Summer 23: Global Fitting Update

Emrys Peets

06/20/2023

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Road Map

1. Recap of background, milestones
2. Key Updates to global fitting code infrastructure and methodology
3. Preliminary Automatic Fitting Study
4. Higher Statistics Studies
    a. Rise Only Study
    b. Global Fit Study
    c. Likelihood Study

Important vocab to keep in mind

- Window Range: Range by which a function is fit, generally of the form [WinMin, WinMax]
- WinMax: Maximum value for a given window range
- WinMin: minimum value for a given window range
- IMD: Invariant Mass Distribution

Global Fit to the Invariant Mass Distribution

Initial functions

$$f_{dijet1}(x) = \frac{p_0(1-x)^{p_1}}{x^{p_2}} \qquad f_{dijet2}(x) = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3\log(x)}}$$

$$f_{dijet3}(x) = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3\log(x)+p_4\log^2(x)}} \qquad f_{ATLAS1}(x) = \frac{p_0(1-x^{1/3})^{p_1}}{x^{p_2}}$$

$$f_{ATLAS2}(x) = \frac{p_0(1-x^{1/3})^{p_1}}{x^{p_2+p_3\log^2(x)}} \qquad f_{UA2_1}(x) = p_0 x^{p_1} e^{p_2 x}$$

$$f_{UA2_2}(x) = p_0 x^{p_1} e^{p_2 x+p_3 x^2} \qquad f_{UA2_3}(x) = p_0 x^{p_1} e^{p_2 x+p_3 x^2+p_4 x^3}$$

$$f_{cmsBH1}(x) = \frac{p_0(1+x)^{p_1}}{x^{p_2\log x}} \qquad f_{cmsBH2}(x) = \frac{p_0(1+x)^{p_1}}{x^{p_3+p_2\log x}}$$

$$f_{ATLASBH1}(x) = p_0(1-x)^{p_1}x^{p_2\log(x)} \qquad f_{ATLASBH2}(x) = p_0(1-x)^{p_1}(1+x)^{p_2\log(x)}$$

$$f_{ATLASBH3}(x) = p_0(1-x)^{p_1}e^{p_2\log(x)} \qquad f_{ATLASBH4}(x) = p_0(1-x^{1/3})^{p_1}x^{p_2\log(x)}$$

$$f_{ATLASBH5}(x) = p_0(1-x)^{p_1}x^{p_2 x)} \qquad f_{ATLASBH6}(x) = p_0(1-x)^{p_1}(1+x)^{p_2 x}$$

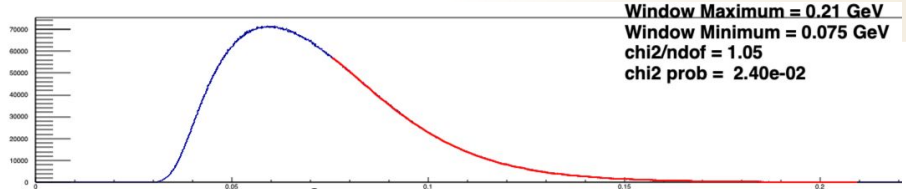
We multiply each function by an error function to fit the rise:

Error function used:

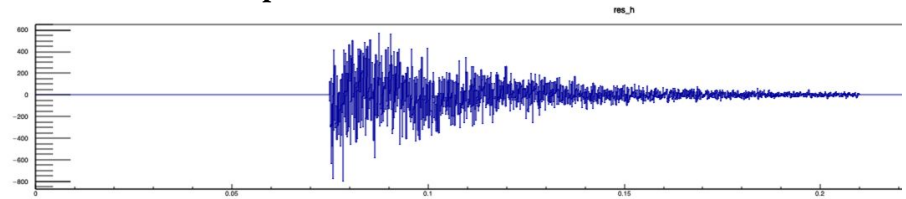
$$\mathrm{Er}(x) = \frac{1}{2}\left(\mathrm{Erf}\left(\frac{(x-[q_0])}{[q_1]}\right)+1\right)$$

C. Bravo. *Thesis linked here*

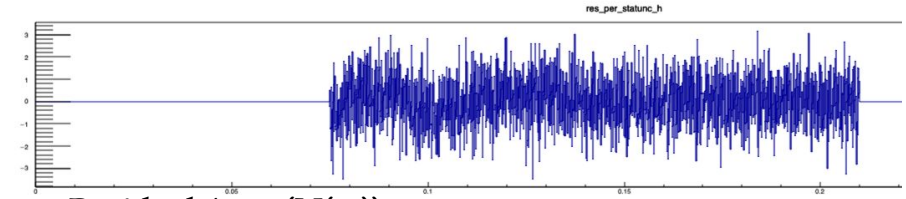
Representative "Good" Fit Using Global Fitting Tool



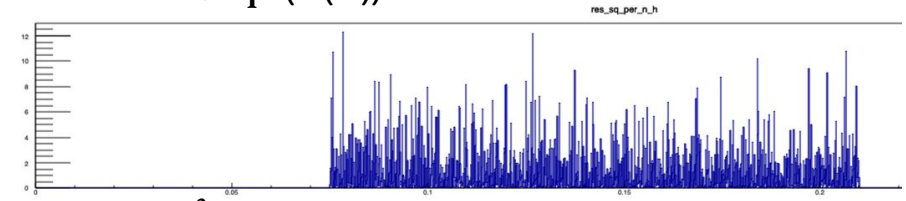
Function on top of IMD



Residual Plot

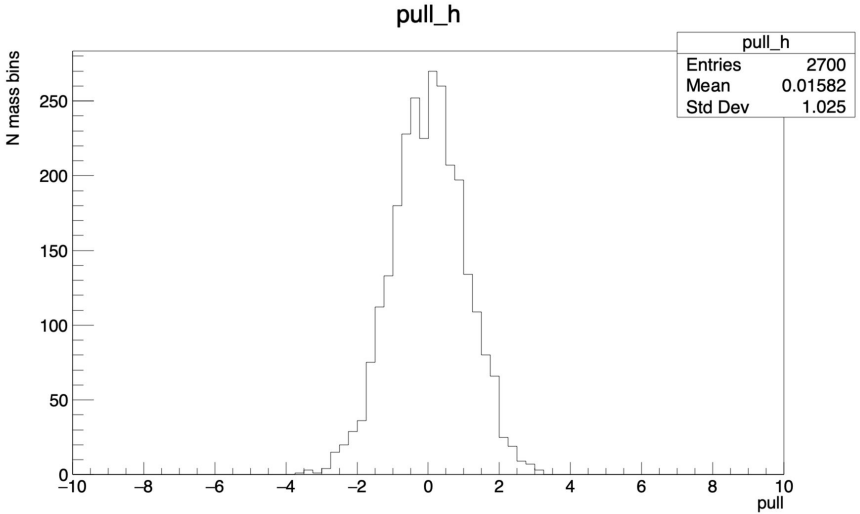


Residual / sqrt(N(m))



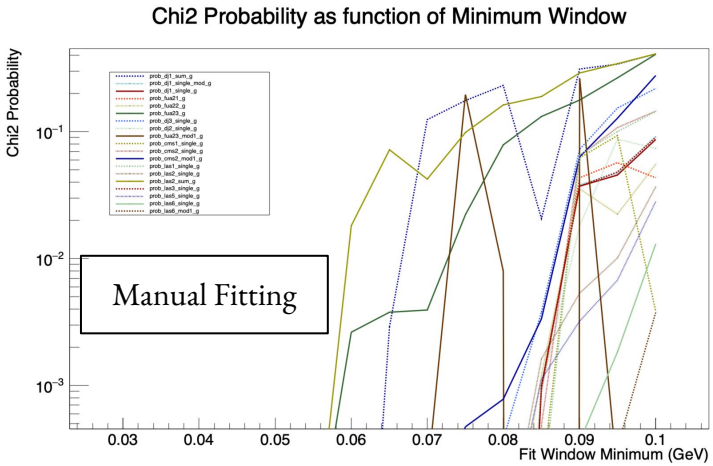
Residual^2 / (N(m))

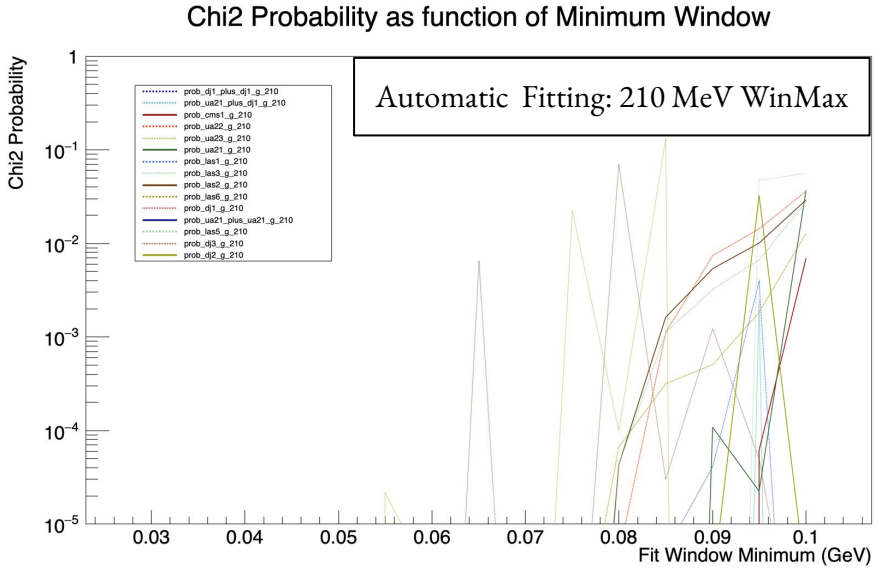
- UA23 Function
- Fit Range: 75 MeV - 210 MeV
- Good ⇒ pvalue >10^{-2}



Recap on Results from Collaboration Meeting

- Manual Fitting
  - 19 functions, 200-300 iterations, <11 parameters, 14 Window Ranges (1 WinMax)
- low stats automated fitting (15 Functions, 250 iterations, 126 Window Ranges)





What's new? (1/2)

- removed sum function generator from global fitting script
  - now able to create input parameter and function files for the sum of two independent functions before running fitting (allows massive scale up of total functions testable)
- store best fit parameters for each window in txt file (great for viewing parameters)
  - [win_min win_max best_param1 best_param2 best_param3 ... chi2/ndf pvalue]
- changed fitting logic to extend beyond local minimums
  - for each iteration, width of generated gaussian increases by .01*(iteration number)*(initial mean)
- Modified terminal input to utilize additional parameter txt file for every function
- integrated workflow into SSH to generate fitting script for each function to run remotely

Sum Function Generator

Terminal Input:

- python3 sum_fun_gen.py -i ./functions/[function1.txt] -f ./functions/[function2.txt] -d ./functions/ -e ./parameters/

Expected Output:

- generates function1_plus_function2.txt file in /resonance_fitting/functions/
  - with m (= f1+ f2) many parameters of the form [0],[1],..,[m-1]
- generates function1_plus_function2.txt file in /resonance_fitting/parameters/
  - created using starting parameters of summands of the form [p1 p2 ... pm]

Making global fitting scripts for every function

Terminal Input

python3 /sdf/group/hps/users/epeets/run/resonance_fitting/makeGlobalFitScripts.py -d /sdf/group/hps/users/epeets/run/resonance_fitting/sh/ -m 28 40 1 -x 40 72 2 -F /sdf/group/hps/users/epeets/run/resonance_fitting/functions/

(WinMin,WinMax)

Expected Output

- resonance_fitting/sh/subJob_28_70.sh (to sbatch each function)
- resonance_fitting/sh/sh_28_70/[function.sh]

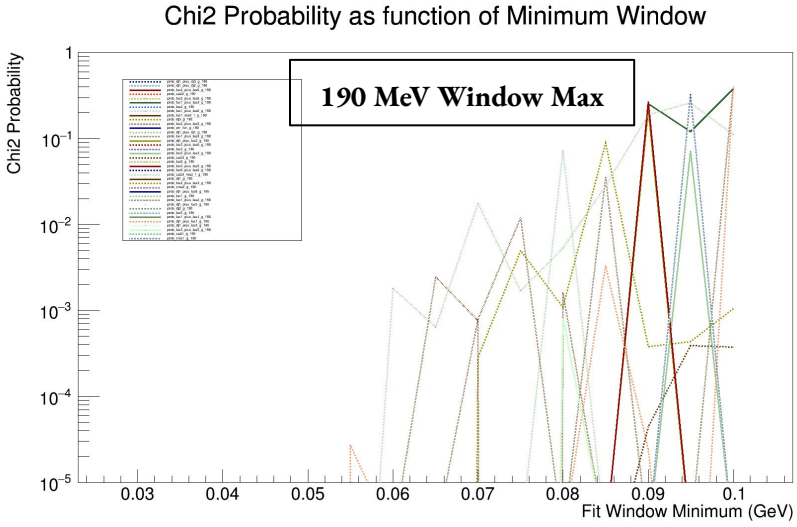
Automated fitting terminal input

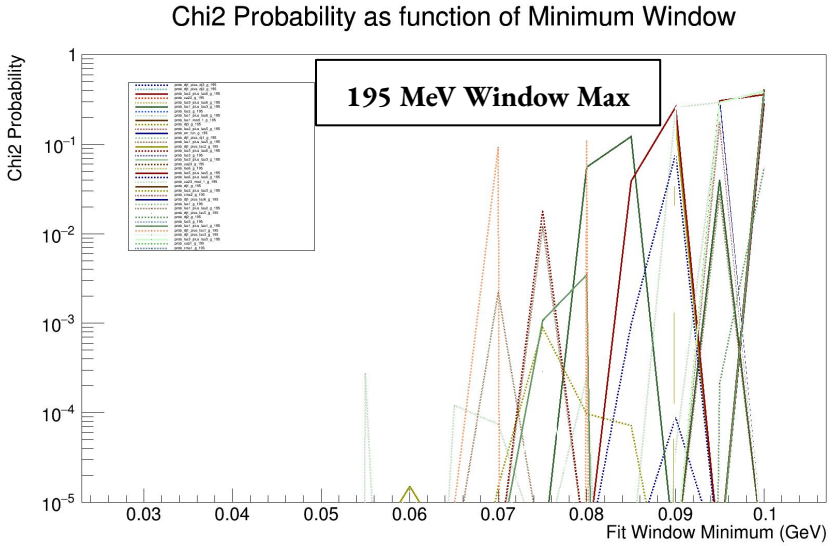
What's new? (2/2)

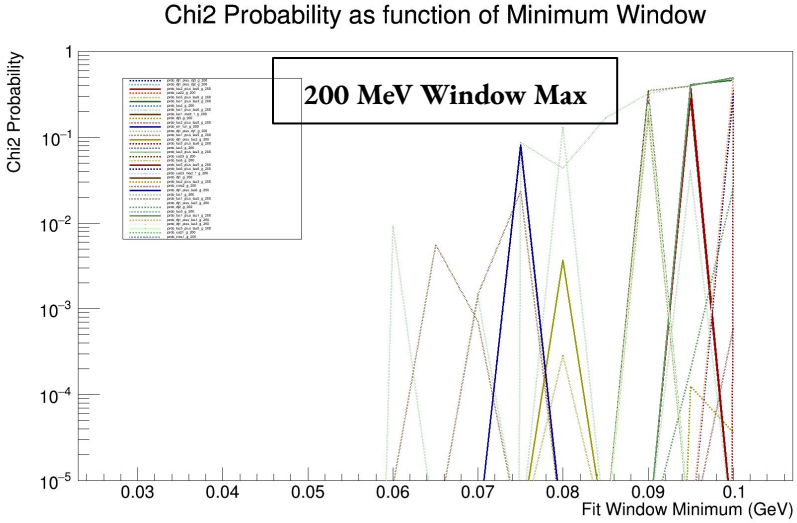
- Discovered bug that caused the failure of all >10 parameter fits (*thanks Cam*)
  - offers solid strategy towards finding *the one true function*
- cleaned code to run more efficiently
- Started process of performing likelihood fits in addition to chi2 fits
- Generously scaled up total functions being used in tests
  - new class of functions without error function
  - mixing and matching functions
  - frankenstein functions
- conducted preliminary study making use of full fitting infrastructure
- began higher statistics study for global range and only rise range

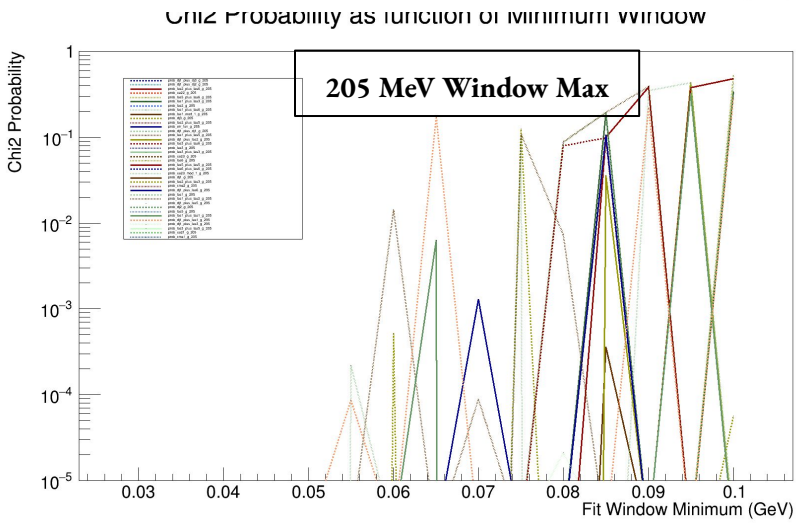
Preliminary Automated Fitting Study

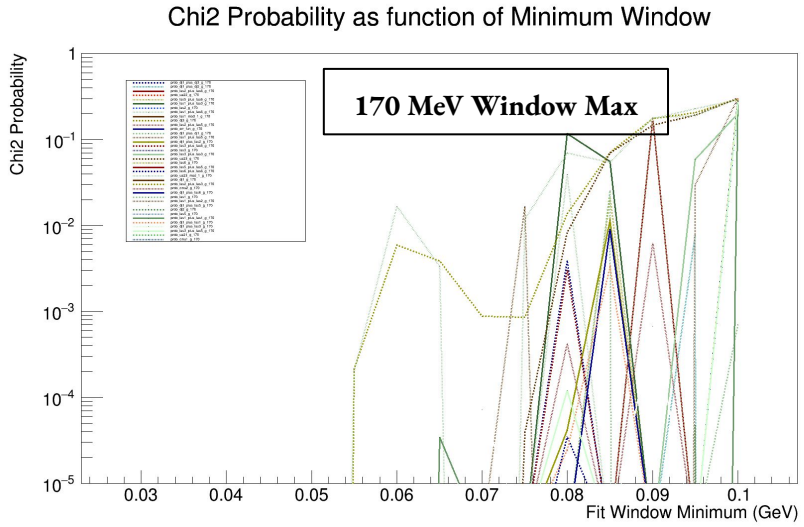
- ~40 functions
- 300 iterations for each window range
  - window range varied win_min from (30,100) 5 MeV Steps
  - window range varied win_max from (170,210) 5 MeV Steps
  - fit parameters with best Chi2/Ndof are stored and used as seed for final fit of each window range
  - Chi2 Probability (pvalue) stored for each window and plotted about the window range for each function
- Utilized new parameter seeding logic

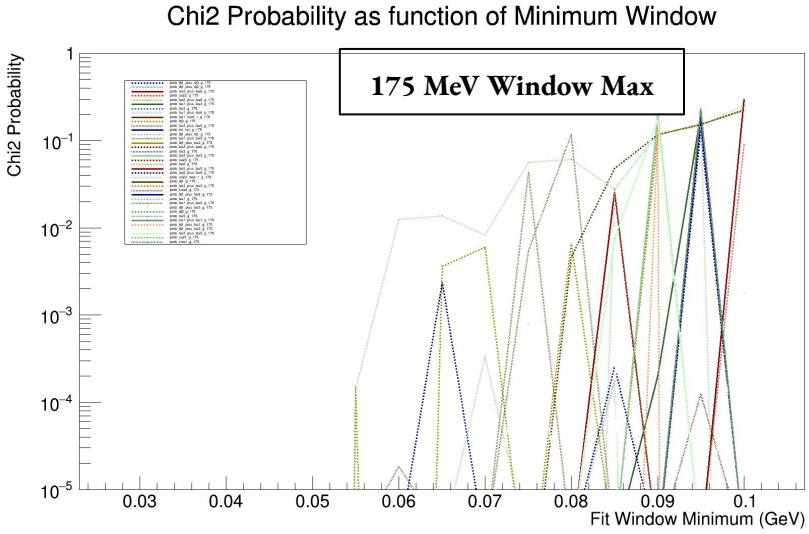

Chi2 Probability as function of Minimum Window
190 MeV Window Max
Chi2 Probability
Fit Window Minimum (GeV)

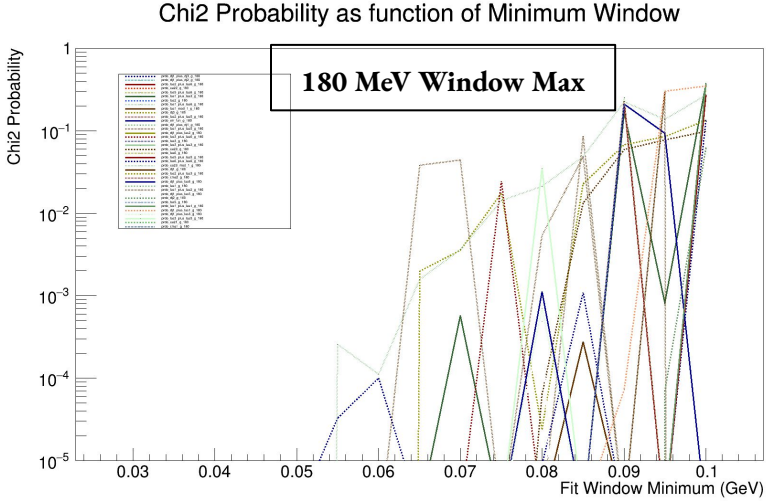

Chi2 Probability as function of Minimum Window
195 MeV Window Max
Chi2 Probability
Fit Window Minimum (GeV)

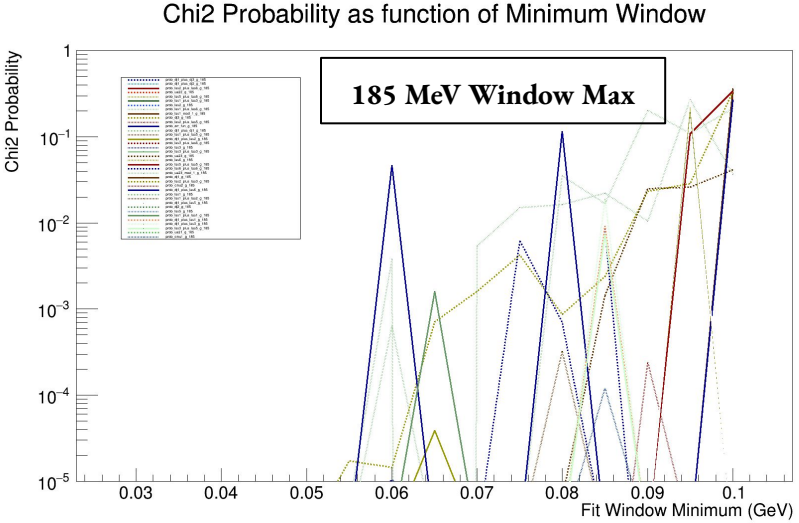

Chi2 Probability as function of Minimum Window
200 MeV Window Max
Chi2 Probability
Fit Window Minimum (GeV)


205 MeV Window Max
Chi2 Probability
Fit Window Minimum (GeV)

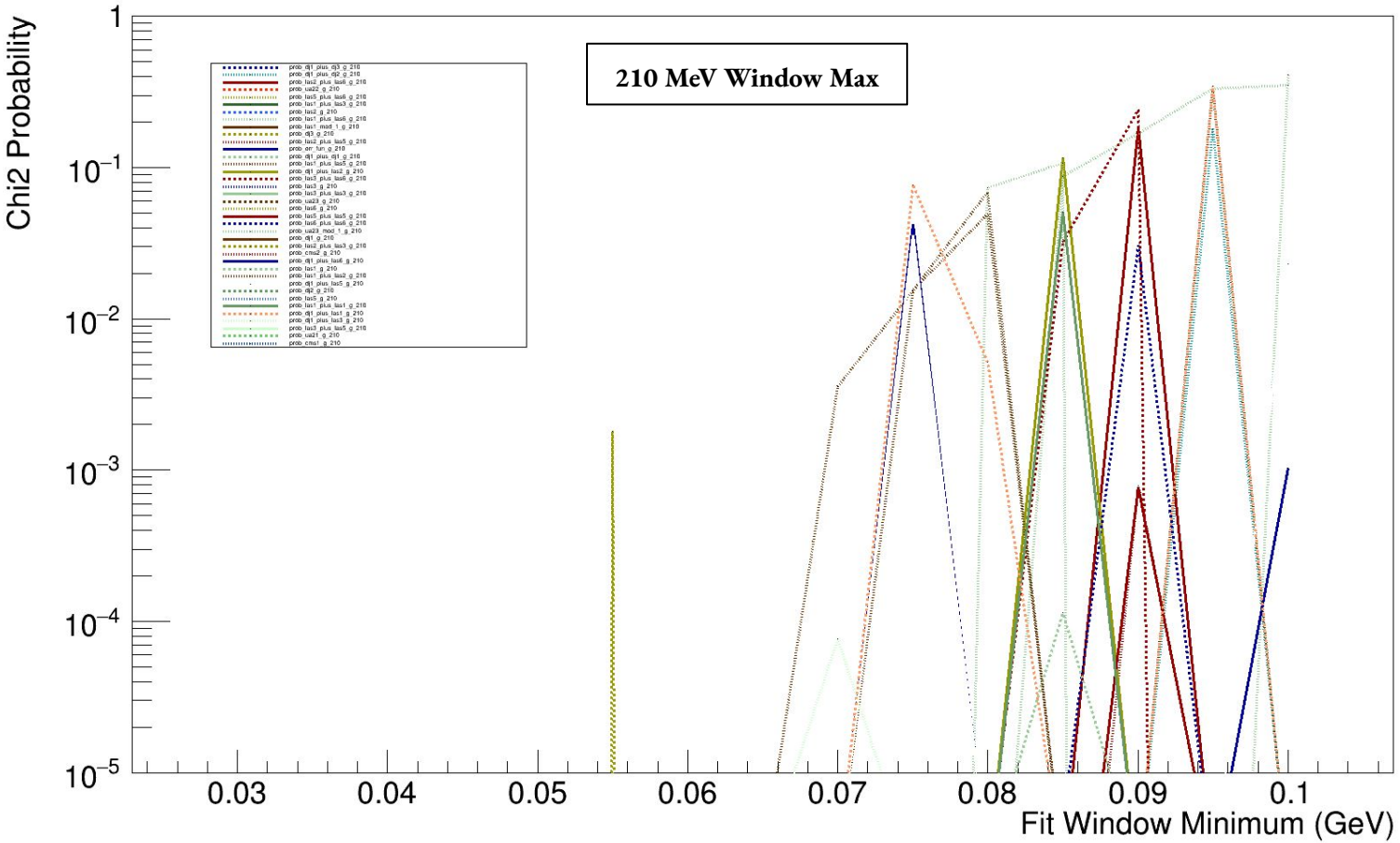

Chi2 Probability as function of Minimum Window
170 MeV Window Max
Chi2 Probability
Fit Window Minimum (GeV)


Chi2 Probability as function of Minimum Window
175 MeV Window Max
Chi2 Probability
Fit Window Minimum (GeV)


Chi2 Probability as function of Minimum Window
180 MeV Window Max
Chi2 Probability
Fit Window Minimum (GeV)


Chi2 Probability as function of Minimum Window
185 MeV Window Max
Chi2 Probability
Fit Window Minimum (GeV)

Chi2 Probability as function of Minimum Window



Three Studies Ongoing

- Two Chi2 fit studies (1000 iterations)
  - Rising Edge Fit Study

    WinMin: [28 MeV - 40 MeV] 1 MeV Steps

    WinMax: [40 MeV - 72 MeV] 2 MeV Steps
  - Global Fit Study (1000 iterations)

    WinMin: [30 MeV - 99 MeV] 3 MeV Steps

    WinMax: [180 MeV - 210 MeV] 3 MeV Steps
- One likelihood fit study (600 iterations)

  WinMin: [30 MeV - 84 MeV] 3 MeV Steps

  WinMax: [180 MeV - 210 MeV] 3 MeV Steps

Existing Functions (64 functions)

cms1_plus_dj3.txt	dj1_plus_dj3.txt	dj2.txt	las1_plus_ua22.txt	las5_plus_las6.txt
cms1.txt	dj1_plus_err_fun.txt	dj3_mod_er_plus_ua23_mod_er.txt	las1_plus_ua23.txt	las5.txt
cms2.txt	dj1_plus_las1.txt	dj3_mod_er.txt	las1.txt	las6_plus_las6.txt
dj1_mod_er_plus_dj1_mod_er.txt	dj1_plus_las2.txt	dj3.txt	las2_plus_las2.txt	las6.txt
dj1_mod_er_plus_dj2_mod_er.txt	dj1_plus_las3.txt	err_fun.txt	las2_plus_las3.txt	rise_logi_plus_err_fun.txt
dj1_mod_er_plus_dj3_mod_er.txt	dj1_plus_las5.txt	las1_mod_1.txt	las2_plus_las5.txt	rise_logi.txt
dj1_mod_er_plus_ua23_mod_er.txt	dj1_plus_las6.txt	las1_plus_err_fun.txt	las2_plus_las6.txt	ua21_plus_ua21.txt
dj1_mod_er.txt	dj1_plus_ua21.txt	las1_plus_las1.txt	las2.txt	ua21.txt
dj1_plus_cms1.txt	dj1_plus_ua22.txt	las1_plus_las2.txt	las3_plus_las3.txt	ua22.txt
dj1_plus_dj1_mod_er.txt	dj1_plus_ua23_mod_er.txt	las1_plus_las3.txt	las3_plus_las5.txt	ua23_mod_1.txt
dj1_plus_dj1.txt	dj1_plus_ua23.txt	las1_plus_las5.txt	las3_plus_las6.txt	ua23_mod_er.txt
dj1_plus_dj2.txt	dj1.txt	las1_plus_las6.txt	las3.txt	ua23.txt
dj1_plus_dj3_mod_er.txt	dj2_mod_er.txt	las1_plus_ua21.txt	las5_plus_las5.txt	

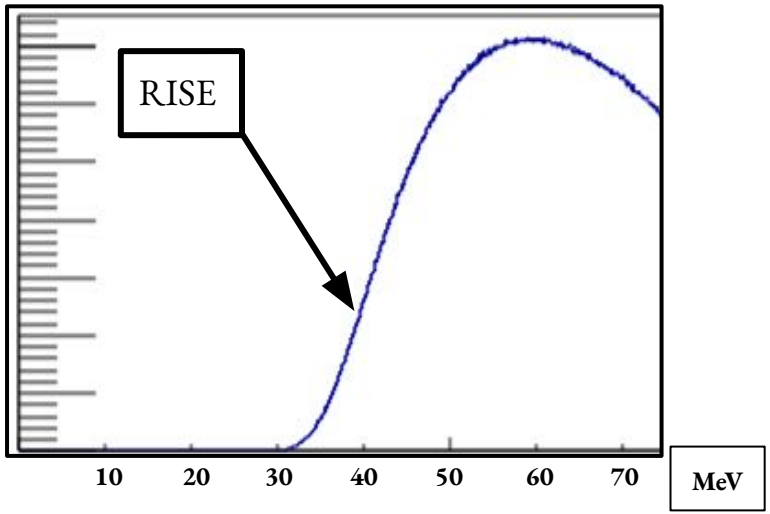
mod_er ⇒error function not used
plus_ ⇒ sum of functions and concatenating parameters
err_fun ⇒ error function
rise_logi ⇒ logistic function (error function alternative)
mod_1 ⇒ small change to initial function

Fitting the Rise of Background Distribution (28-70 MeV)

Necessary contingency if single function unable to fit global distribution.

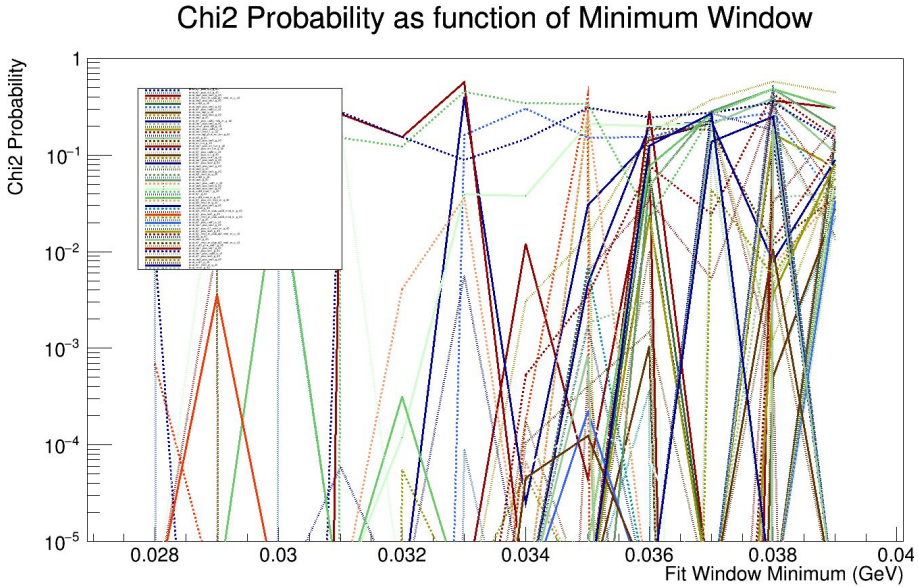
Purpose of rise study is to determine the component of a piecewise function dedicated to fitting only the rise of distribution.

Finer granularity (step size) useful for rapid rise of data collected.

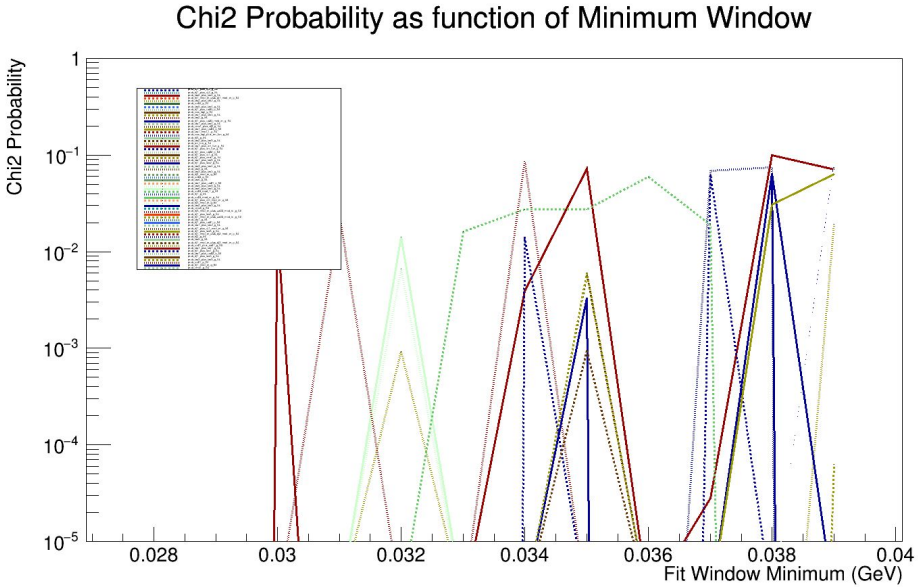


Fitting the Rise of Background Distribution (28-70 MeV)

40 MeV Fixed Win Max

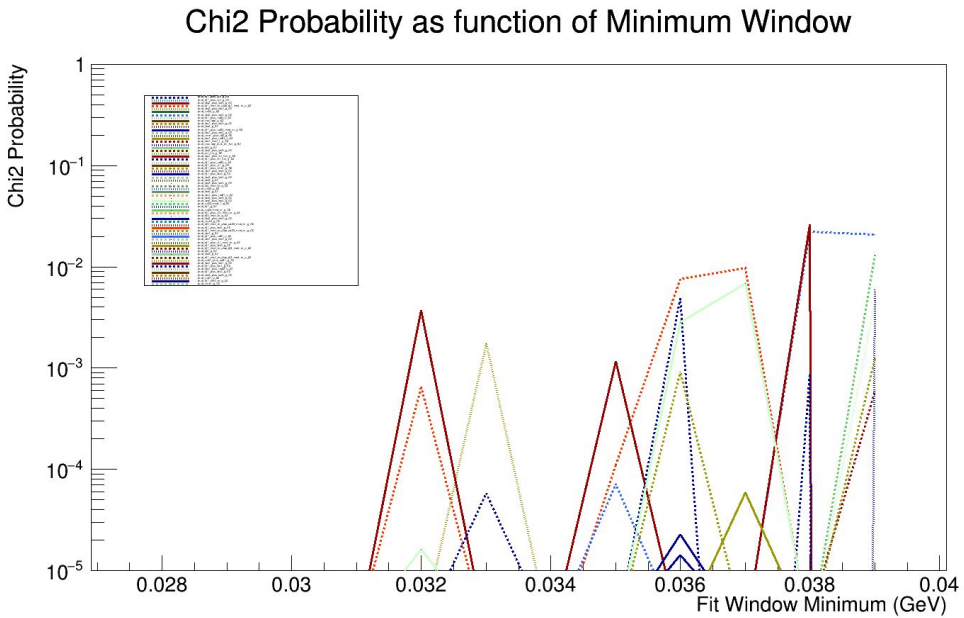


54 MeV Fixed Win Max

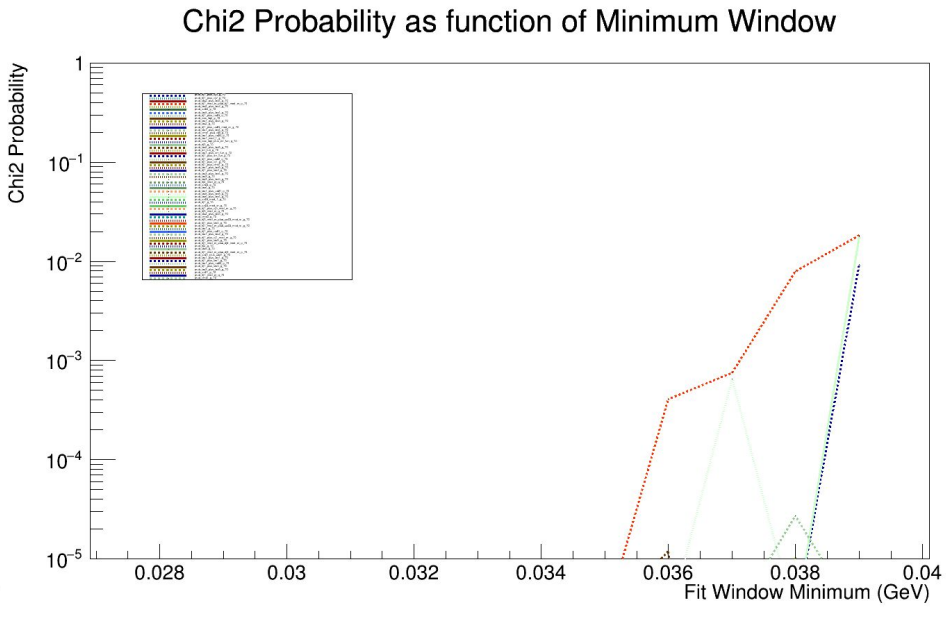


Fitting the Rise of Background Distribution (28-70 MeV)

62 MeV Fixed Win Max



70 MeV Fixed Win Max

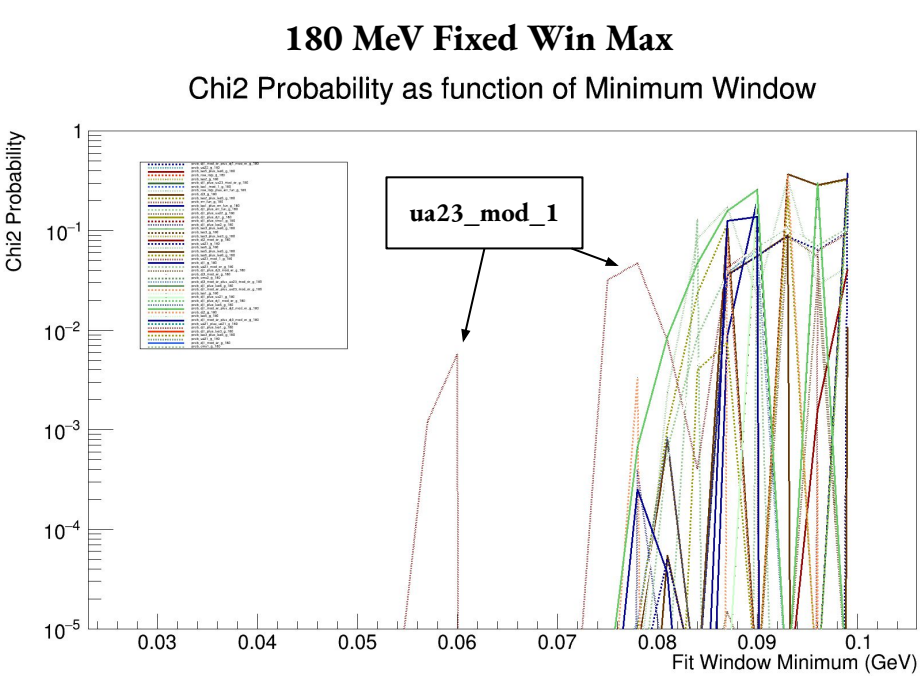


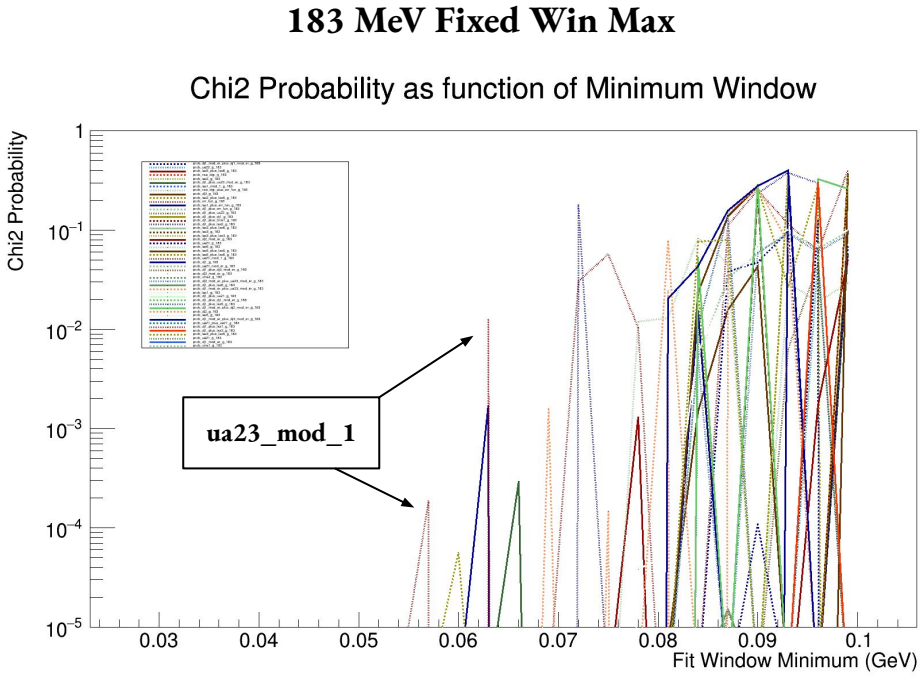
Key Observation! Broader range ⇒ less overall good fits

Higher Statistics Global Fitting

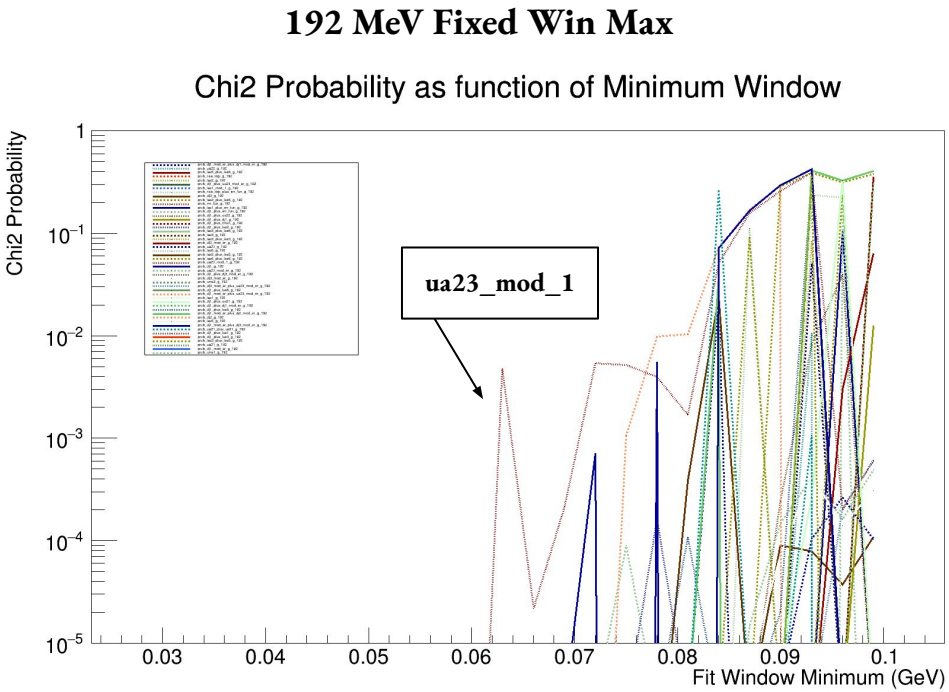
- ran into issues with fitting time
- ~10 functions w/ > 10 parameters completely failed
  - likely consequence of this being the first study where they have been fitting at all
  - other high parameter functions gave nonsensical parameters
- necessary to improve parameter logic to have higher rate of convergence

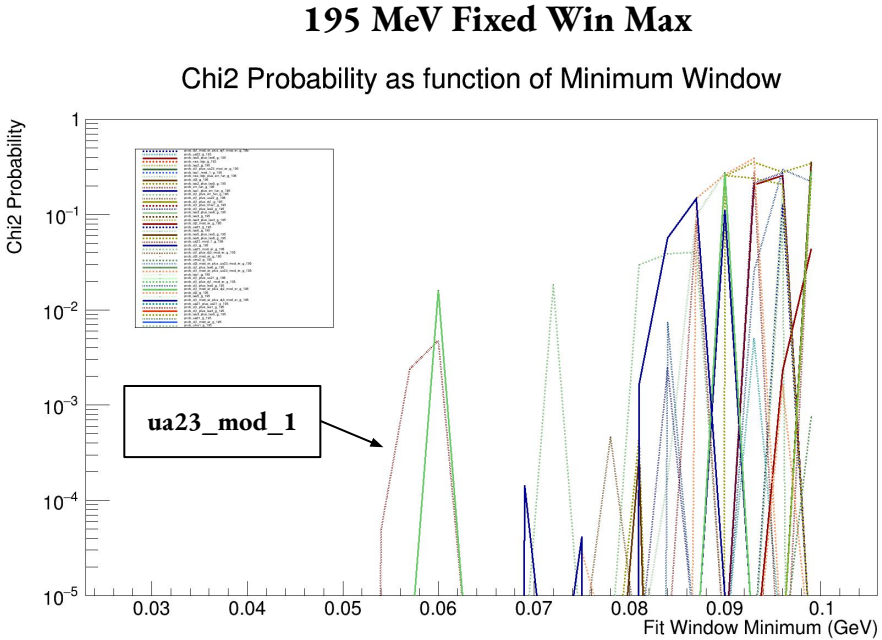
Selection of Higher Statistics Global Fits





Selection of Higher Statistics Global Fits

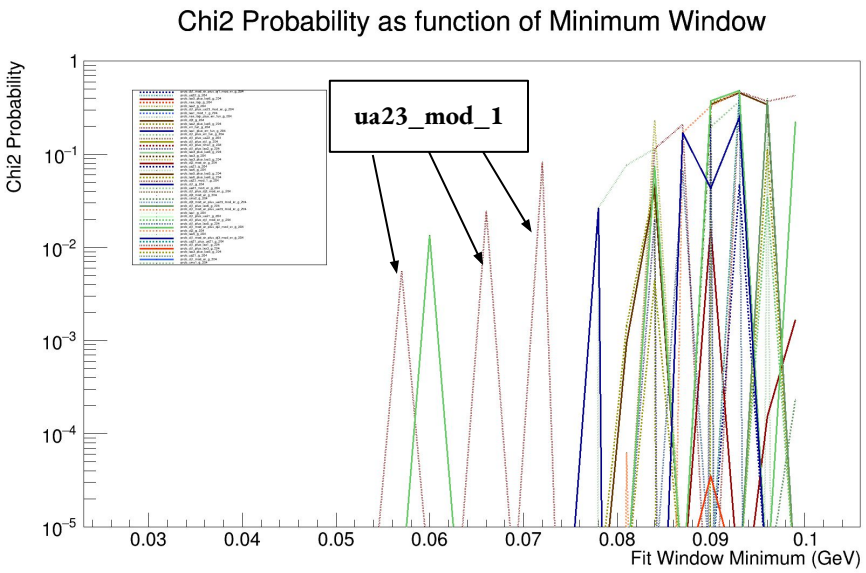




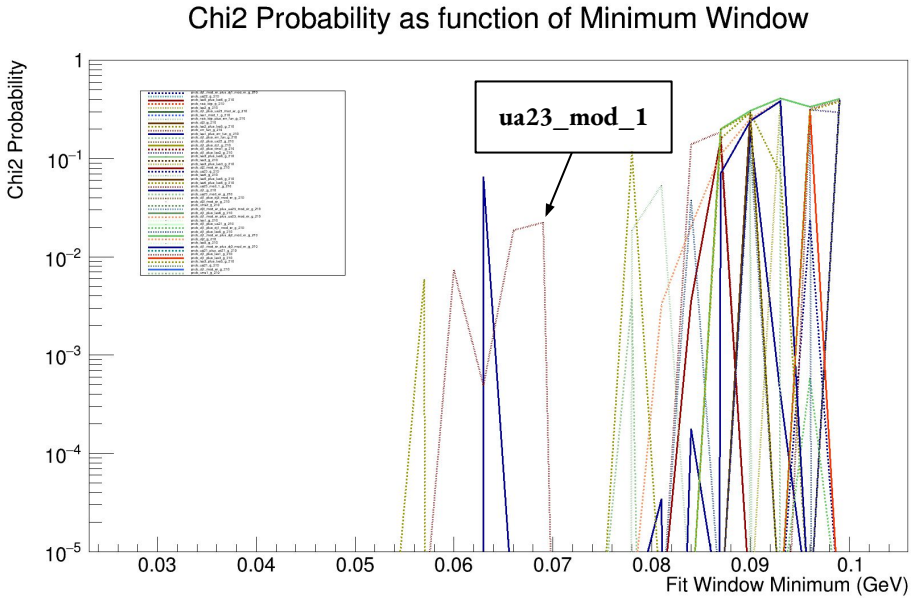
Selection of Higher Statistics Global Fits

SLAC

204 MeV Fixed Win Max



210 MeV Fixed Win Max



Tentatively Promising Functions (subject to change)

From the rise only study:

dj1_mod_er_plus_dj1_mod_er (no error function!)

[0.036, 0.062, 2.7275, -12.0245, 3.641, -386976.40, 65.5842, 1.1578, 0.007565]

[0.037, 0.062, 3.511, -8.7734, 3.7145, -20874.21, 47.8108, 1.15455, 0.0097726]

Note similar parameters for each window range

[0.036, 0.07, 426.865, 12.5755, 2.5965, -113100.192, 49.262, 1.19239, 0.000410]

[0.037, 0.07, 435.970, 12.2184, 2.5459, -212454.257, 53.0277, 1.18481, 0.000751]

[0.038, 0.07, 1494.30692, 17.2085, 2.2636, -263171.685, 51.5685, 1.140395, 0.00797777]

[0.039, 0.07, 3175.1608, 19.230, 1.9868, -1098238.979, 58.6093, 1.12298202, 0.0182566]

From Global Fit Study (incomplete study at the moment):

UA23_mod_1 (as illustrated in previous plots)

```
(TMath::Erf([8]*([7]*x-[1])/[0])+1)/2)  * [2]*TMath::Power(x,[3])*TMath::Exp([4]*x + [5]*x*x+ [6]*x*x*x)
```

mod

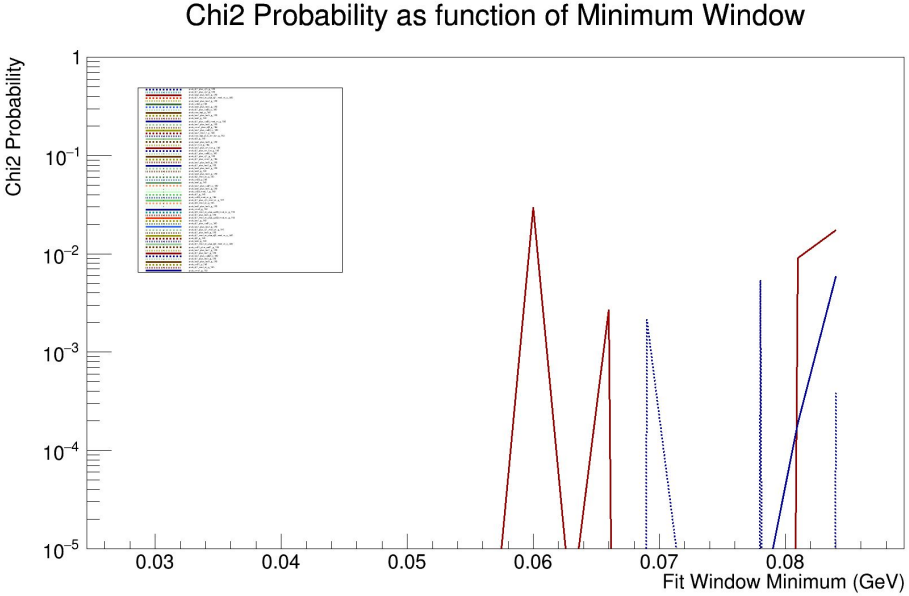
Depending on window range: this function consistently produces "good fits" from (57 MeV - 210 MeV)

Likelihood Study

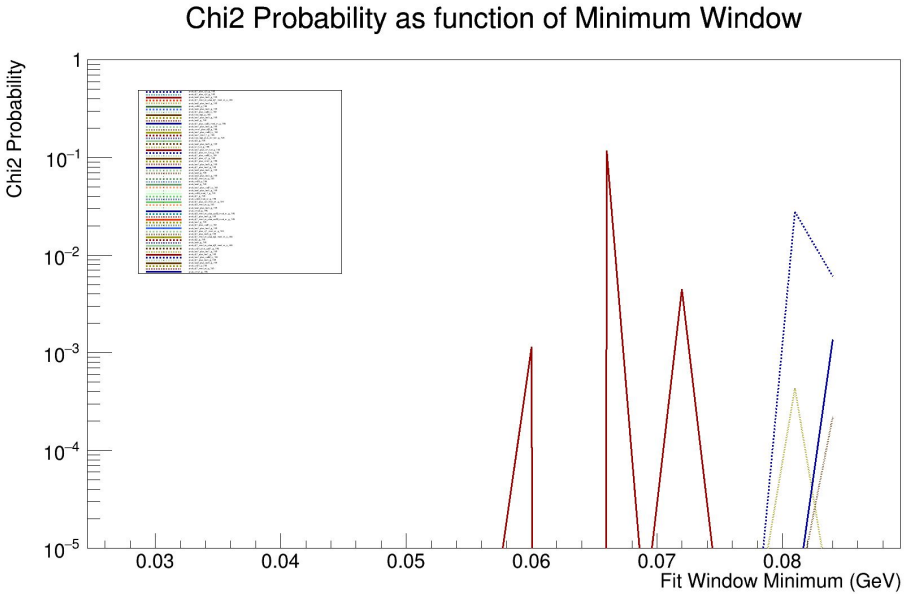
- now using “S” + “L” fitting option to produce likelihood fits of background
- Failed!
  - found that this fitting option is more sensitive to initial input parameters
  - have much room for improvement as the majority of functions failed to fit in any given window range

Small Selection of Likelihood Fits

180 MeV Fixed Win Max



195 MeV Fixed Win Max



What's Next?

- reimplementation of full global fitting toolset previously developed
  - summary plots, etc.
- high parameter function optimization
  - necessary for search of fabled *"one true function"*
- streamline code for efficiency
  - to assist in further scaling of functions, ideas
- construct "function checklist" to determine best way to filter insufficient functions
  - "good fit," parameter convergence across multiple window ranges, etc.