Current Goal – Smearing Matrix

Outline

- Unfolding algorithm building blocks
- Results for my current unfolding status

Benard Mulilo
RIKEN Radiation Laboratory
Mar. 13, 2019 @ 9 pm JST

Bayesian Unfolding Algorithm - Building Blocks

$$P(\theta|D,M) \propto P(D|\theta,M)P_0(\theta|M)$$

posterior $\propto$ likelihood $\times$ prior

- Default
- User defined

- Create Model
- Read Data

Define model

- Parameters, (λ)
- Likelihood, $p(D|\lambda)$
- Priori, $p(\lambda)$

Read Some Data

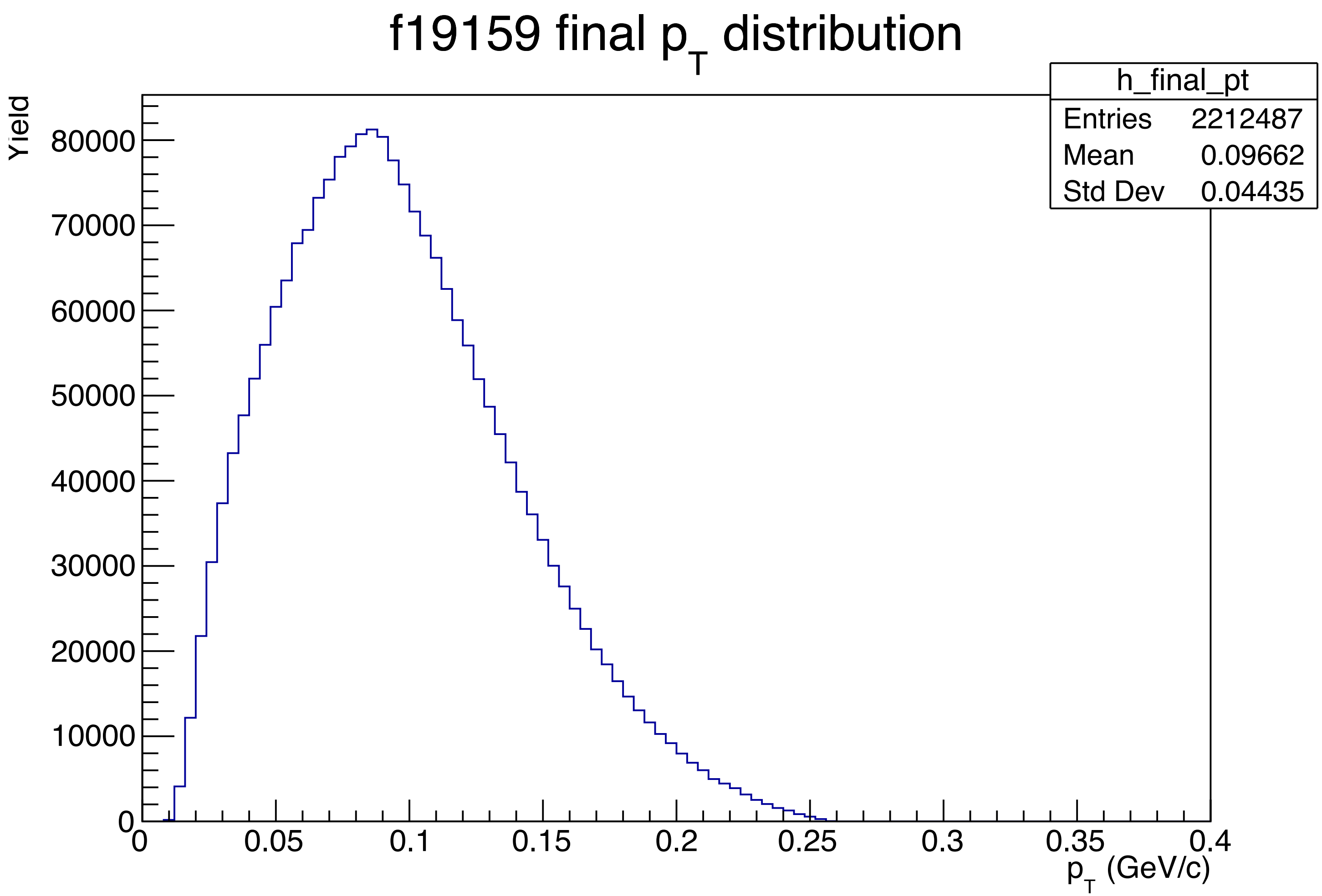
- User defined data, e.g. random generated data.
- Data stored in root tree or text file
- N-tuple data, etc..

Tools

- Find mode
- Fit
- Compare models
- Output

Transverse Momentum Distribution from Data

Need Knowledge data.



Determination of Inverse Smearing Matrix

Procedure:

1. Defining Model:
   Default root models or user defined models.
   Clear knowledge of data is required.

2. Defining Parameters:
   Depend on model whether user defined or default in root.

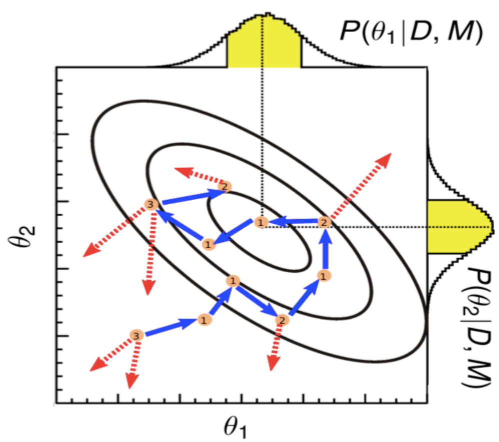
3. Defining Likelihood:
   Heart of model is likelihood function. BAT works with log-likelihood functions. BAT calculates log of Poisson.

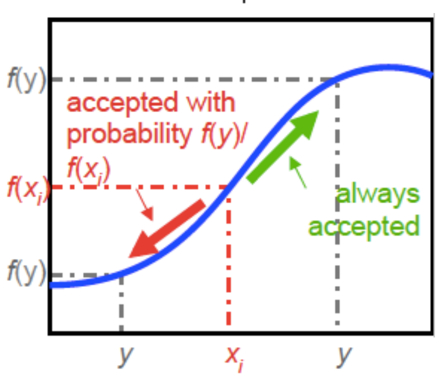
4. Setting prior distributions.

Determination of Inverse Smearing Matrix

Marcov Chain Monte Carlo (MCMC)

- MCMC carries out integration and marginalization of posteriors without requiring manual tuning of parameters. It is implemented on Metropolis Hastings Algorithm.
- Maps positive $f(x)$ function by a random walk towards higher probabilities.





1. Start at some random point, x_i
2. Randomly generate y around x_i
3. If $f(y) \geq f(x_i)$, set $x_i+1 = y$
4. If $f(y) < f(x_i)$, set $x_i+1 = y$ with prob. $f(y)/f(x_i)$
5. If y not accepted, stay where you are, i.e., set $x_i+1 = x$
6. Generate new y , repeat. For infinite steps, distribution converges to $f(x)$.

BAT Program Execution and Outputs

- From transverse momentum, (pT) distribution from experimental data, I know important information: entries, mode and standard deviation.

- So I begin with this information to define a model that fits a Gaussian distribution to my randomly generated data and so has three parameters: the mean, sigma and some scaling factor.

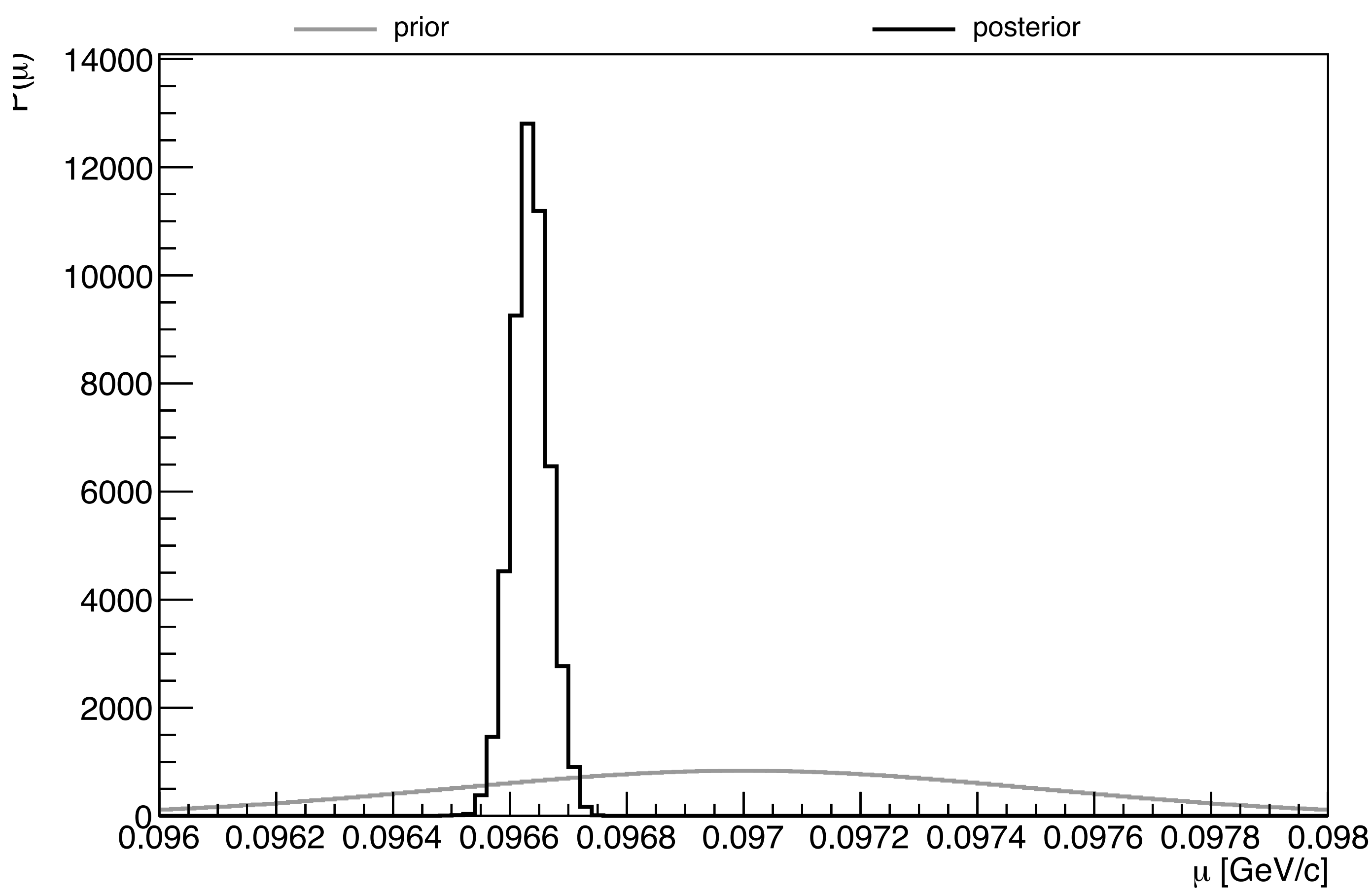
- I then randomly generate data to model the experimental data based on the number of entries from pT spectrum (2 212 487), mean (0.09662) and sigma (0.04435) using a TF1 Gaussian from ROOT and execute my code.

- Finally, I generate results using ROOT after convergence of a set of 4 MCMC chains.

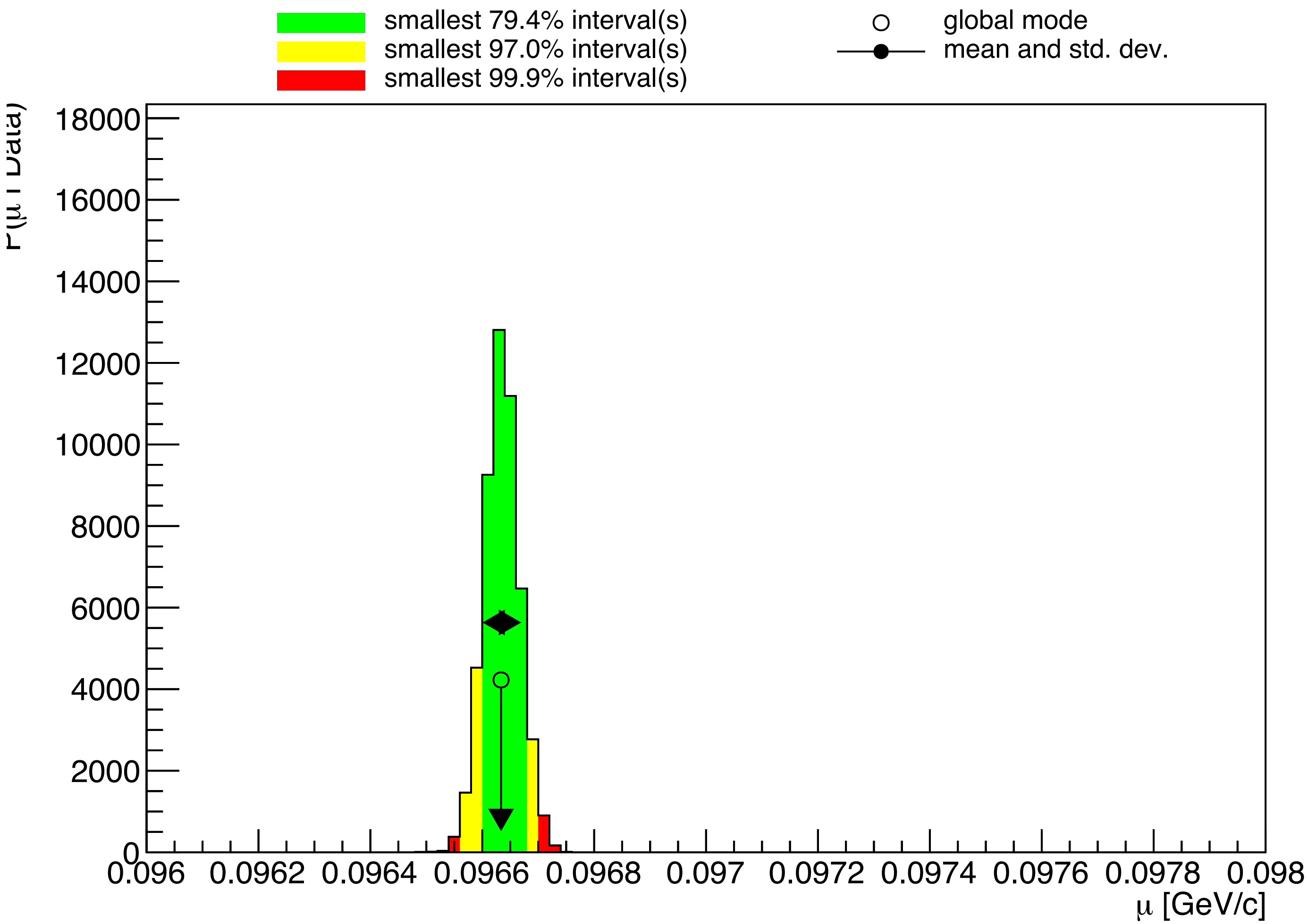
BAT Program Execution and Outputs


```
Summary :  Status of the MCMC
Summary :  ==================
Summary :  Convergence reached:                    yes
Summary :  Number of iterations until convergence: 9500
Summary :  Number of chains:                       4
Summary :  Number of iterations per chain:         100000
Detail  :  Scale factors and efficiencies (measured in last 100000 iterations):
Detail  :  Chain : Scale factor    Efficiency
Detail  :      0 :       0.001916        20.0 %
Detail  :      1 :       0.001916        19.3 %
Detail  :      2 :       0.001916        18.3 %
Detail  :      3 :       0.002875        21.2 %
Summary : Exiting
```

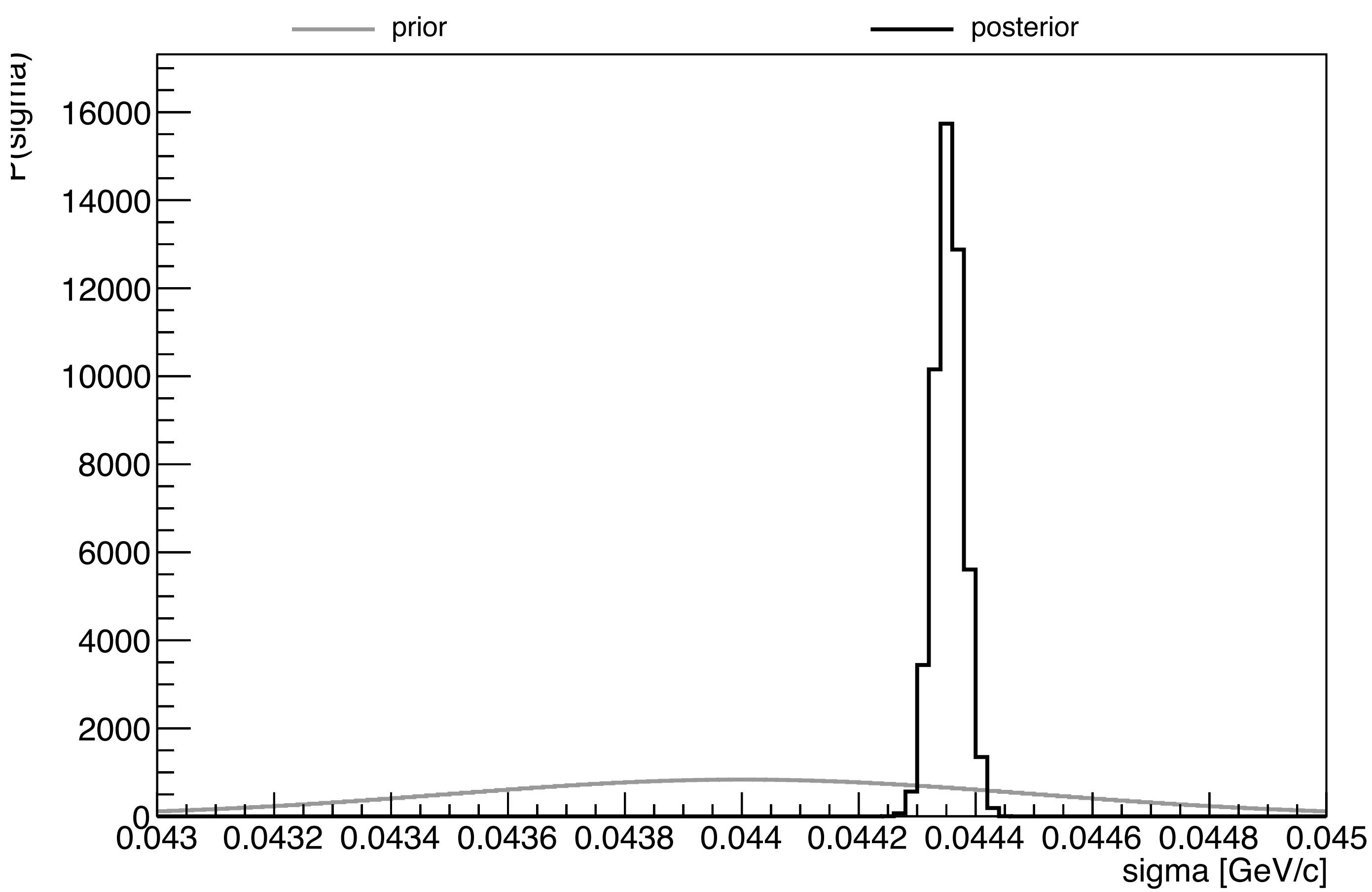
Output – 1D Knowledge Update Plot for Mean Parameter



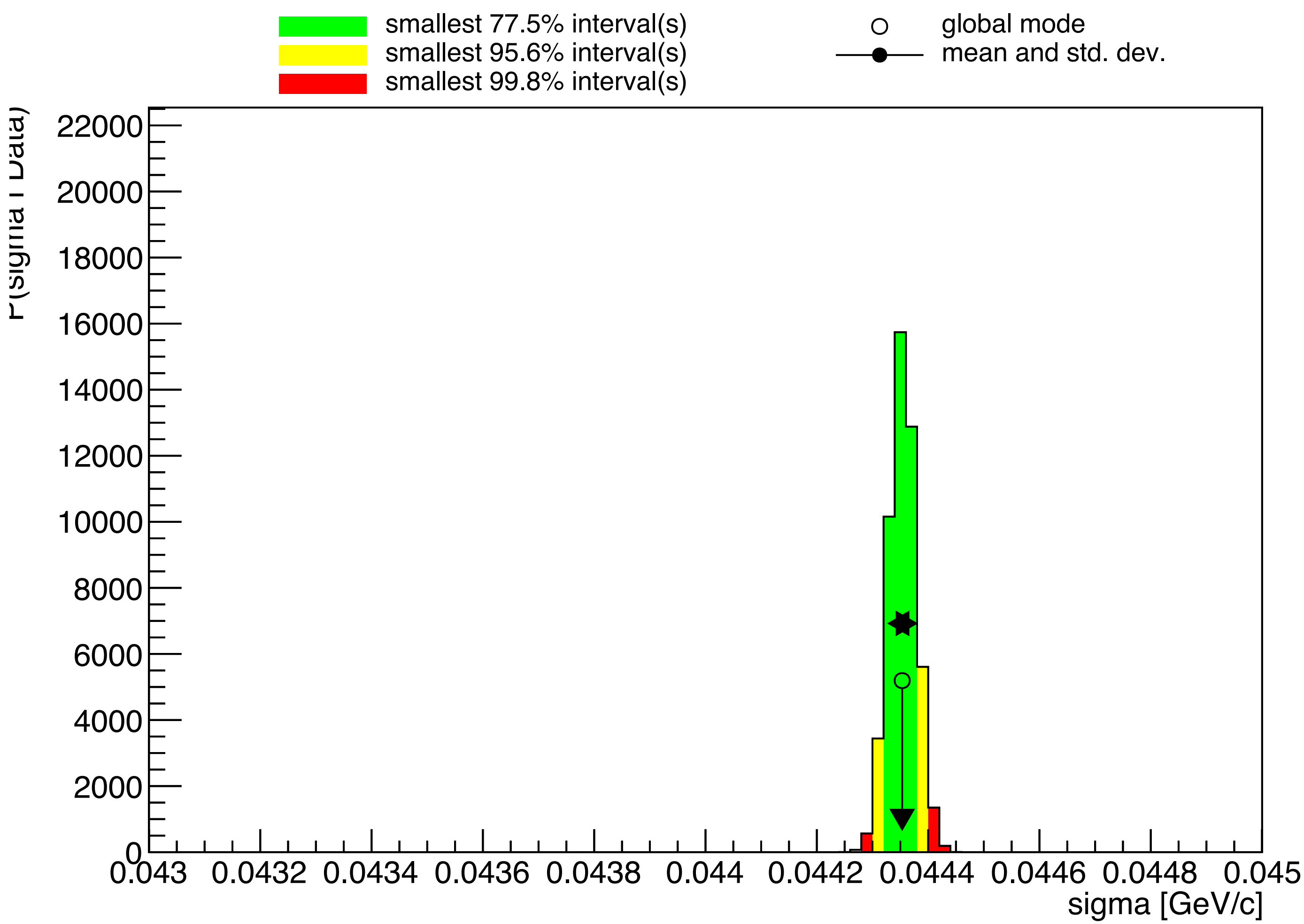
Output – 1D Marginalized Plot for Mean Parameter



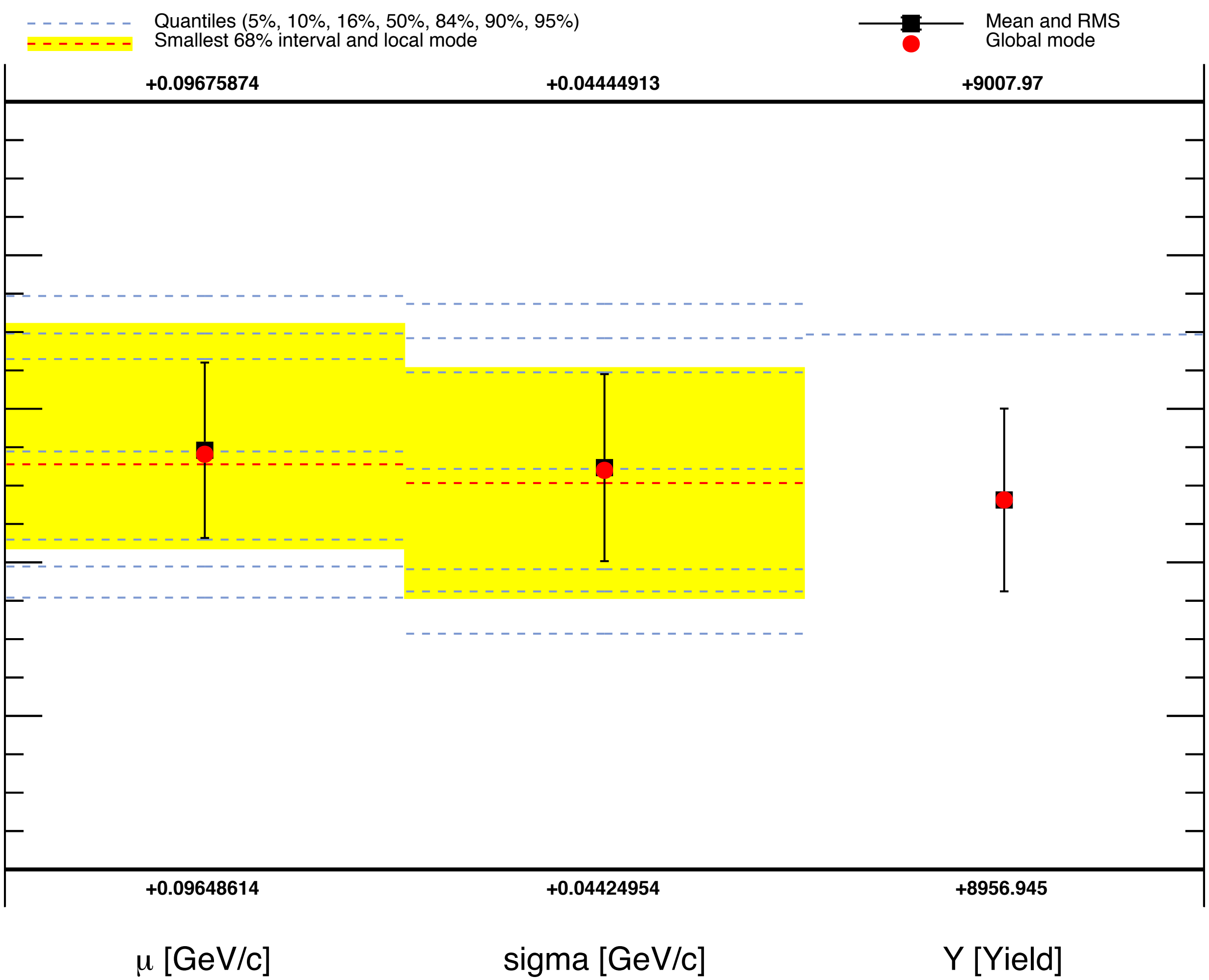
Output – 1D Knowledge Update Plot for Sigma Parameter



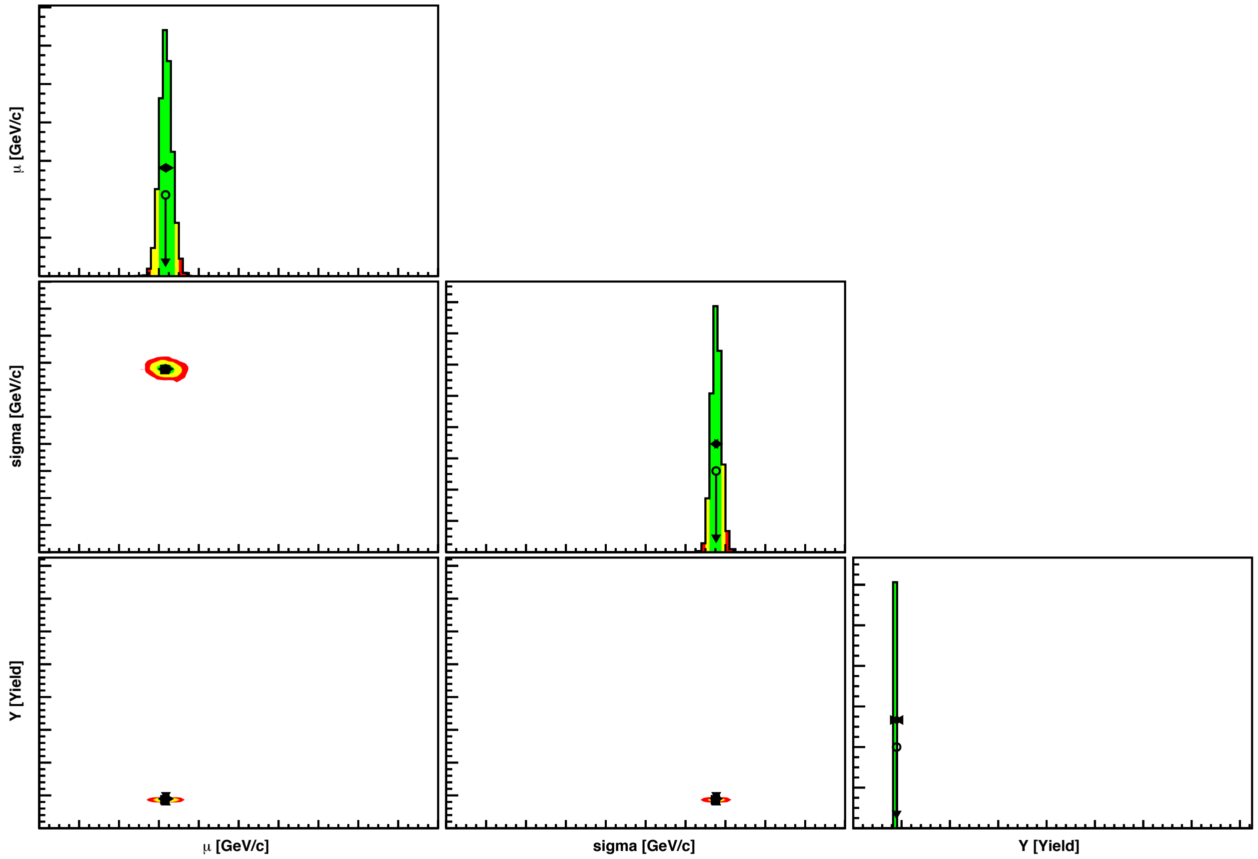
Output – 1D Marginalized Plot for Sigma Parameter



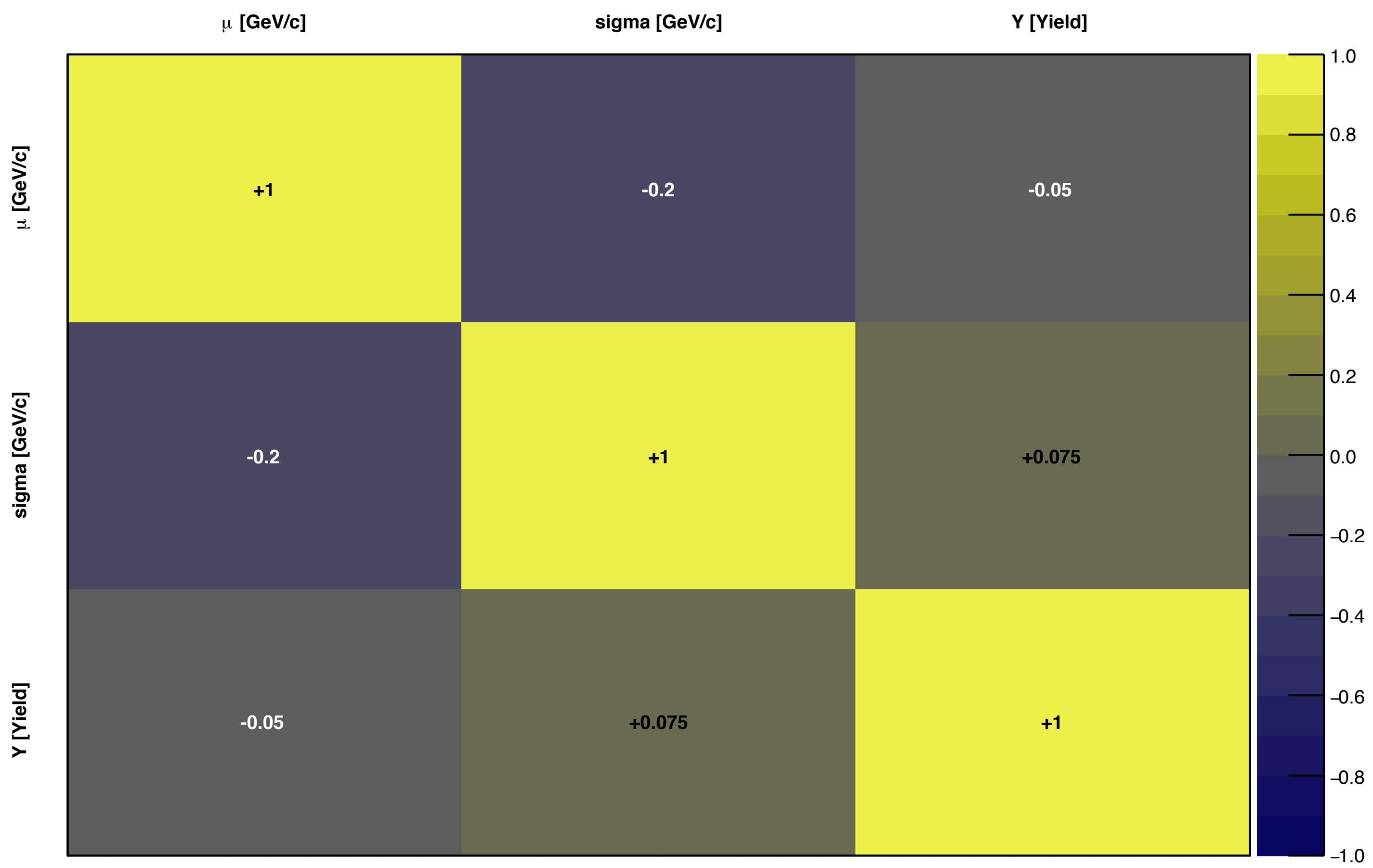
Output – Mean and Sigma Parameters In Same Plot



Output – Parameter Correlation of 1D and 2D Plots



Output – Correlation Matrix For All 3 Parameters



Current Tasks

① Make a user-defined function model that will best fit our experimental pT distribution data using Monte Carlo.

② Apply the model to fit the pT distribution using smeared monte carlo and extract the inverse smearing matrix that will reproduce step 1.

③ Apply the smearing matrix to pT experimental data for actual computation of asymmetries.

Summary of BAT Useful Output


```
Results of the marginalization
==============================
List of variables and properties of the marginalized distributions:

 (0) Parameter "mu" :
     Mean +- sqrt(Variance):         +0.09663498 +- 3.125486e-05
     Median +- central 68% interval: +0.09663455 + 3.262283e-05 - 3.11483e-05
     (Marginalized) mode:            +0.09663
      5% quantile:                   +0.09658266
     10% quantile:                   +0.0965937
     16% quantile:                   +0.0966034
     84% quantile:                   +0.09666717
     90% quantile:                   +0.09667645
     95% quantile:                   +0.09668976
     Smallest interval containing 79.4% and local mode:
     (0.0966, 0.09668) (local mode at 0.09663 with rel. height 1; rel. area 1)

 (1) Parameter "Sigma" :
     Mean +- sqrt(Variance):         +0.04435401 +- 2.4358e-05
     Median +- central 68% interval: +0.04435368 + 2.499739e-05 - 2.5956e-05
     (Marginalized) mode:            +0.04435
      5% quantile:                   +0.04431083
     10% quantile:                   +0.04432182
     16% quantile:                   +0.04432773
     84% quantile:                   +0.04437868
     90% quantile:                   +0.04438767
     95% quantile:                   +0.04439658
     Smallest interval containing 77.5% and local mode:
     (0.04432, 0.04438) (local mode at 0.04435 with rel. height 1; rel. area 1)

 (2) Parameter "Y" :
     Mean +- sqrt(Variance):         +8981.502 +- 6.074507
     Median +- central 68% interval: +8662.5 + 280.5 - 280.5
     (Marginalized) mode:            +8662.5
      5% quantile:                   +8291.25
     10% quantile:                   +8332.5
     16% quantile:                   +8382
     84% quantile:                   +8943
     90% quantile:                   +8992.5
     95% quantile:                   +9033.75
     Smallest interval containing 100.0% and local mode:
     (8250, 9075) (local mode at 8662.5 with rel. height 1; rel. area 1)
```

[Back up]
BAT Functionality

2. Working of BAT

③ Log Output

- Detailed information on BAT running can be output to the screen or file using static class called BCLog. File or screen output of information is specified at the beginning.

- BC::LogLevel has the following values:

  (i) debug: (Lowest output level)
  (ii) detail: (Functions and status)
  (iii) summary: (Results, including best fit values and normalization)
  (iv) warning: (Warning messages) and
  (vi) nothing: (No output)

[Back up]
BAT Functionality

3. Working of BAT

④ Miscellaneous

- Through CERN's ROOT classes, it handles storage of marginalized distributions and results of MCMC used to generate them.

- BCH1D and BCH2D have several methods for obtaining useful information from marginalized distributions such as:
  (i) Access mean.
  (ii) Access median.
  (iii) Access mode.
  (iv) Limits at given credibility levels

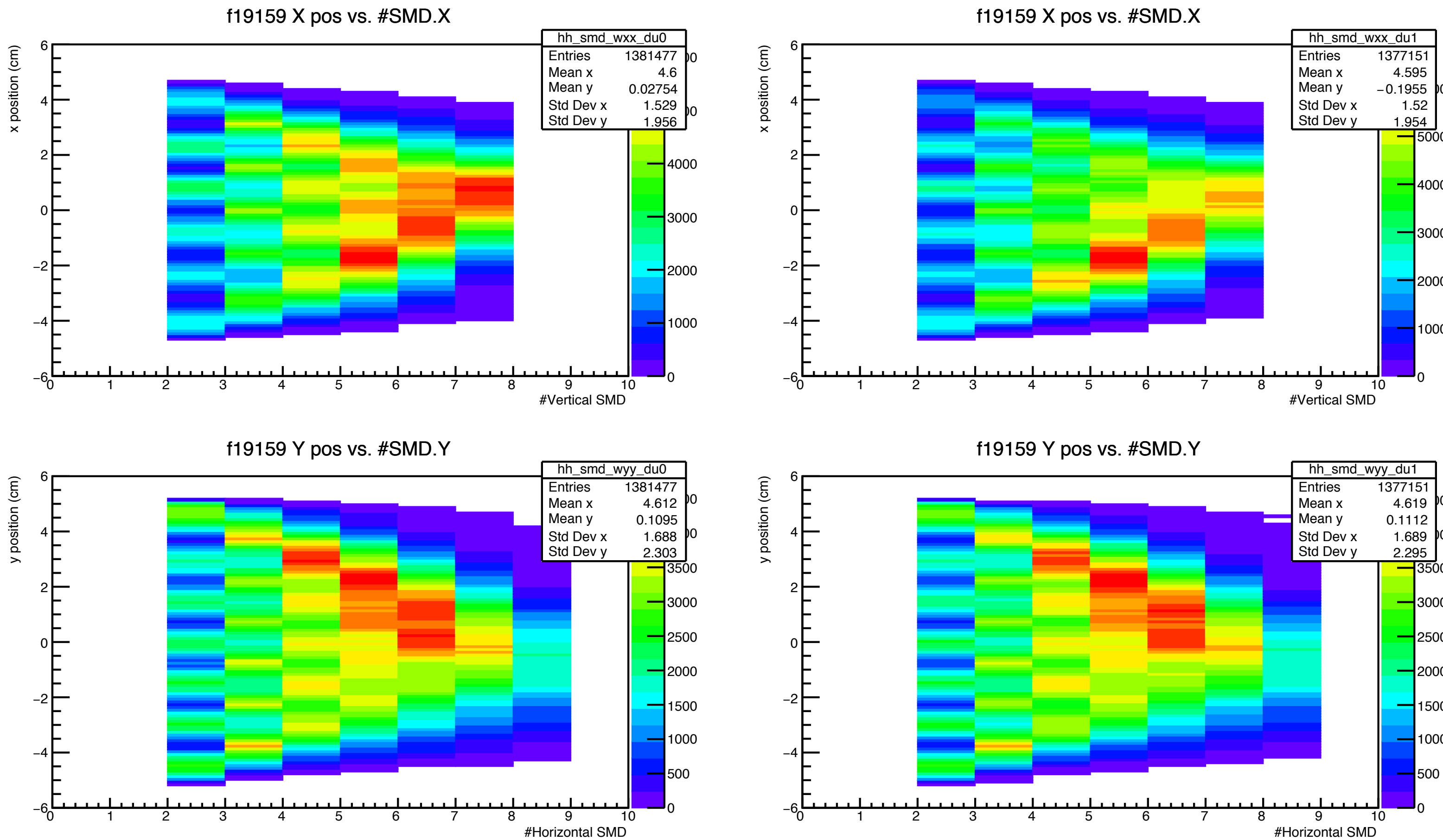
<u>To Do List</u>

- Optimize the setup.

- Start unfolding of x_F and p_T distributions

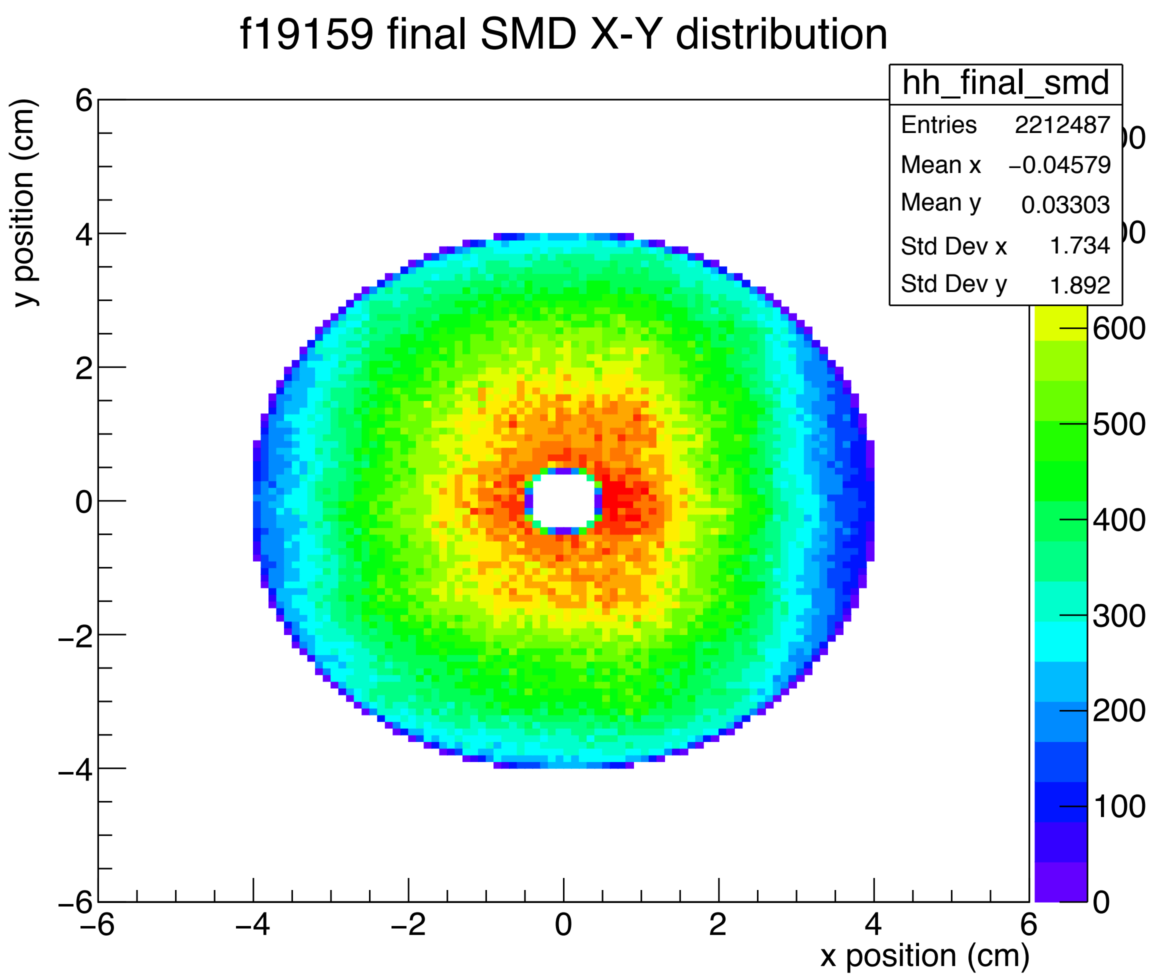
[Back up]

SMD Data – Position Reconstruction



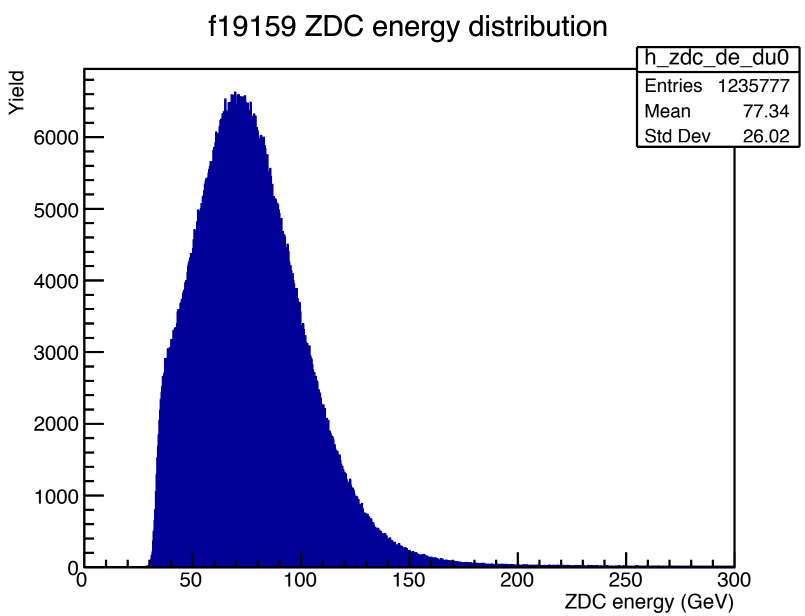
X and Y neutron position reconstruction using the shower maximum detector (SMD)

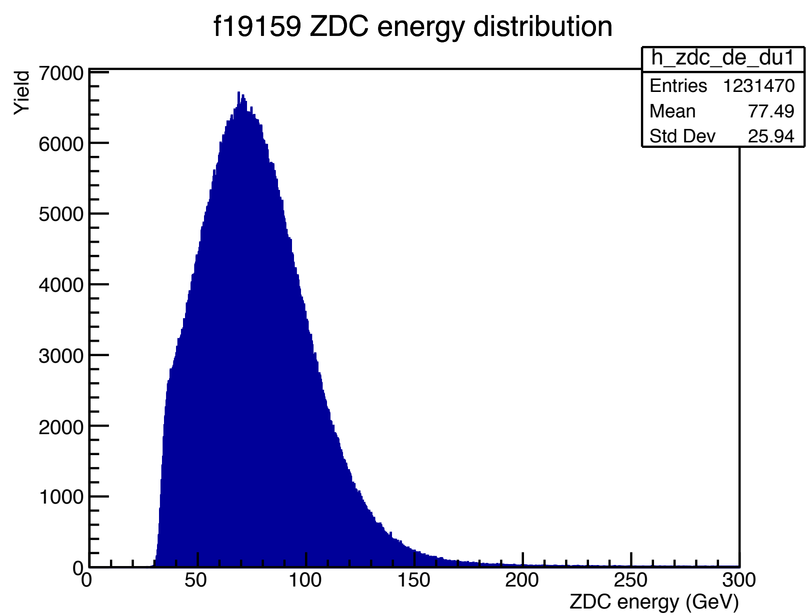
[Back up]
SMD Data: Final X-Y Distribution



[Back up]
ZDC Energy for Spin down and Spin up

Slide 5





Spin Arrow Downward

Spin Arrow Upward

[Back up]
ZDC Energy – Final Energy Distribution

Slide 6