chi2 and u

Piecewise Fitting

Emrys Peets
Stanford University

dont read tea leaves

I.   RooFit now fitting (auto normalization)

    A.   rebinning fits predictably lowers chi2

        1.   Rebinning changes DOF and respective bin by bin contribution to chi2

            a)   chi2 probability for similar bkg parameters could be "much worse" but still low chi2

    B.   chi2 with rebinning of 40 mostly agrees with Gaussian Processes

II.   Fit display

    A.   New and Improved, modified TJs display scripts for roofit functionality

    B.   fixed incorrect fit being displayed with best chi2 per set of iterations

III. Signal Injector

    A.   Made signal injector that tracks mass resolution

IV.   "High Mass" Resonance Search

Updates since last update

I.   Success in blinding procedure

    A.   using piecewise method, only tested on rising edge of IMD

    B.   Albeit flawed in parameter projection, piecewise blinding converges well

II. RooFit Functionality for 2015 dataset

    A.   Wrote RooFit framework and codebase for fitting 2015

    B.   success in restricted range fits

Piecewise Blinding

Procedure (n=1, rising edge)

I. Fit on 6.5% within the range [33, 57] MeV.

    A. Store parameter and parameter errors from a "good fit".

II. Fit on 100% IMD within the range [34, 54] MeV

- Generate 100% parameter seeds

In each fit iteration

1. vary 6.5% parameters using a gaussian of width parameter error
2. set limit on acceptable 100% parameter by size of original parameter error
3. If 6.5% parameter was fixed

        → loosen limits to be within 40% above or below stored value

- Display fit with lowest chi2 (highest chi2 probability)
  - store parameters and parameter errors

Background Model Candidate Function Fit over 6.5% 2016 IMD

Fit Info

Bin Size: 50 keV
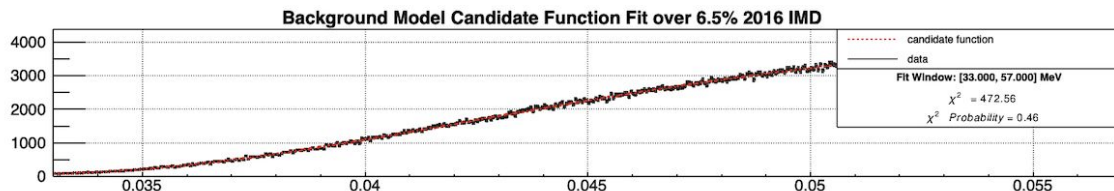
Range: [33, 57] MeV

Chi2_Prob = 0.46

Stored Parameters

```
[p10] = 490.6685532617223
[p1]  = 0.04240522602748158
[p2]  = 0.09136078735108702
[p3]  = 43.68193585247294
[p4]  = 15.584080643412133
[p5]  = -2.082364814201879
[p6]  = 0.012060290557698417
[p7]  = 0.03853733548225469
[p8]  = -750083.4855897021
[p9]  = 179.33614202803608
```
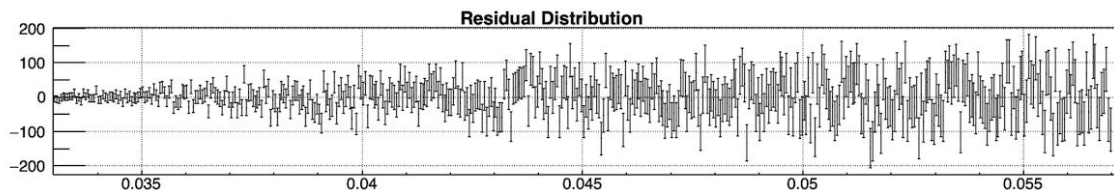
Stored Errors

```
[p10]_error = 0.6516690055867969
[p1]_error = 7.2522612743763926e-06
[p2]_error = 9.823004084683062e-06
[p3]_error = 0.0326775275677349
[p4]_error = 3.67050508957257
[p5]_error = 0.0
[p6]_error = 1.1582017141353873e-05
[p7]_error = 5.4019955336021e-06
[p8]_error = 0.0
[p9]_error = 0.021569718263066306
```
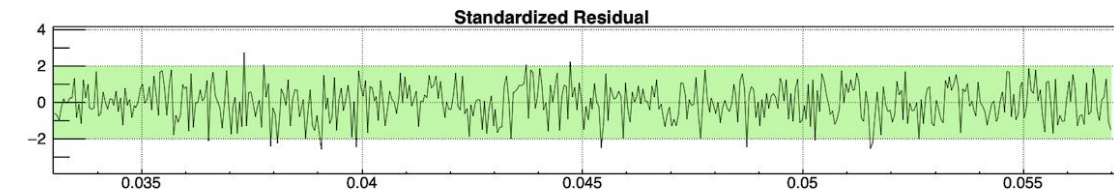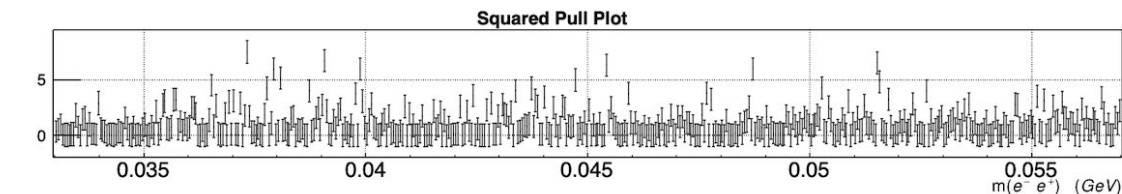
5

Piecewise Blinding   100% Rising Edge

Background Model Candidate Function Fit over 100% 2016 IMD

Residual Distribution

Standardized Residual

Squared Pull Plot

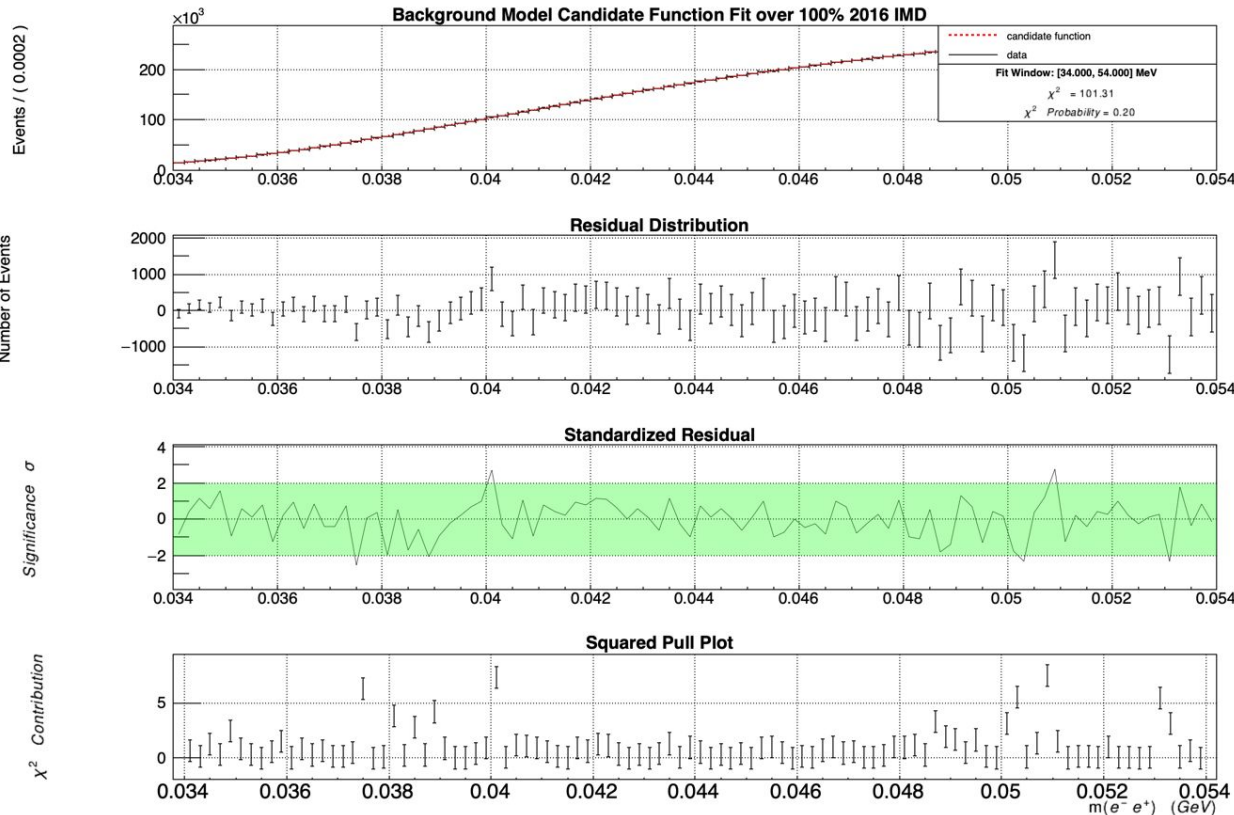## Fit Info

Bin Size: 200 keV

Range: [34, 56] MeV

Chi2_Prob = 0.20

Stored Parameters

```
[p10] = 490.27553602420863
[p1]  = 0.043907581659763495
[p2]  = 0.08774562074969872
[p3]  = 44.19496553033662
[p4]  = 14.09409491546085
[p5]  = -2.0608113099430136
[p6]  = 0.010018876062215187
[p7]  = 0.034984061109657903
[p8]  = -750082.8446750462
[p9]  = 179.93550550243555
```

Stored Errors

```
[p10]_error = 1.2799090630799697
[p1]_error  = 3.831848641555613e-06
[p2]_error  = 4.807991830724223e-06
[p3]_error  = 0.013652475025935473
[p4]_error  = 0.008890706489388833
[p5]_error  = 9.813823693516355e-05
[p6]_error  = 1.327521268384979e-05
[p7]_error  = 2.765154465400599e-06
[p8]_error  = 1.249310351035092
[p9]_error  = 0.0008165923241563178
```

6

Piecewise Blinding *Comparisons*

6.5% Fit Info

Bin Size: 50 keV

Range: [33, 57] MeV

Chi2_Prob = 0.46

Stored Parameters

```
[p10] = 490.6685532617223
[p1] = 0.04240522602748158
[p2] = 0.09136078735108702
[p3] = 43.68193585247294
[p4] = 15.584080643412133
[p5] = -2.082364814201879
[p6] = 0.01206029057698417
[p7] = 0.03853733548225469
[p8] = -750083.4855897021
[p9] = 179.33614202803608
```

Stored Errors

```
[p10]_error = 0.6516690055867969
[p1]_error = 7.2522612743763926e-06
[p2]_error = 9.823004084683062e-06
[p3]_error = 0.0326775275677349
[p4]_error = 3.67050508957257
[p5]_error = 0.0
[p6]_error = 1.1582017141353873e-05
[p7]_error = 5.4019955336021e-06
[p8]_error = 0.0
[p9]_error = 0.021569718263066306
```

100% Fit Info

Bin Size: 200 keV

Range: [34, 56] MeV

Chi2_Prob = 0.20

Stored Parameters

```
[p10] = 490.27553602420863
[p1] = 0.043907581659763495
[p2] = 0.08774562074969872
[p3] = 44.19496553033662
[p4] = 14.09409491546085
[p5] = -2.0608113099430136
[p6] = 0.010018876062215187
[p7] = 0.034984061109657903
[p8] = -750082.8446750462
[p9] = 179.93550550243555
```

Stored Errors

```
[p10]_error = 1.2799090630799697
[p1]_error = 3.831848641555613e-06
[p2]_error = 4.807991830724223e-06
[p3]_error = 0.013652475025935473
[p4]_error = 0.008890706489388833
[p5]_error = 9.813823693516355e-05
[p6]_error = 1.327521268384979e-05
[p7]_error = 2.765154465400599e-06
[p8]_error = 1.249310351035092
[p9]_error = 0.0008165923241563178
```

Comparison Takeaways

1. p4 largest discrepancy [15.5→14.09]

2. p4 error significantly decreased [3.6 → .009]

3. p5, p8 auto fixed in 6.5%, not fixed in 100%

7

Reminder: Global Fit of 2015

Good expansion of fitting methods to independent dataset.
Helps standardize procedures.
- Will summarize progress on 2015 dataset from the summer (pre-roofit)

Restricted Range Fits on 2015
- build RooFit functionality for processing 2015 IMD
- determine if restricted fitting works okay

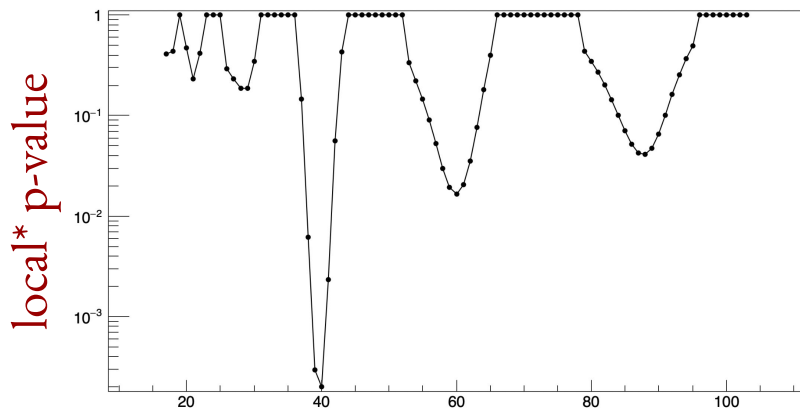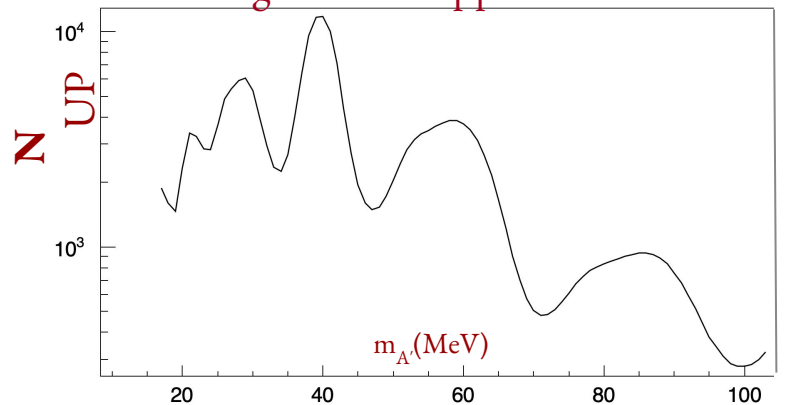**Goal:** generate scaled parameters from 2016 functional form to fit 2015 distribution.

TJ's fits / display from summer
- parameters stored for use in bkg+signal model
- **chi2 probability = 0.81**
- **chi2/dof = 0.97**
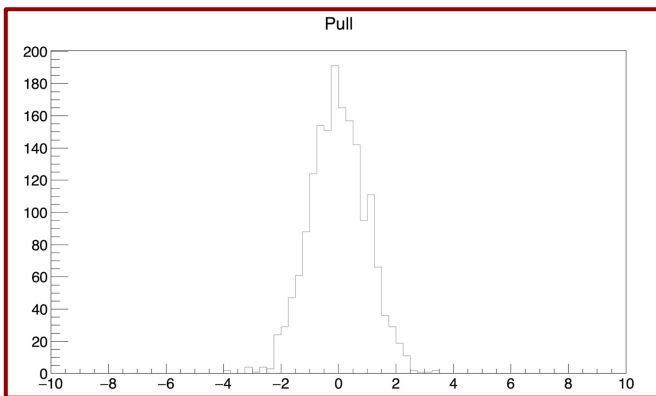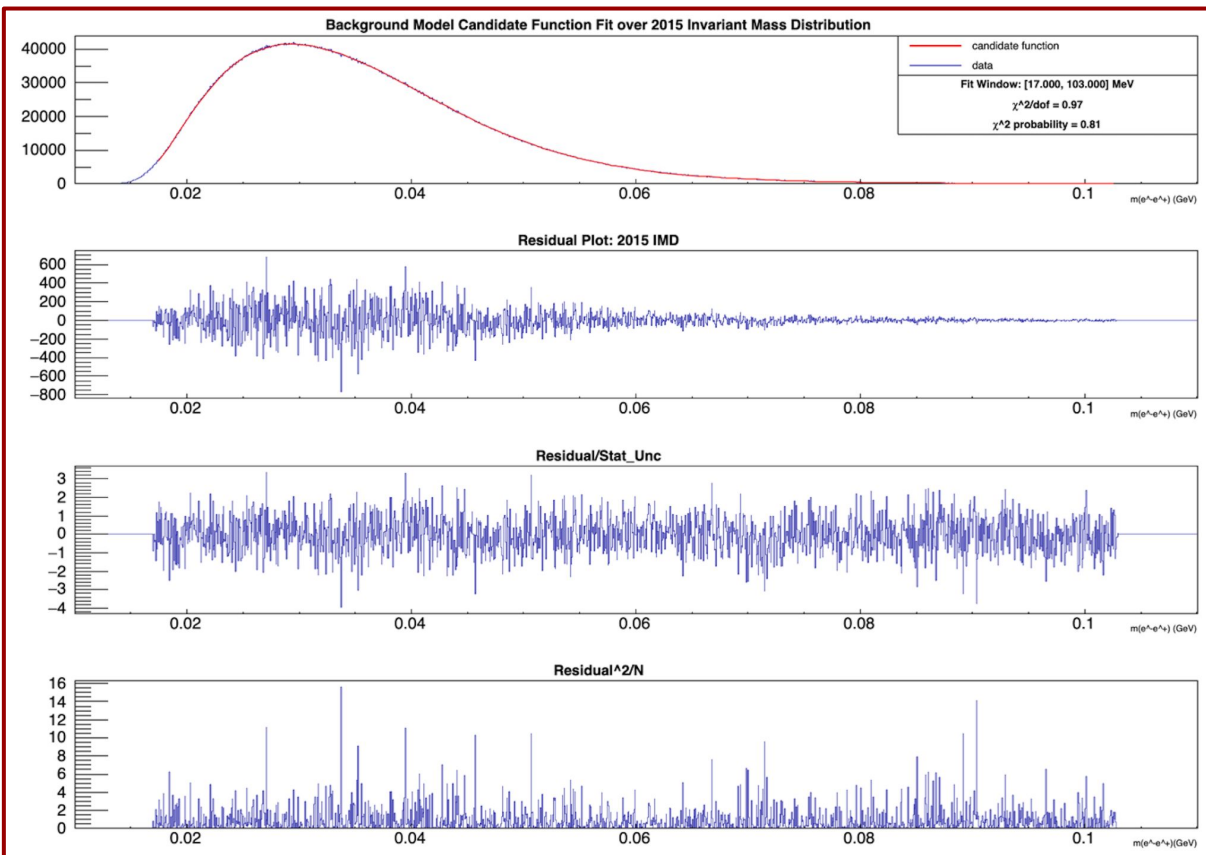
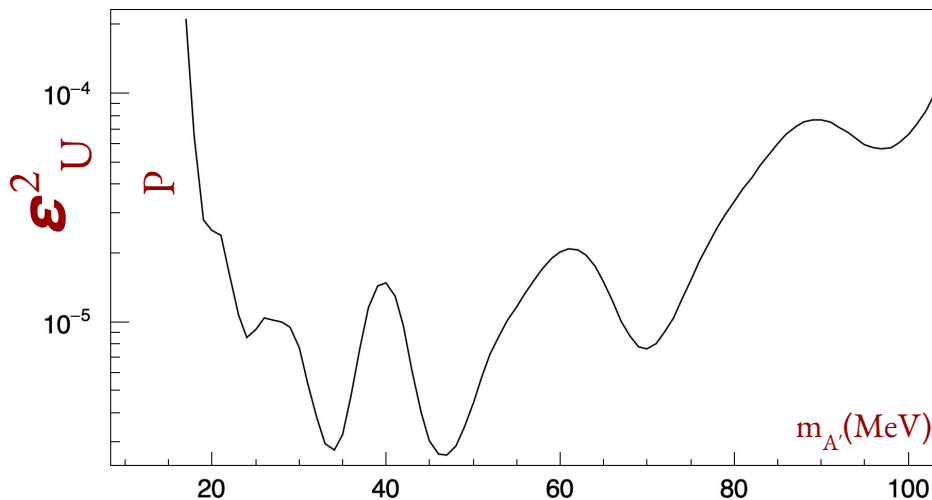# Corresponding 2015 Upper limits and pvalues

Signal Yield Upper Limit

N_{UP}

$m_{A'}(MeV)$

ε^2 Upper Limit

ε^2_{UP}

$m_{A'}(MeV)$

local* p-value

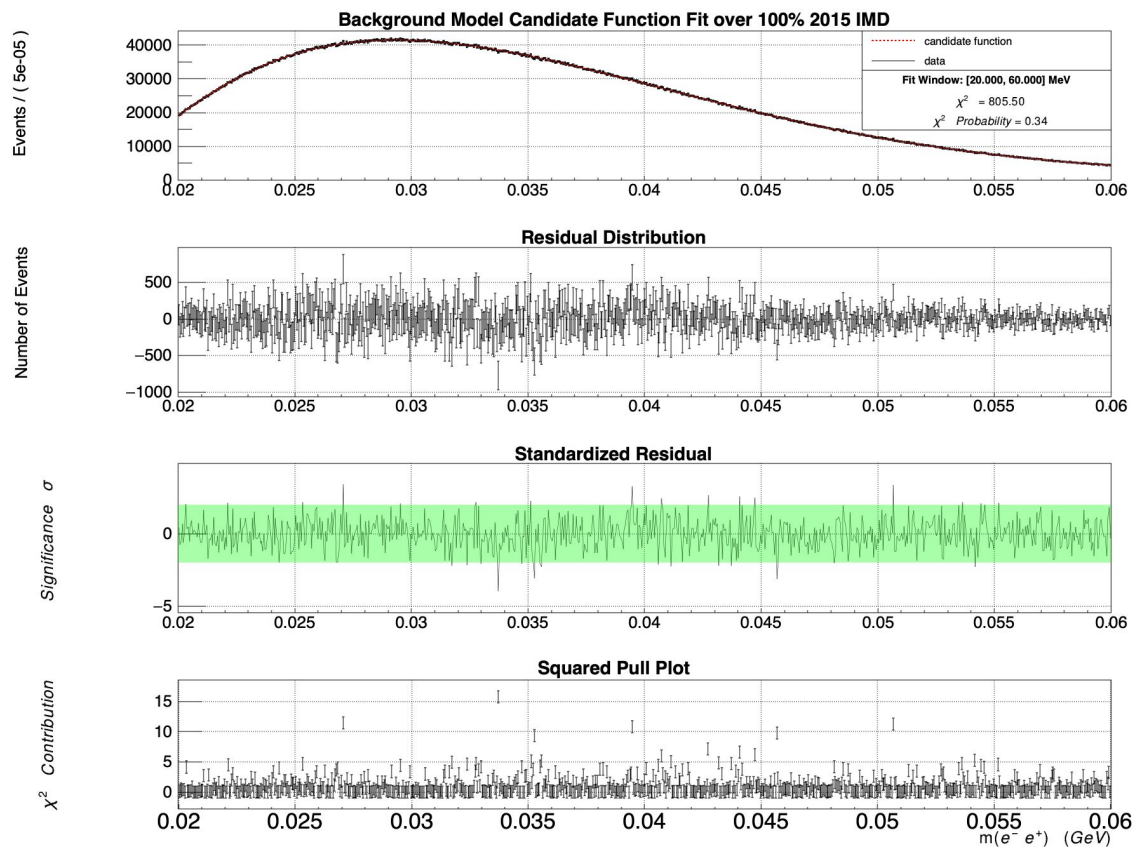*Global pvalues to be calculated when bumphunter rewritten.

Restricted Range Fit of 2015 (1/2)
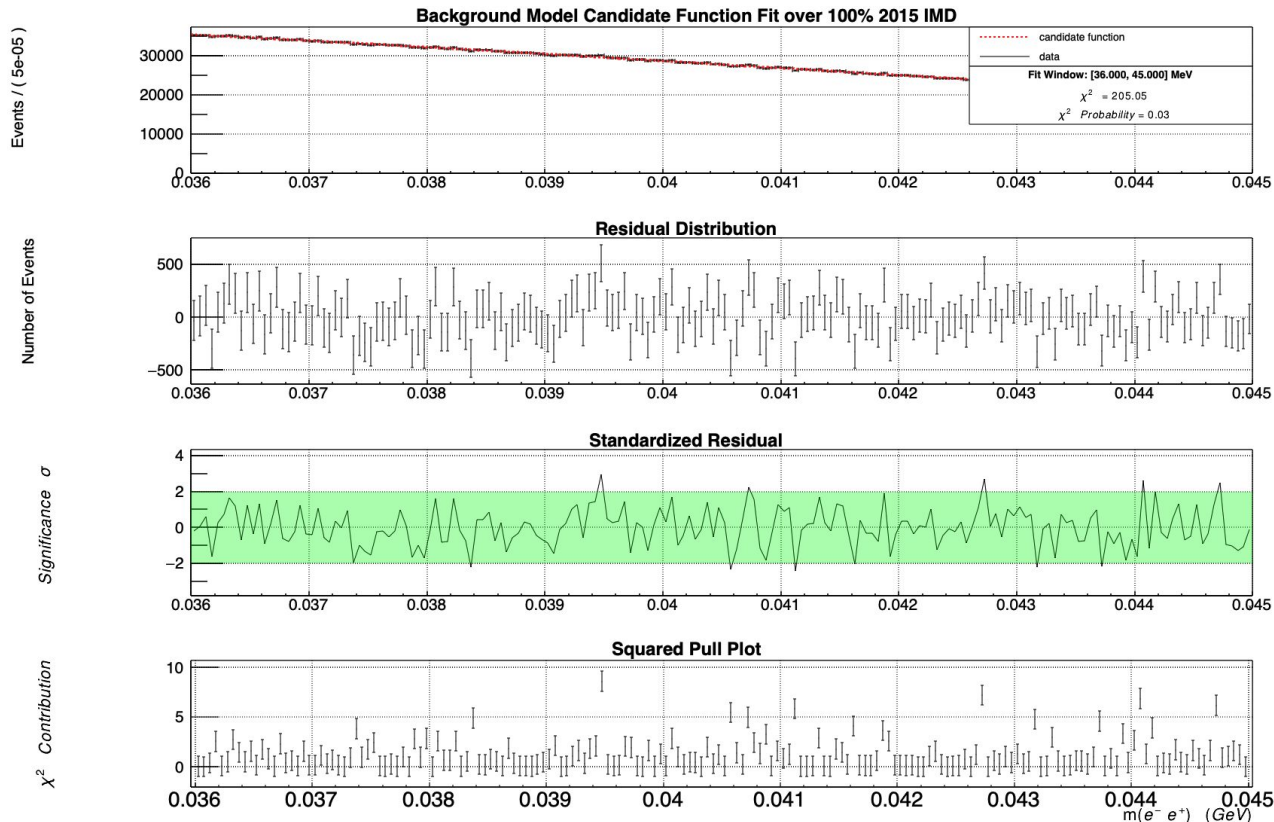

Fit Info

Bin Size: 50 keV

Range: [20, 60] MeV

Chi2_Prob = 0.34

Restricted Range Fit of 2015 (2/2)

Fit Info

Bin Size: 50 keV

Range: [36, 45] MeV

Chi2_Prob = 0.03

Seems more difficult to fit smaller window in 2015.

Next Steps (the pre-analysis workshop meeting grind)

I. Look into proper parameter constraints

    A. correlated gaussians and covariance matrices

    B. Requires fit of 6.5% and 100% in same fitting framework

II. Rewrite BumpHunter to incorporate RooFit expected values

    A. run 6.5% params first and then 100%

II. Signal Injection Studies with improved signal_injectorinator.py

III. Test additional functions

IV. Test feature comparing script for independent background models, different datasets, and different functions.

V. 1% 2021 asap

RooFit Outputs

```
EIGENVALUES OF SECOND-DERIVATIVE MATRIX:
      -1.3349e-03  7.6037e-05  1.2005e-03  2.5386e-03  7.0602e-03  2.7104e-01  5.0418e-01  7.2152e+00
```

```
EXT PARAMETER                  APPROXIMATE     INTERNAL        INTERNAL
NO.   NAME        VALUE           ERROR        STEP SIZE        VALUE
 1  [p10]       3.50185e+02     3.53317e+00    4.19254e-06    -5.62201e-01
 2  [p1]        2.95819e-02     1.90605e-05    3.29714e-07     1.61103e-03
 3  [p2]        8.14971e-02     3.11089e-05    1.86700e-07    -3.42042e-02
 4  [p3]        7.15171e+01     3.20510e-01    2.59714e-06     3.44162e-01
 5  [p4]        3.05917e+01     3.34937e-02    2.02217e-06     1.85304e-01
 6  [p5]       -2.42447e+00        fixed
 7  [p6]        1.08352e-02     3.38710e-05    3.68695e-06    -2.11428e-02
 8  [p7]        3.81382e-02     2.82470e-05    4.00985e-07     6.77836e-02
 9  [p8]       -8.39900e+05        fixed
10  [p9]        1.70768e+02     1.22582e-01    3.51350e-07     5.77961e-02
                              ERR DEF= 0.5
```

```
EXTERNAL ERROR MATRIX.    NDIM=  25     NPAR=  8     ERR DEF=0.5
 1.248e+01  2.400e-05 -1.523e-05  2.572e-01 -1.279e-02  4.646e-05 -6.970e-08  1.711e-01
 2.400e-05  3.633e-10  2.474e-10 -5.783e-07 -1.424e-08  1.281e-10  9.783e-12 -9.063e-08
-1.523e-05  2.474e-10  9.678e-10  1.982e-06  7.042e-07 -4.361e-11  8.553e-11  1.007e-06
 2.572e-01 -5.783e-07  1.982e-06  1.027e-01  2.950e-03 -5.027e-07 -3.089e-06 -6.659e-03
-1.279e-02 -1.424e-08  7.042e-07  2.950e-03  1.122e-03  1.468e-08  2.509e-07  1.056e-03
 4.646e-05  1.281e-10 -4.361e-11 -5.027e-07  1.468e-08  1.147e-09  7.277e-10 -5.774e-07
-6.970e-08  9.783e-12  8.553e-11 -3.089e-06  2.509e-07  7.277e-10  7.979e-10 -5.477e-07
 1.711e-01 -9.063e-08  1.007e-06 -6.659e-03  1.056e-03 -5.774e-07 -5.477e-07  1.503e-02
```

```
Fit 10: Chi2 = 2643.78624600157, Chi2 Probability = 0.31744790335896744
Best fit: Chi2 = 2642.581200034163, Chi2 Probability = 0.32336539263888797
```

```
PARAMETER   CORRELATION COEFFICIENTS
   NO.  GLOBAL      1       2       3       4       5       7       8      10
    1  0.99426   1.000   0.356  -0.139   0.227  -0.108   0.388  -0.001   0.395
    2  0.99347   0.356   1.000   0.417  -0.095  -0.022   0.198   0.018  -0.039
    3  0.99382  -0.139   0.417   1.000   0.199   0.676  -0.041   0.097   0.264
    4  0.99456   0.227  -0.095   0.199   1.000   0.275  -0.046  -0.341  -0.169
    5  0.92859  -0.108  -0.022   0.676   0.275   1.000   0.013   0.265   0.257
    7  0.96403   0.388   0.198  -0.041  -0.046   0.013   1.000   0.761  -0.139
    8  0.99317  -0.001   0.018   0.097  -0.341   0.265   0.761   1.000  -0.158
   10  0.99437   0.395  -0.039   0.264  -0.169   0.257  -0.139  -0.158   1.000
```

Automatic Parameter fixing under investigation.