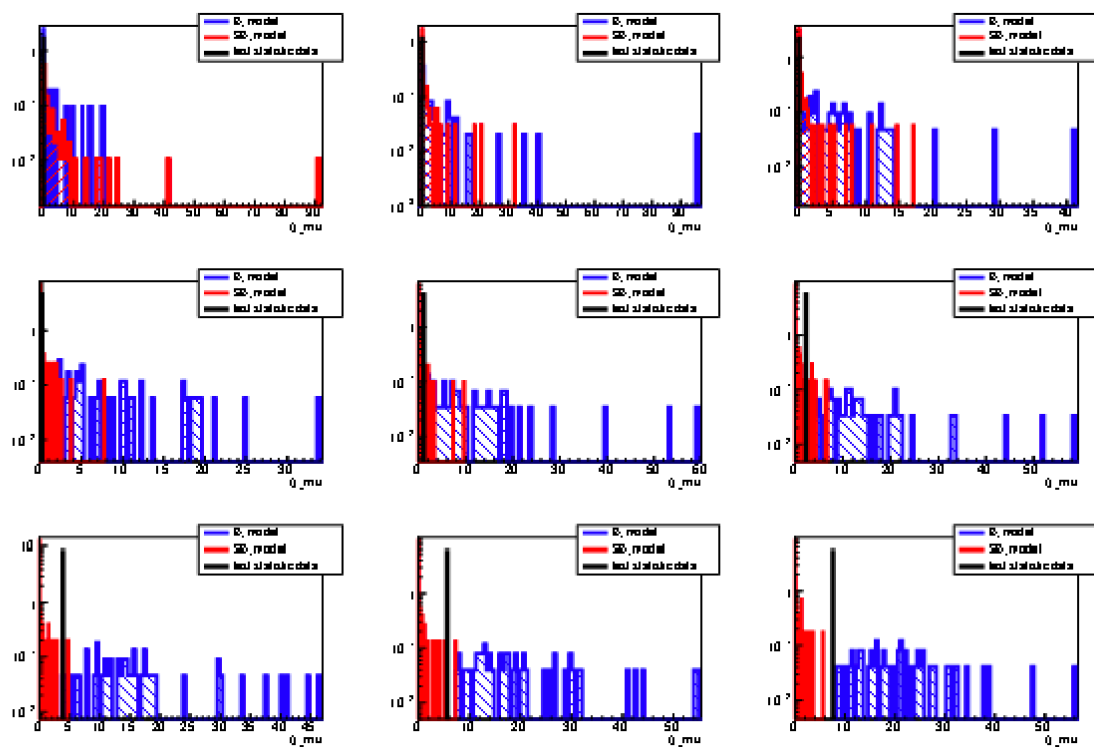
Constraint: Gaussian
distribution of
(x=mu_pp
mean = a_pp,
Sigma = sigma_pp)

$$\sim \frac{(x-\mu)^2}{\sigma^2}$$

- Total Event Model:
 - Event Model * Total Constraints

	Before	After
Event Model	8.09546e-13	4.80763e-13
Total Event Model	4.62575e-32	4.80763e-13

WIMP (40GeV)



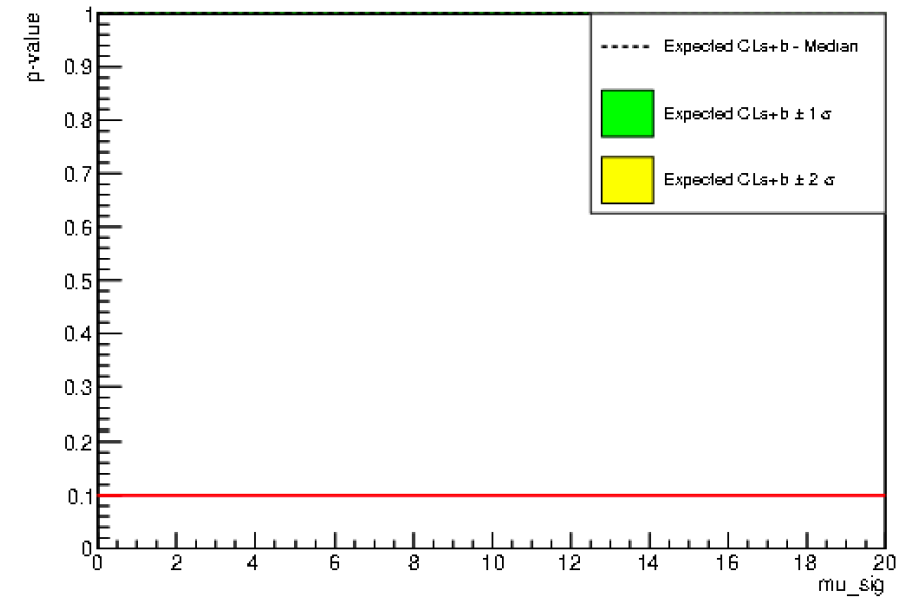
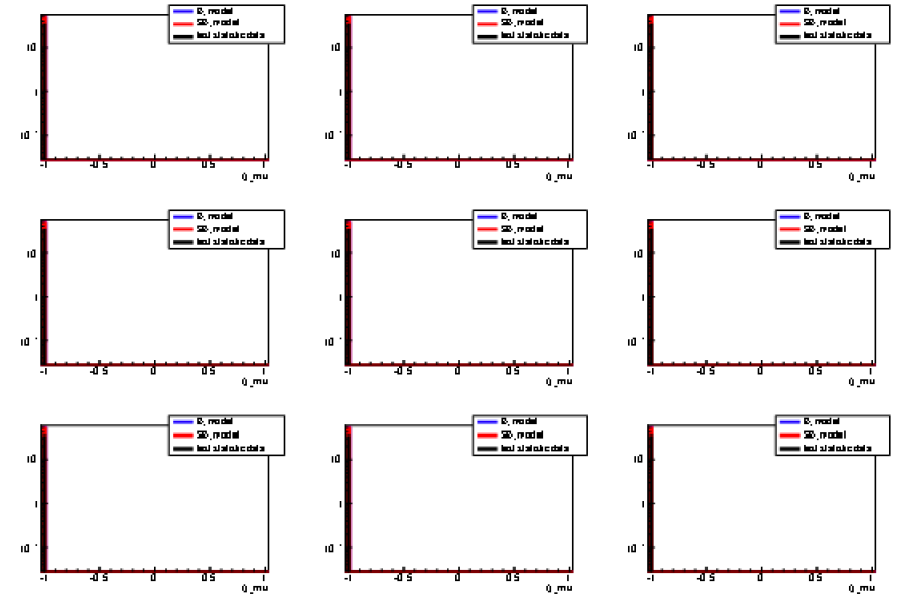
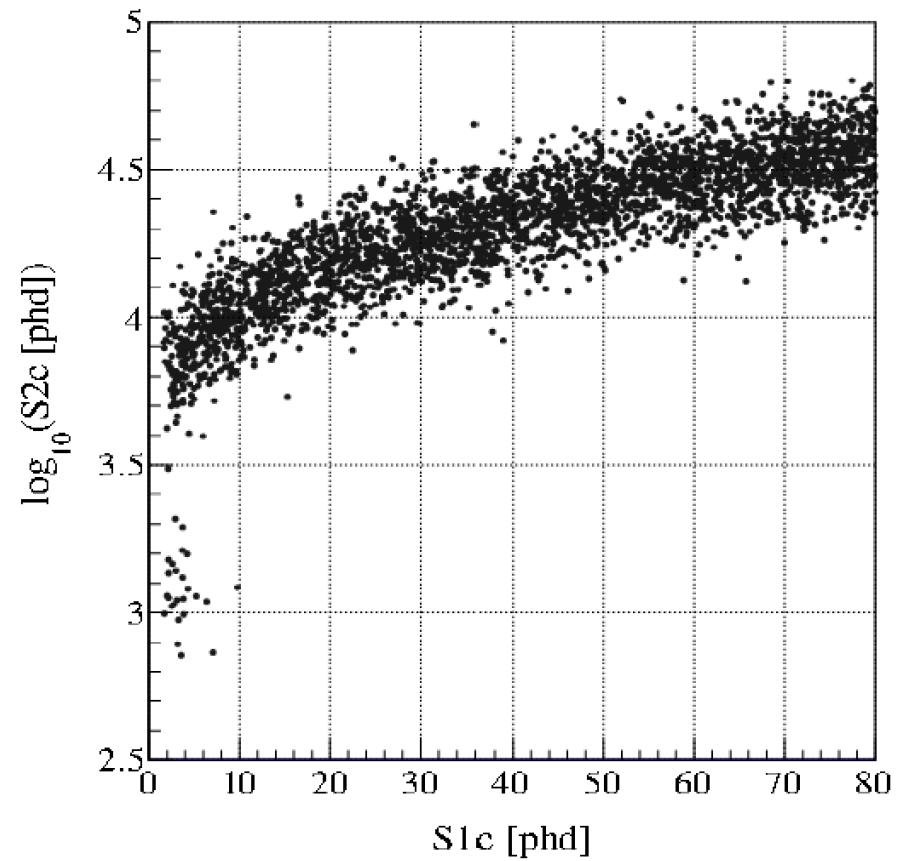
What's done

- LZNESTUtils: Change to a new branch
- PdfMakers: compile the background code successfully
- LowR_NR: passed 12 root files (1 signal +11 backgrounds) to make a workspace (1 root file)
- LZStats: run the code locally with the workspace generated from the LowR_NR

Running LZStats

- Frequentist Analysis (limit setting)
 - Confidence level: 90%
 - Fiducial mass (kg): 5600
 - Livetime(days) : 1000
 - Paritcle mass (GeV): 40
 - MC toys per point : 100
 - Number of Points: 9
 - Min POI: 0
 - Max POI: 20

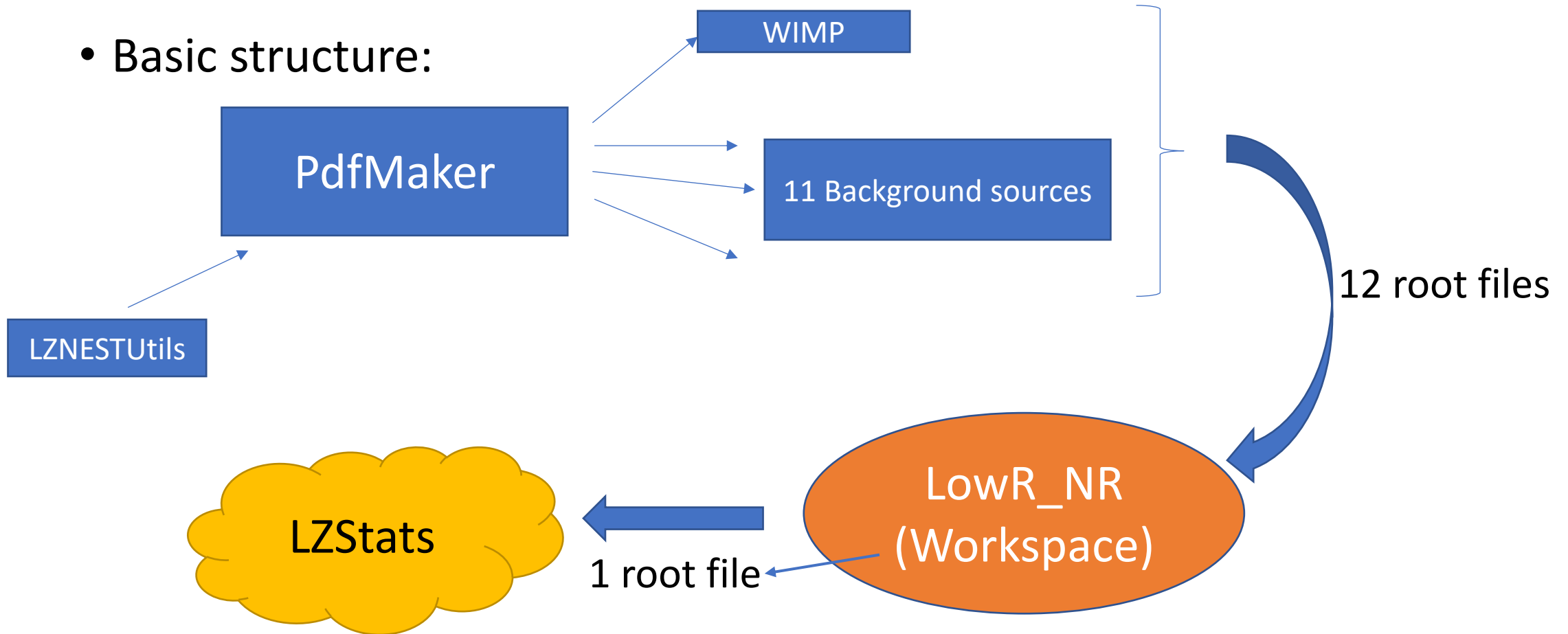
Problem



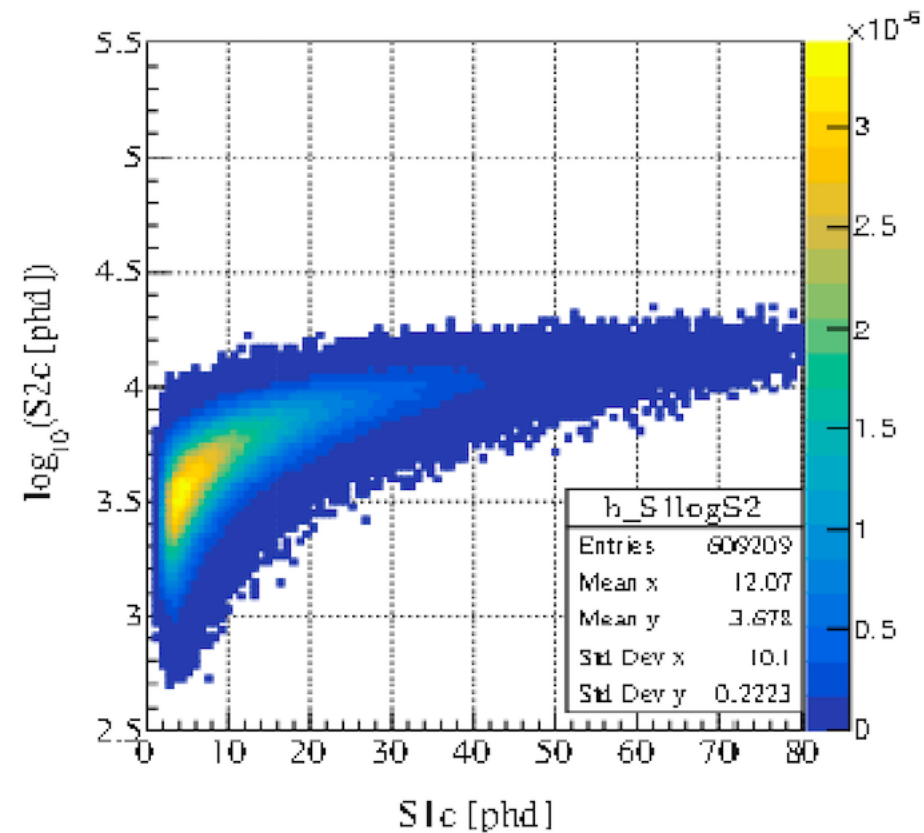
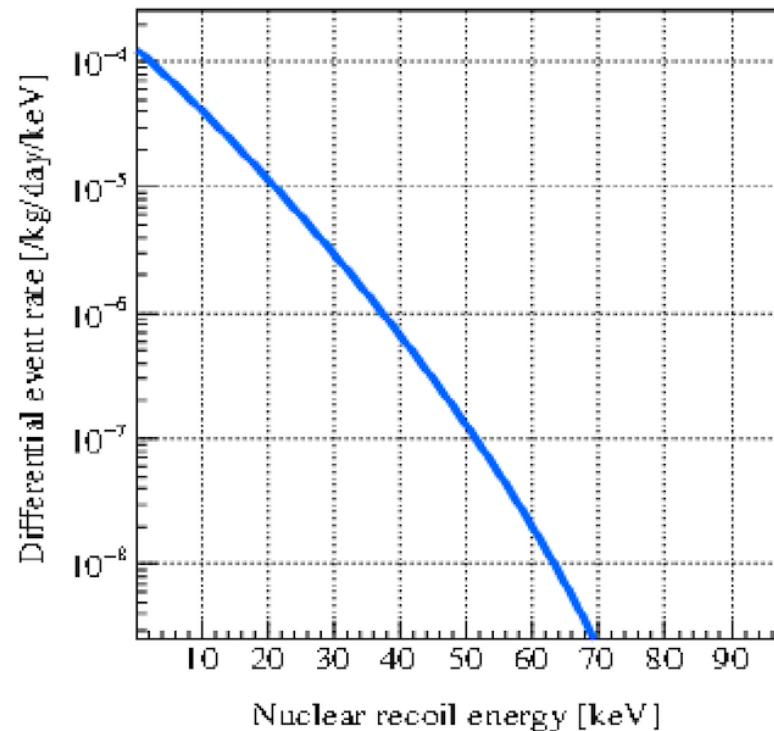
```
Error in <ROOT::Math::FitResult>: FitConfiguration and Minimizer result are not  
consistent  
Number of free parameters from FitConfig = 12  
Number of free parameters from Minimizer = 11
```

Changed LZStats

- Basic structure:



Output of signal (40GeV WIMP) from PdfMaker



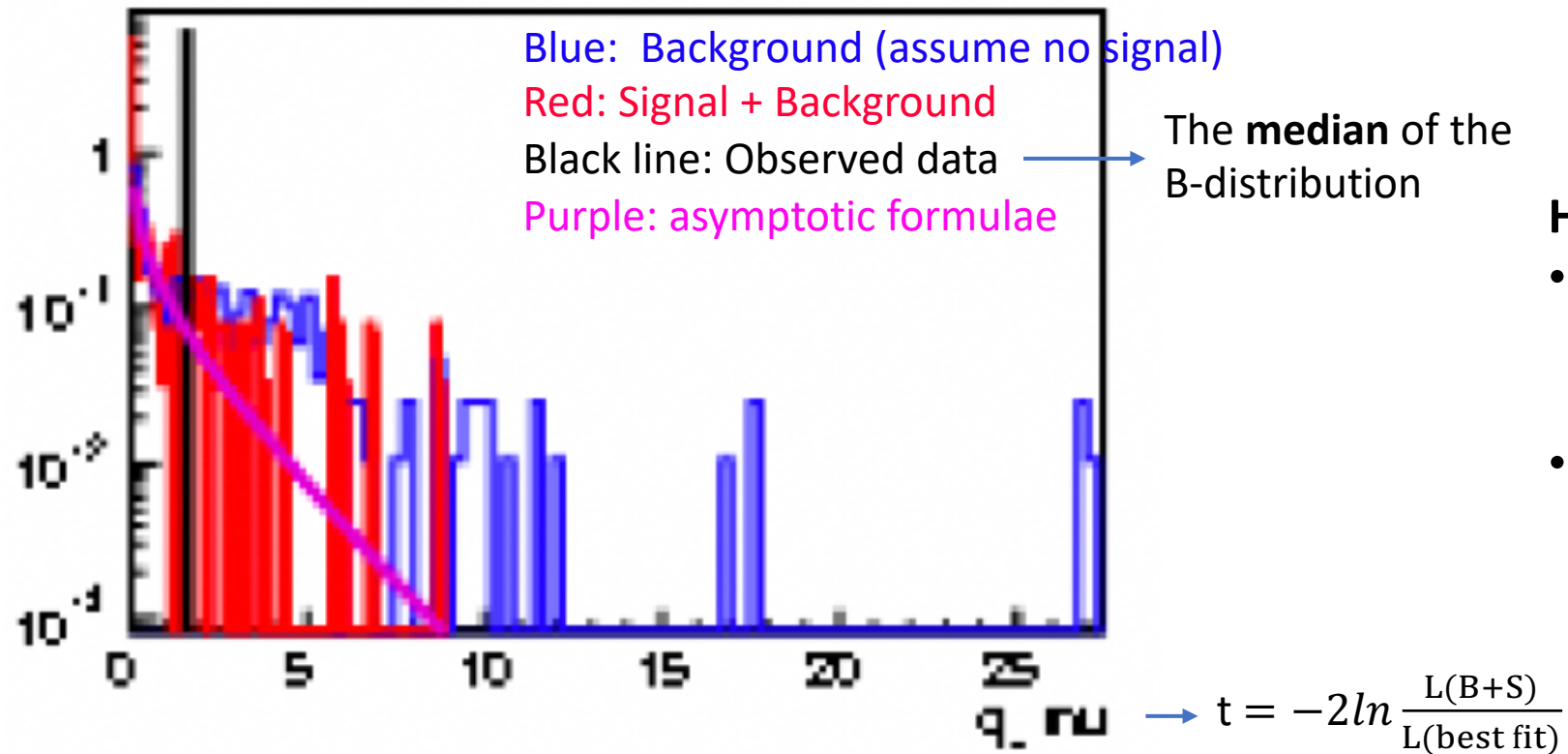
Workspace → RooWorkspace

- Settings
 - WIMP_mass = 40
 - Interaction type = Spin Independent
 - Livedays = 1000
 - Fiducial (kg) = 5600

Problem/To do

- Have trouble compiling the background code, one function in the background code is not defined in the LZNESTUtils → possible solution: change branches in the LZNESTUtils.

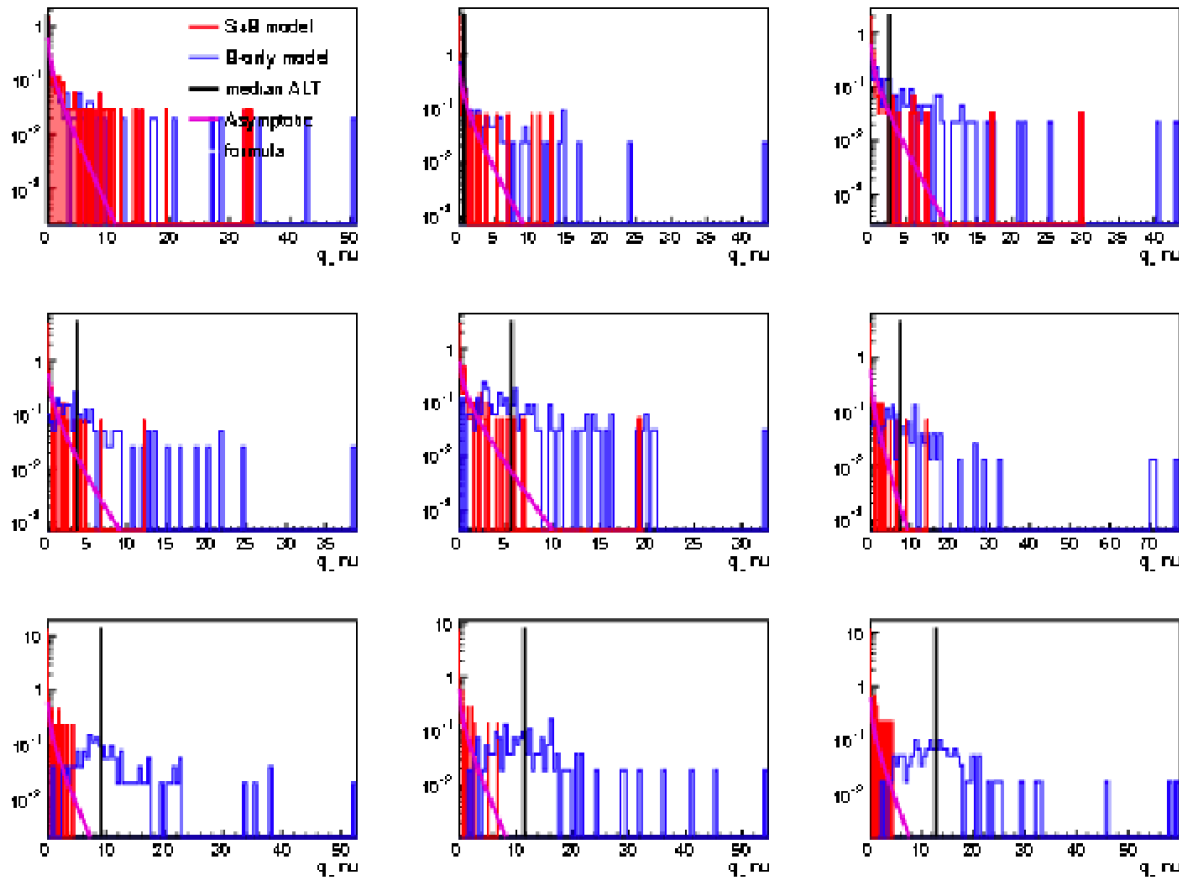
Analysis of a single graph



Hypothesis:

- **H0**: The data has both background and signal events fixed signal
- **H1**: The data only has background events floating

Increasing POI (0~15)



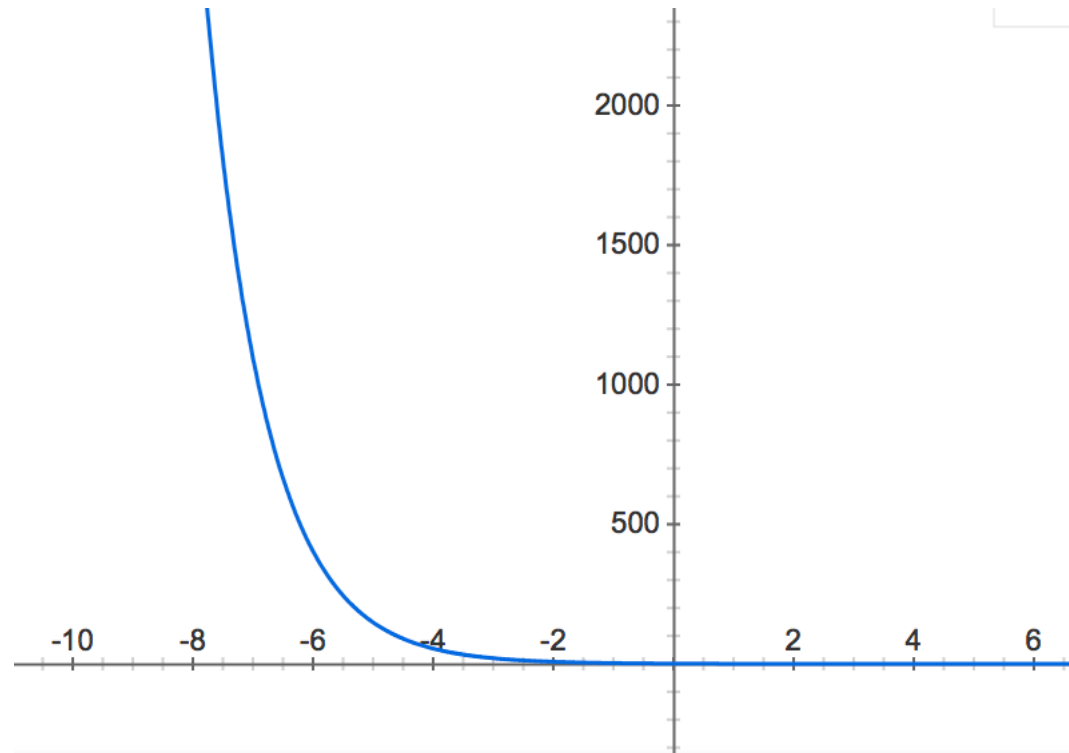
Method

- Background is fixed for different WIMP masses
- For one WIMP mass, increasing values of the parameter of interest (POI) $\rightarrow$ the cross section of the WIMP's interaction with Xenon
- Find the upper limit on the POI

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^N \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^M \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$



For high POI $\rightarrow$ high μs_j



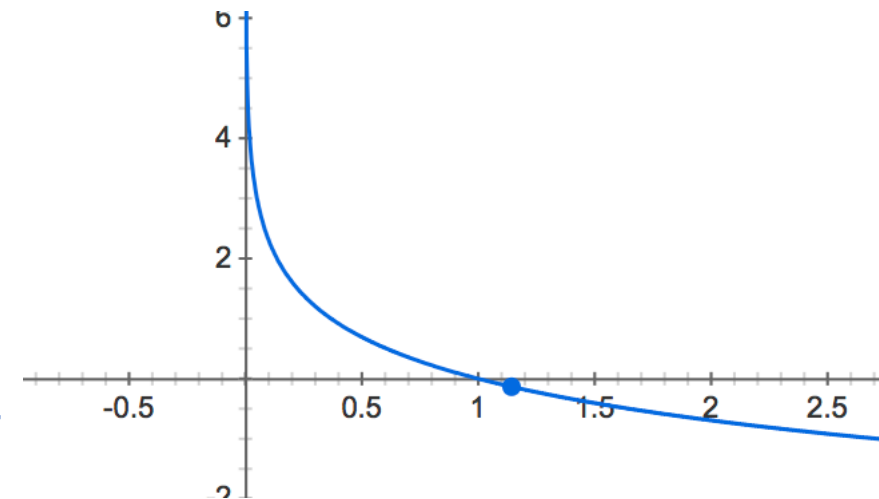
Equations & Graphs

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})}$$

μ : fixed
 $\hat{\mu}$: floating

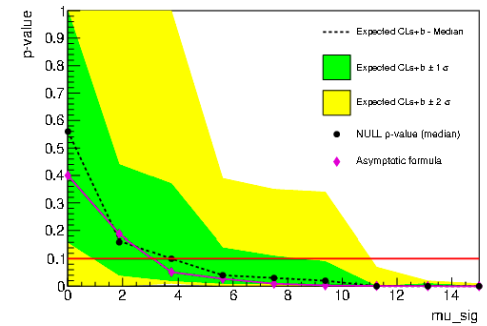
$$t_{\mu} = -2 \ln \lambda(\mu)$$

$\ln x$ vs x



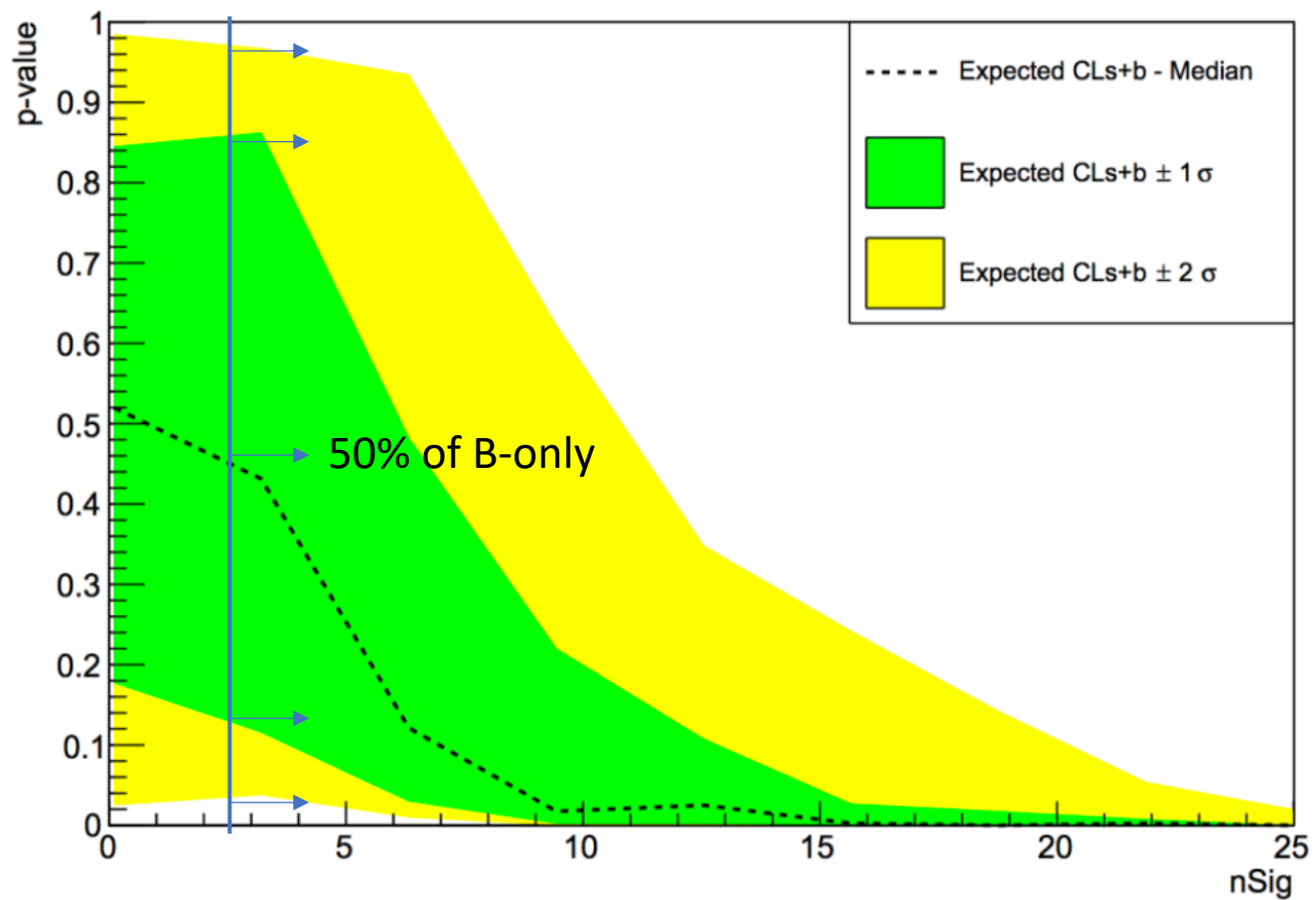
To do

- How to get the upper limit, basically how to get this graph from the nine graphs.
- Analyze the graph

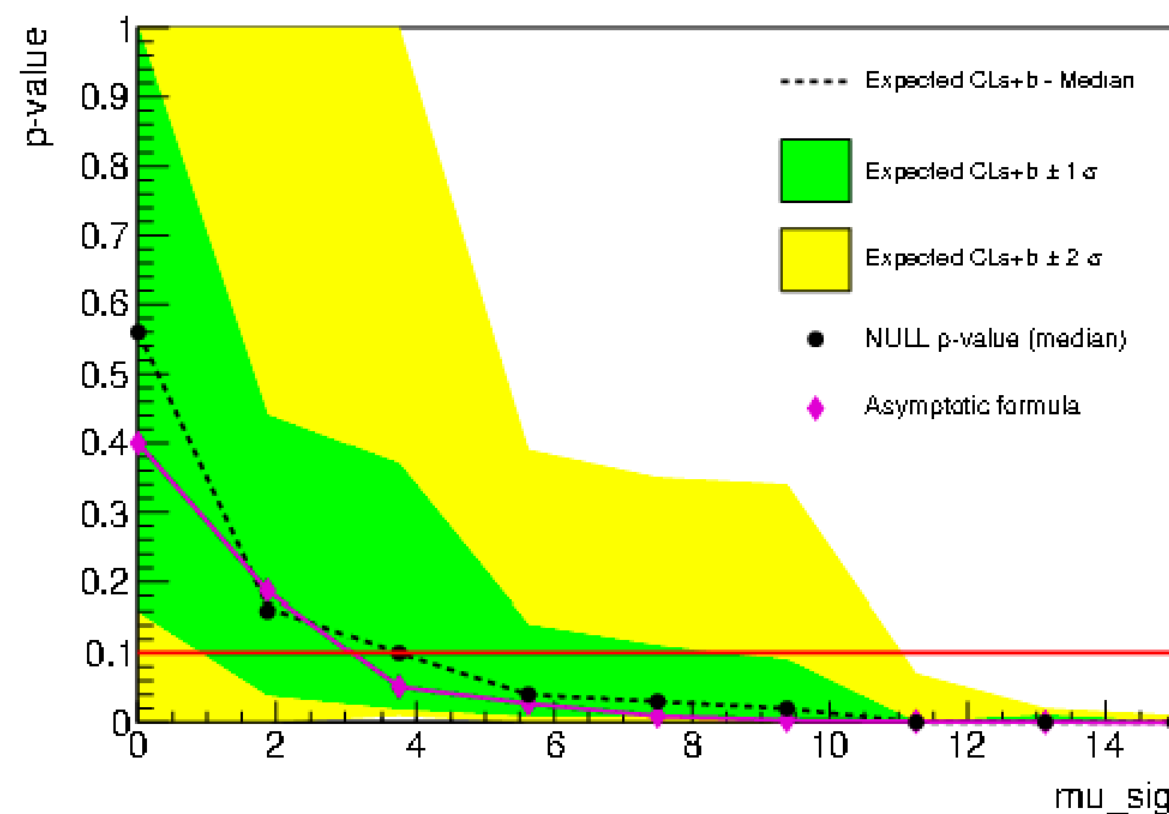
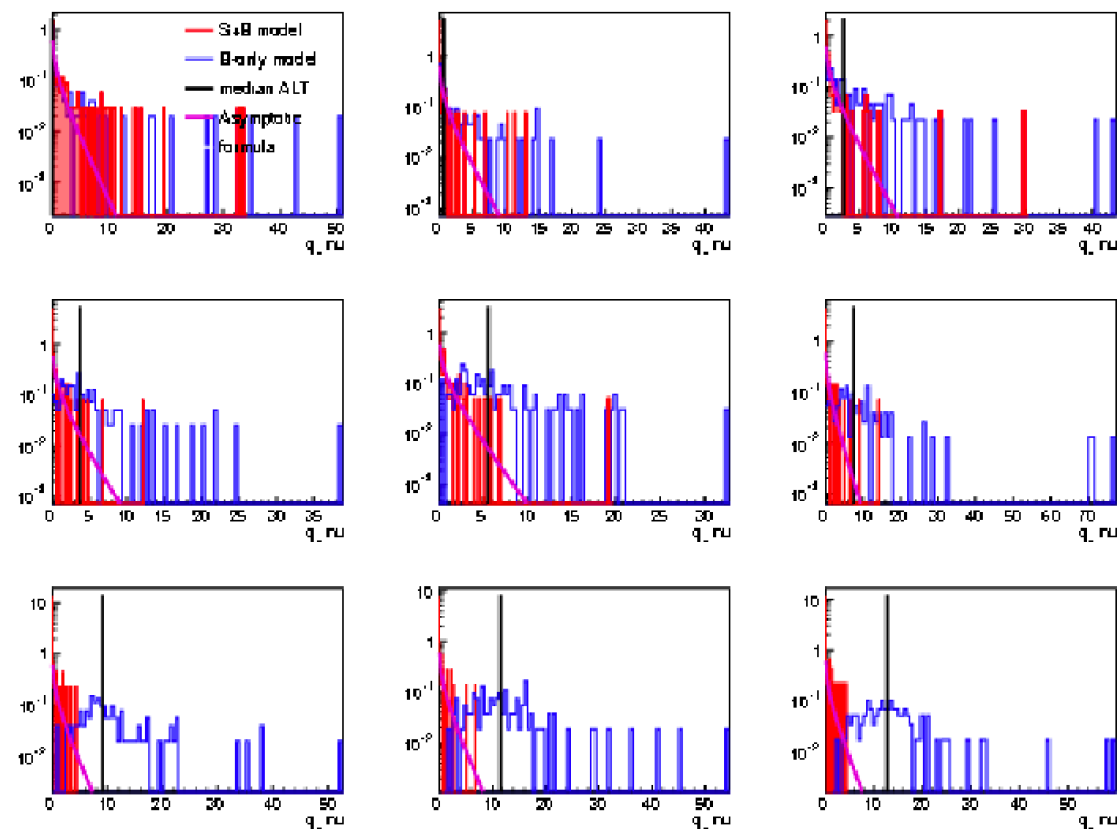


Done:

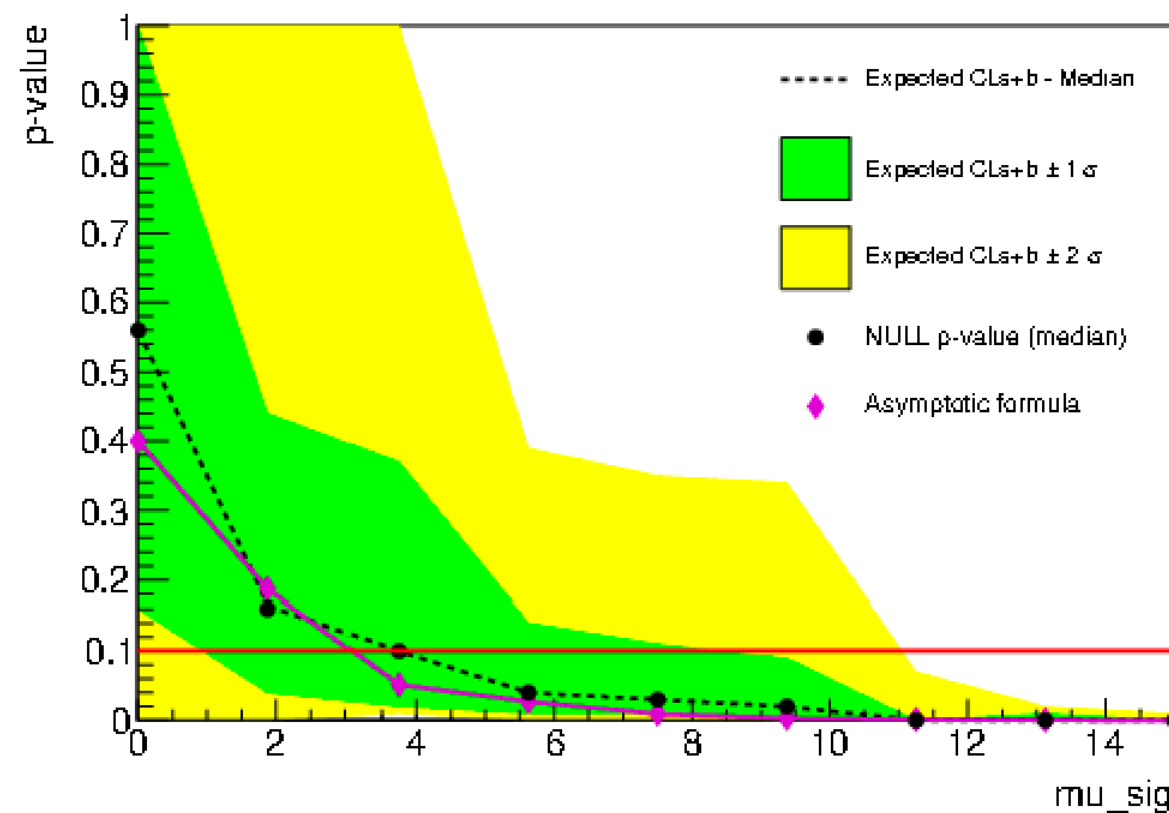
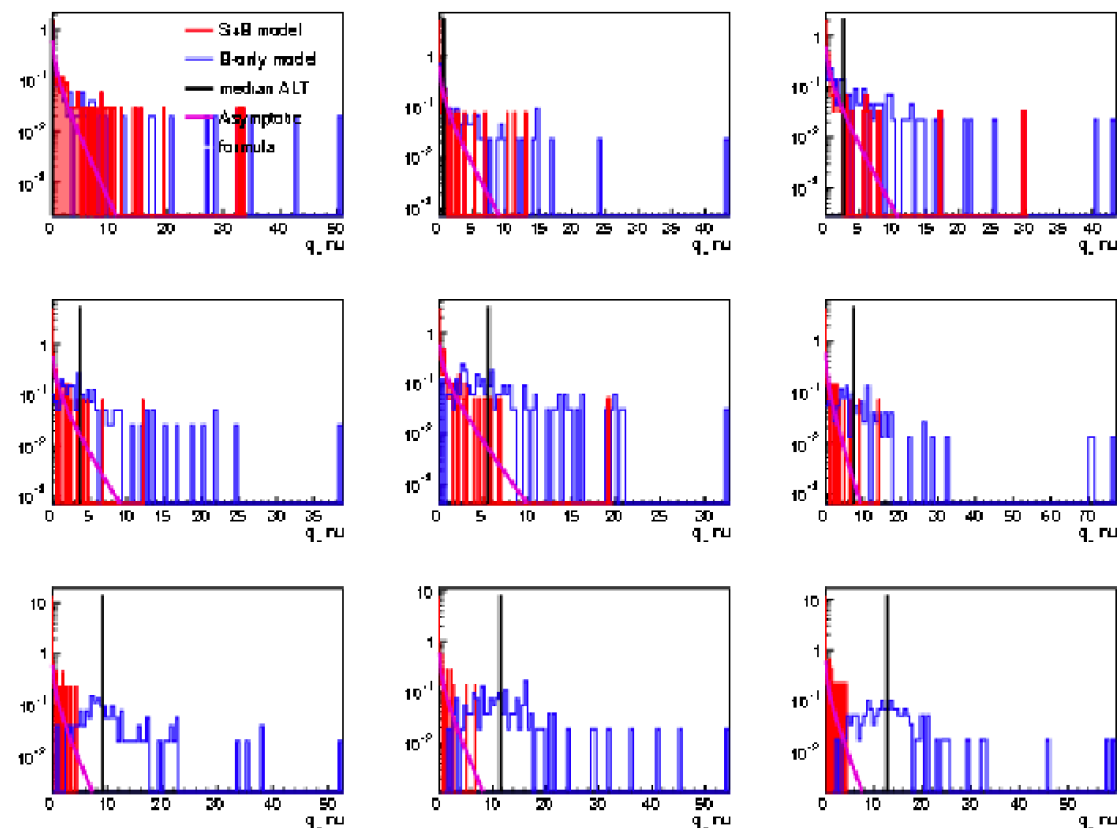
- For mass = 2.0000 GeV
 - ERROR: WIMP histogram is empty for mass 2.0 → Mass 2.0 is too small
- Mass: 10, 16.2222, 27.8256, 46.4159, 66.6667, 100, 200, 1000
- Moved the output from scratch to hdfs.
- Settings: 100 Ntoys, 1e7 samples, take 9 POI, 1000 livedays



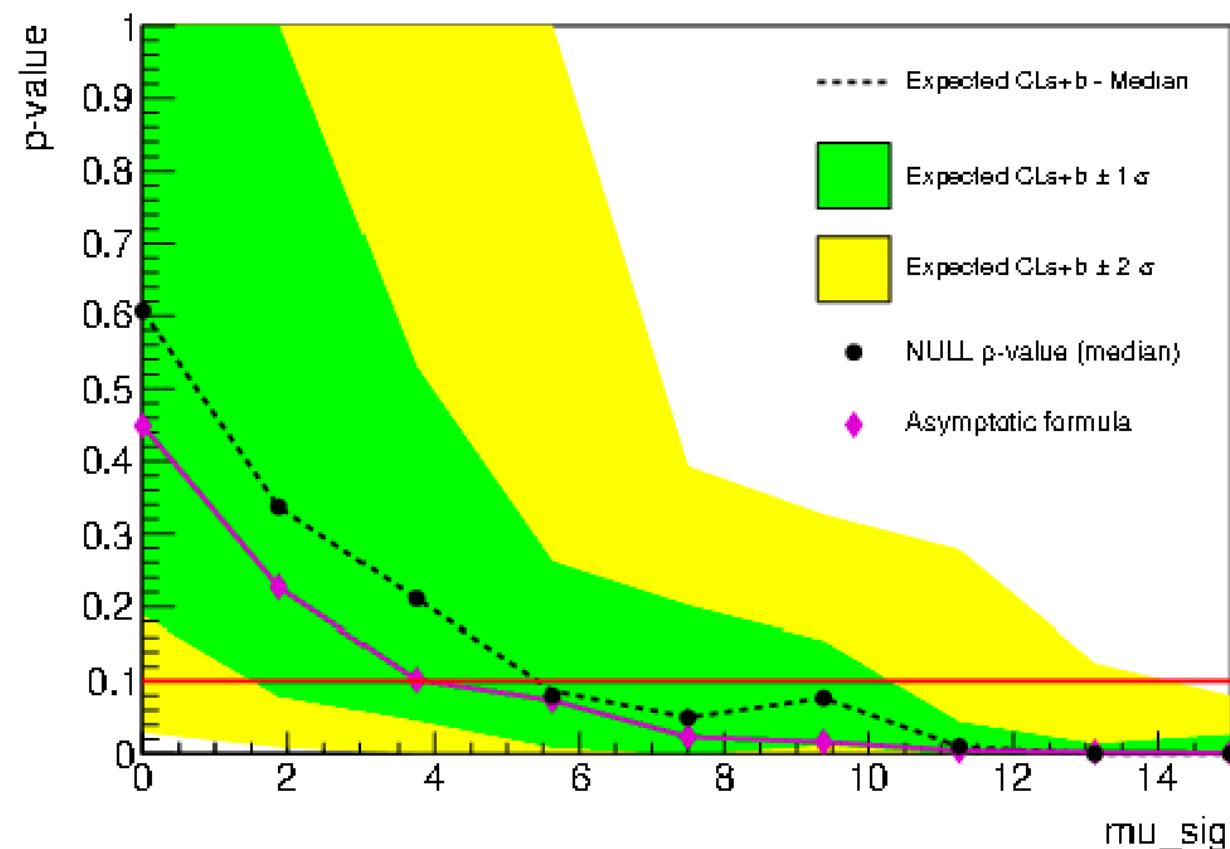
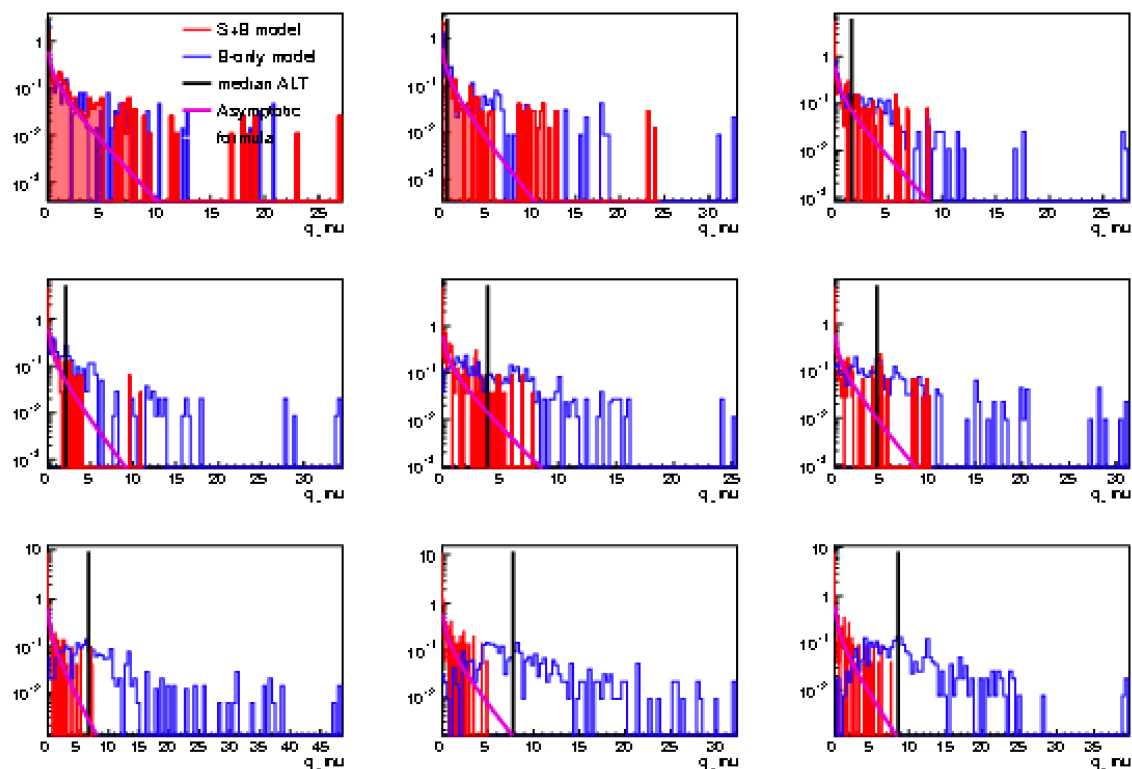
Mass = 16.2222

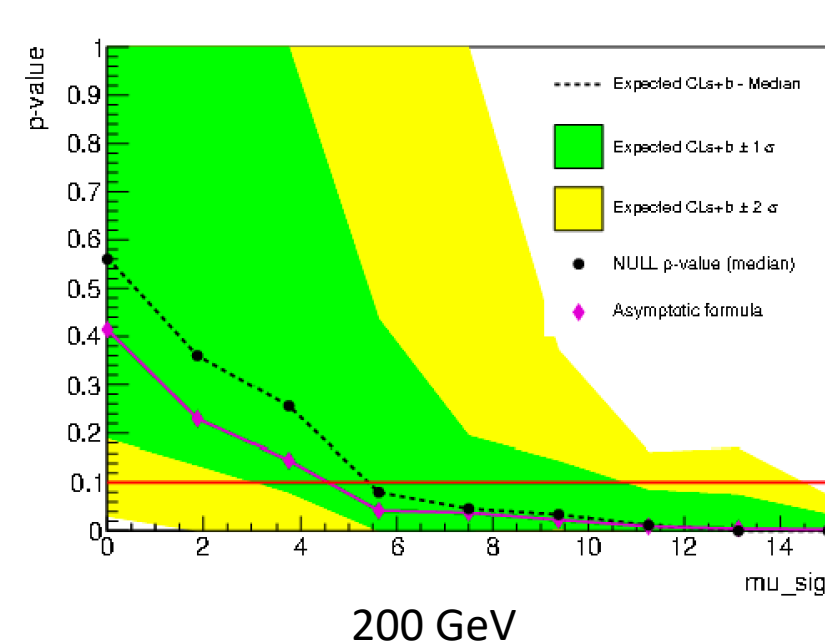
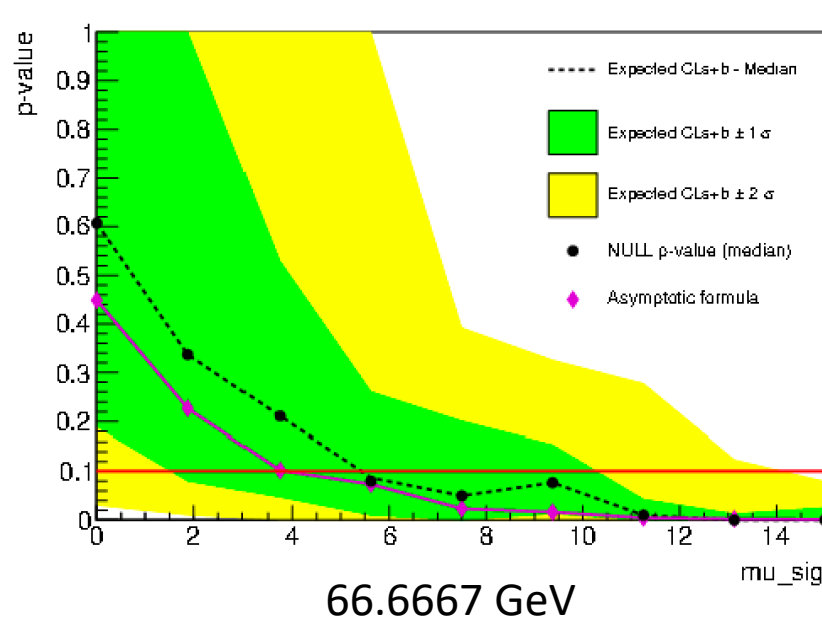
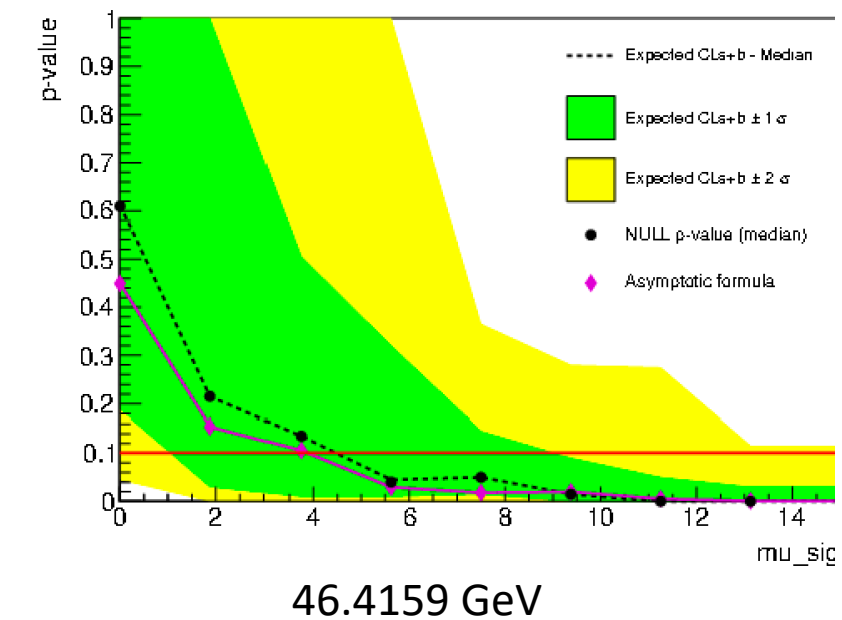
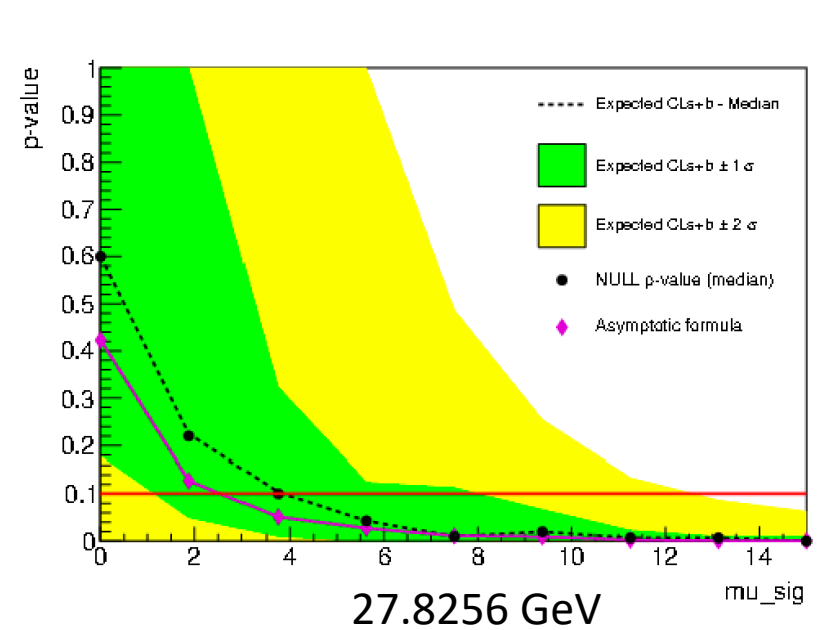
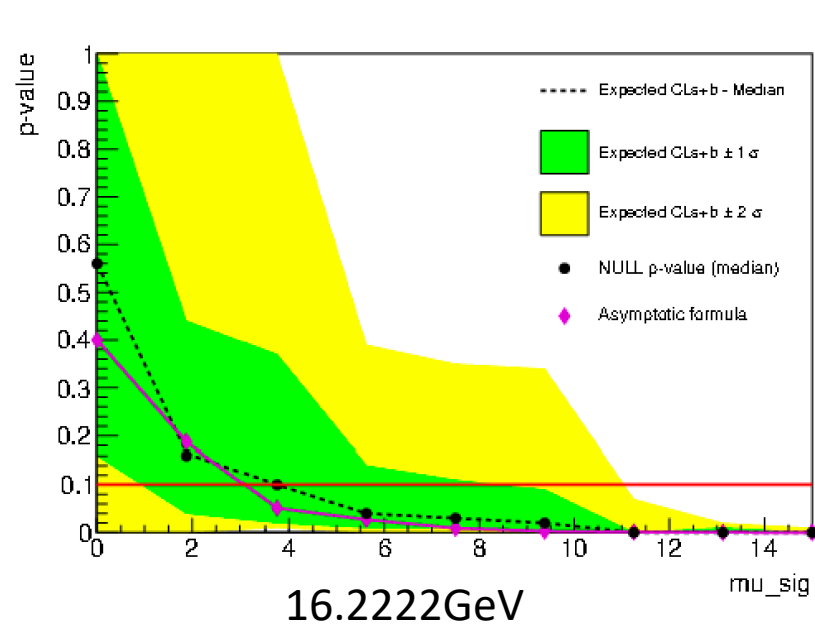
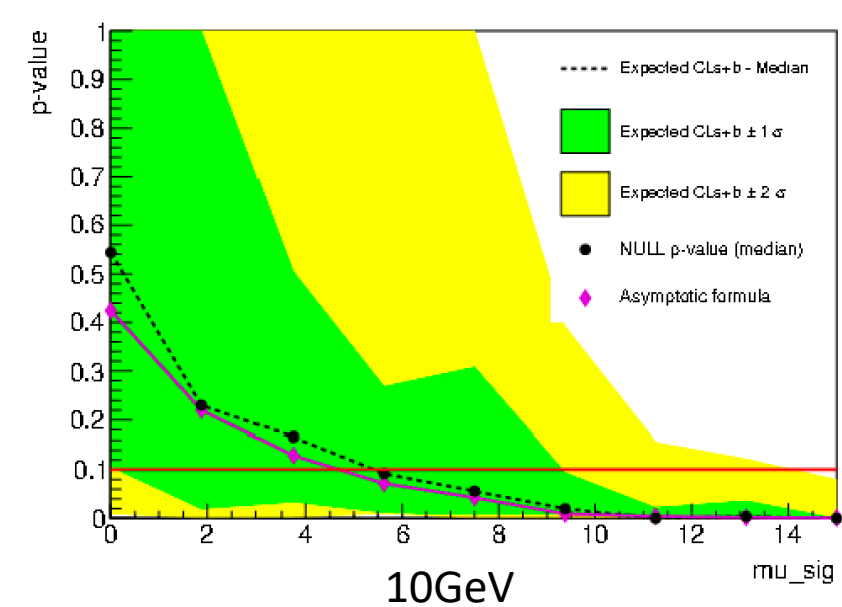


Mass = 16.2222

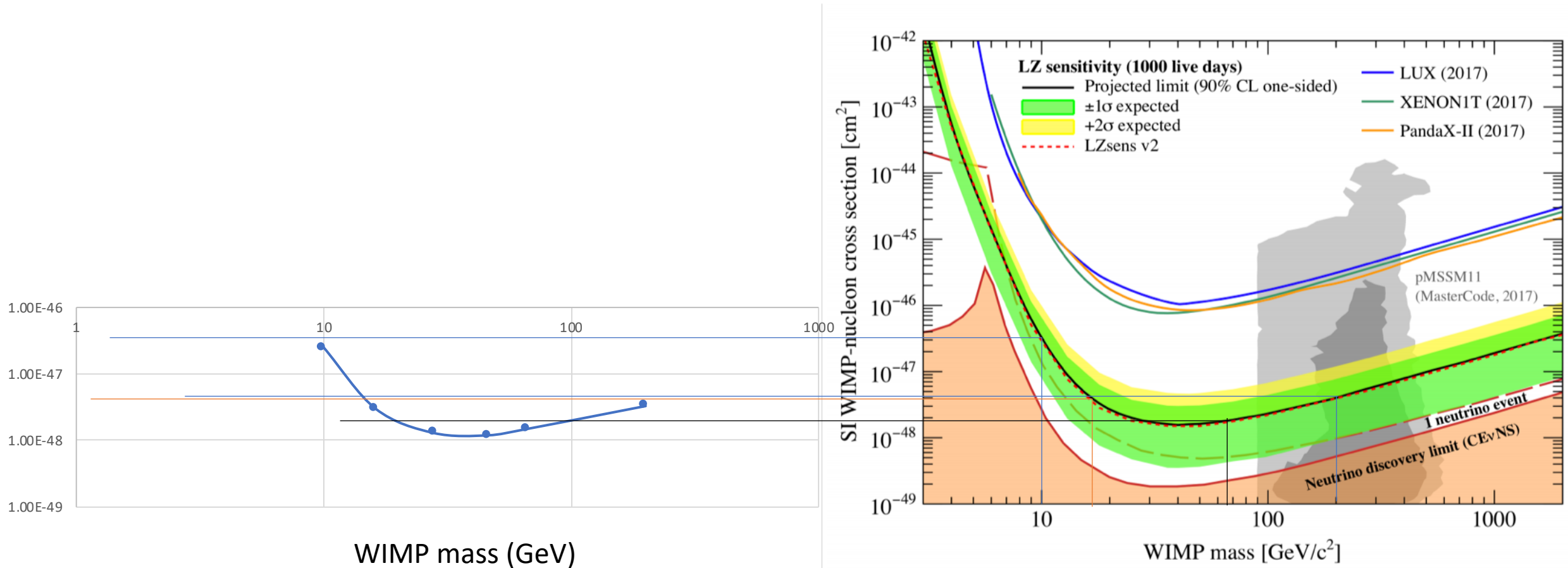


Mass = 66.6667





log scale lz sensitivity :



Understanding the structure of LZStats

- **Submit_MassJobs.sh:**
 - calls for the other two scripts, the one that submits jobs to the cluster
 - Kr mass = 2.0000 & 66.6667
 - #jobs for each mass = 10
- **runWiscJobs.py:**
 - Defines variables and parameters for a job running
- **Submit_lzstats.sh:**
 - When run Submit_MassJobs.sh, each job runs this script on the cluster
- **Setup.sh:**
 - set up the environment

Problem

- For one mass, when extracting 10 tar.gz files, only get one result “...run1”
- Check the error text, ImportError

```
ImportError: /cvmfs/lz.opensciencegrid.org/external/Python/2.7.15/x86_64-centos7-  
-gcc7-opt/lib/python2.7/lib-dynload/array.so: undefined symbol: _PySlice_Unpack  
Info in <RooWorkspace::SaveAs>: ROOT file ./output/lzstats-LZ projected 1e7-WIMP
```

- Ask Jonathan:
 - For the Jobs submitted to the cluster, the python cannot load the yaml package → unable to change the run number → all the runs have the same run number, run1
- The yaml package is exported in setup.sh