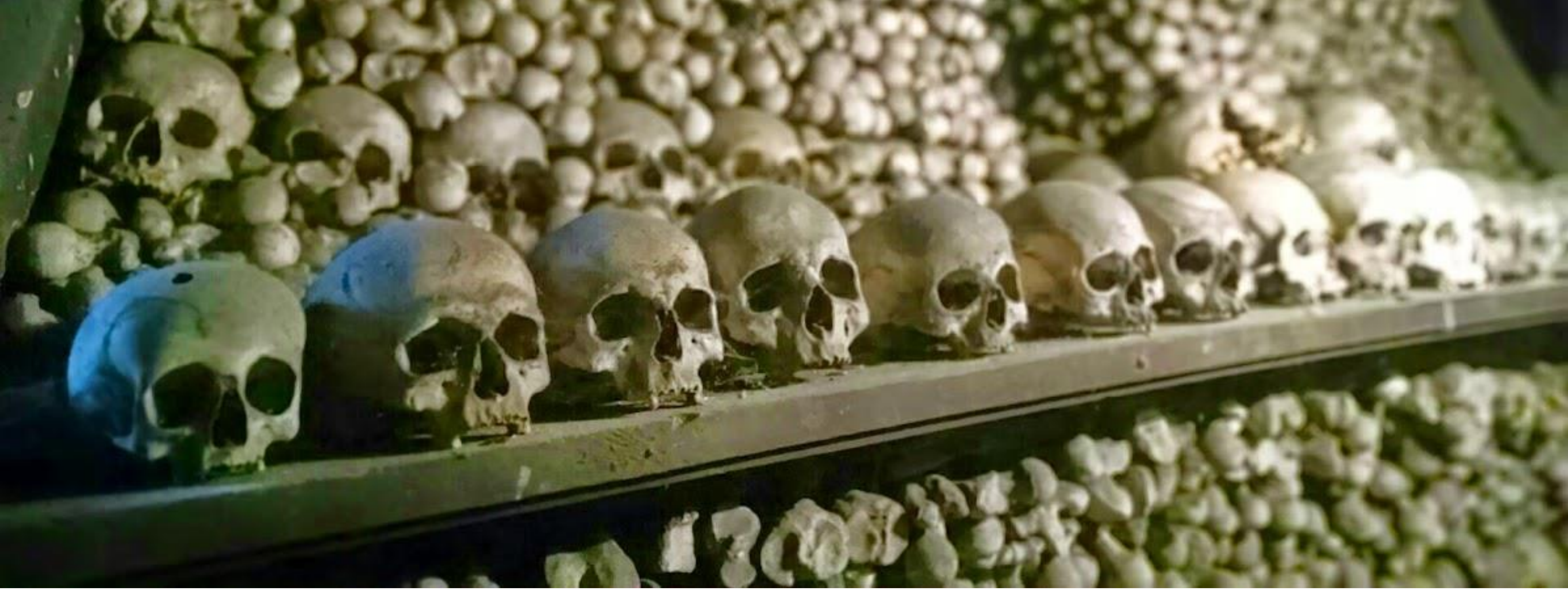


PIDRIX: PARTICLE IDENTIFICATION MATRIX FACTORIZATION

Evan Sangaline
Michigan State University
ACAT 2014

TRADITIONAL PARTICLE IDENTIFICATION

Measure things that carry information about particle identity (dE/dx, 1/β, invariant mass, Čerenkov radiation, calorimeter energy, etc)

Painstakingly model these things for each particle type (crystal ball, normal distribution, Student's t-distribution, etc)

Maximize the likelihood to determine our best estimates of shapes and yields

This step is really hard and error prone.

Can we just skip it?

PIDRIX

Measure things that carry information about particle identity (dE/dx, 1/β, invariant mass, Čerenkov radiation, calorimeter energy, etc)

↓

Maximize the likelihood to determine our best estimates of shapes and yields

Yes!

THE ASSUMPTIONS

- We have a number of different particles that we measure
  - π, K, p, e, etc but could also be “background”
- For a given particle type each dimension of measurement is uncorrelated
  - e.g. ToF measurement error is not correlated with dE/dx measurement error
  - Bin on momentum and pseudorapidity to remove physics correlations
- Our particle yields follow Poisson statistics
  - No yields less than zero

TOWARDS SOMETHING MORE MATHEMATICAL

This means that we expect the observed density to be of the form

$$A_{total}(x, y) = \sum_{i}^{r} v_i(x) * u_i(y)$$ where $v_i(x)$ and $u_i(y)$ are positive semidefinite and real

Get your bearings:

x, y are the measurement variables (dE/dx and 1/ β for instance)

$A_{total}(x, y)$ is a fit function

$v_i(x) * u_i(y)$ is the contribution to the fit function for the i^{th} particle

r is the number of distinct particle types

FURTHER...

We use histograms so let's discretize things

$$A_{total}(x,y) = \sum_{i}^{r} v_i(x) * u_i(y)$$ becomes $$A_{m_x n} = U_{m_x r} V_{r_x n}$$

Now say our observation histogram is T,
then maximizing the log-likelihood is equivalent to
minimizing the generalized Kullback-Leibler divergence:

$$D_{GKL}(T||A) = \sum_{i,j} T_{ij} ln \frac{T_{ij}}{A_{ij}} - T_{ij} + A_{ij}$$

NOW WHAT?

Traditional Fitting

Restrict the space of possible U and V matrices according to our models before minimizing.

e.g. “The columns of U must be Gaussians and the rows of V must be Student’s t-distributions.”

Pidrix

Don’t model anything. Just minimize the KL divergence.

i.e. Skip the hard part.

MINIMIZING THE KULLBACK-LEIBLER DIVERGENCE

These update rules can quickly be seen to be stable when T=A.

$$U_{ip} \leftarrow U_{ip} \frac{\sum_\alpha V_{p\alpha} T_{i\alpha} / A_{i\alpha}}{\sum_\alpha V_{p\alpha}}$$

Also, if U and V start out positive definite and real then they will also remain so (our non-negative yield constraint).

$$V_{pj} \leftarrow V_{pj} \frac{\sum_\alpha U_{\alpha p} T_{\alpha j} / A_{\alpha j}}{\sum_\alpha U_{\alpha p}}$$

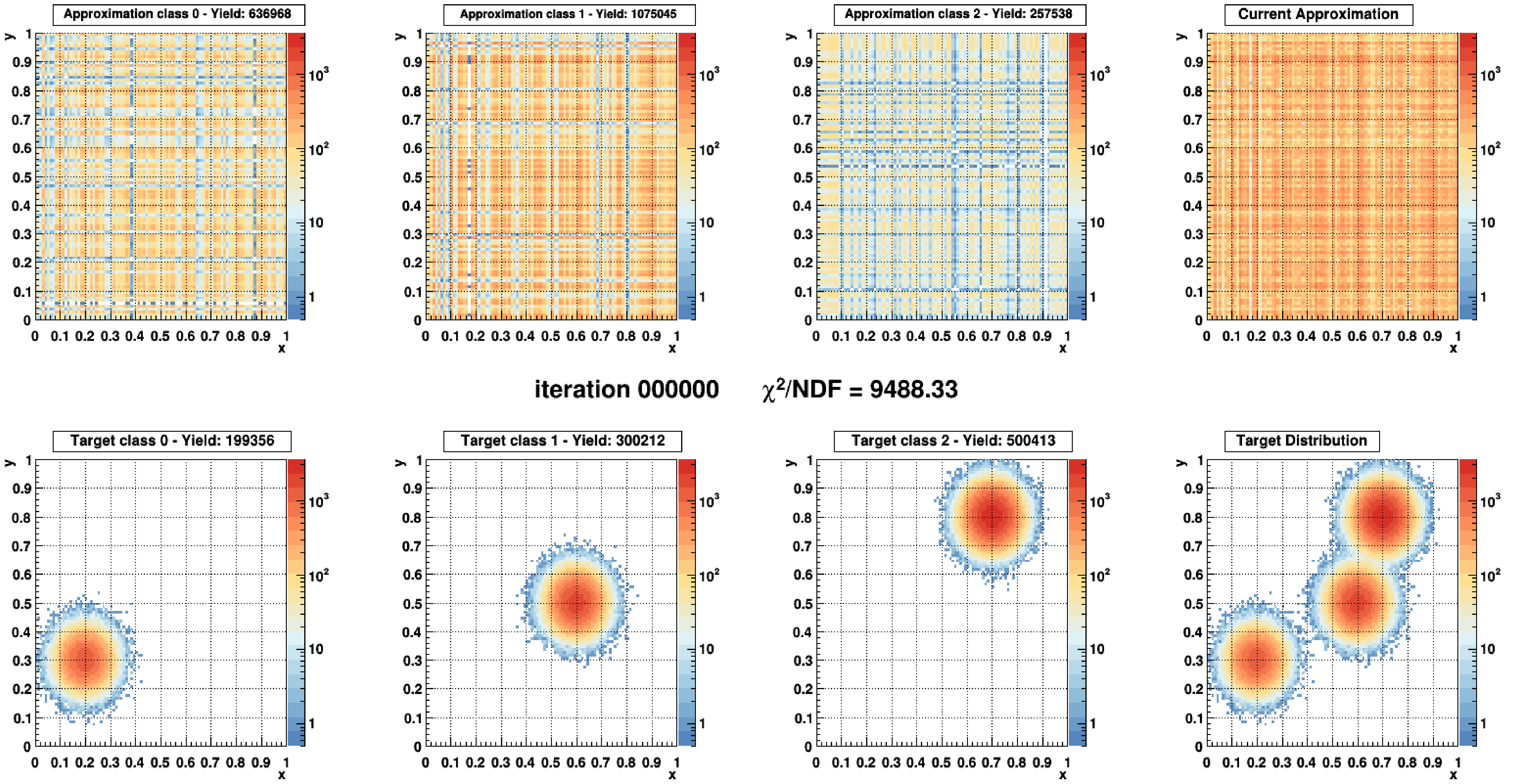
HOW DO THEY WORK?

$$U_{ip} \leftarrow U_{ip} \frac{\sum_\alpha V_{p\alpha} T_{i\alpha}/A_{i\alpha}}{\sum_\alpha V_{p\alpha}}$$

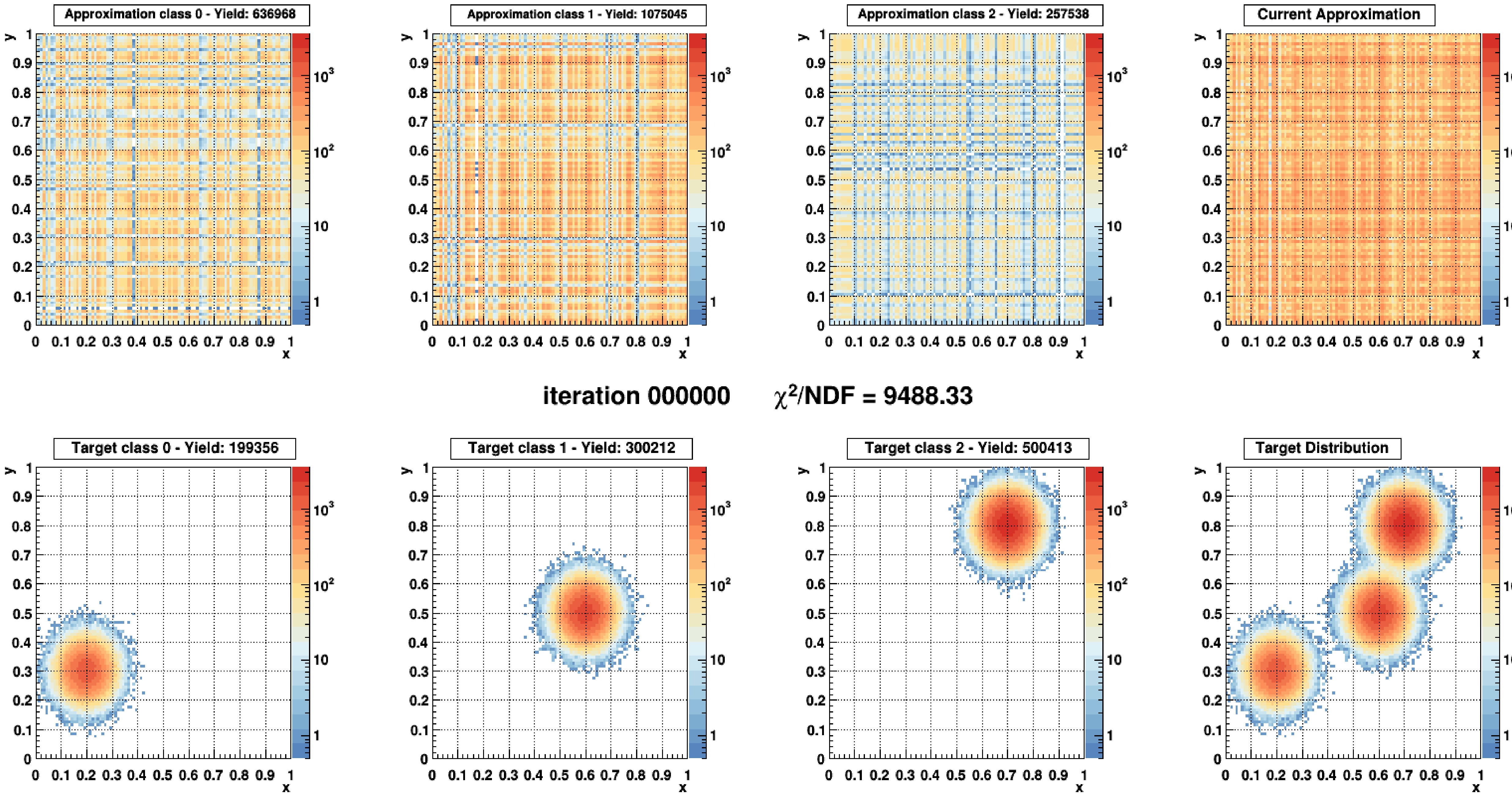
- First: think of T/A (element-wise) like a residual
  - If an entry is less than one then we are underestimating that bin with our approximation, greater than one and we're overestimating
- For a given particle p (proton) we look at a specific i (dE/dx)
- Then we take our current normalized ToF distribution (the p^{th} column of V) corresponding to protons and dot that with the row of T/A that corresponds to our current i value (dE/dx value)
- If this particle's ToF distribution has lots of yield where we are underestimating (overestimating) T then we'll increase (decrease) the dE/dx distribution for this particle at our current i

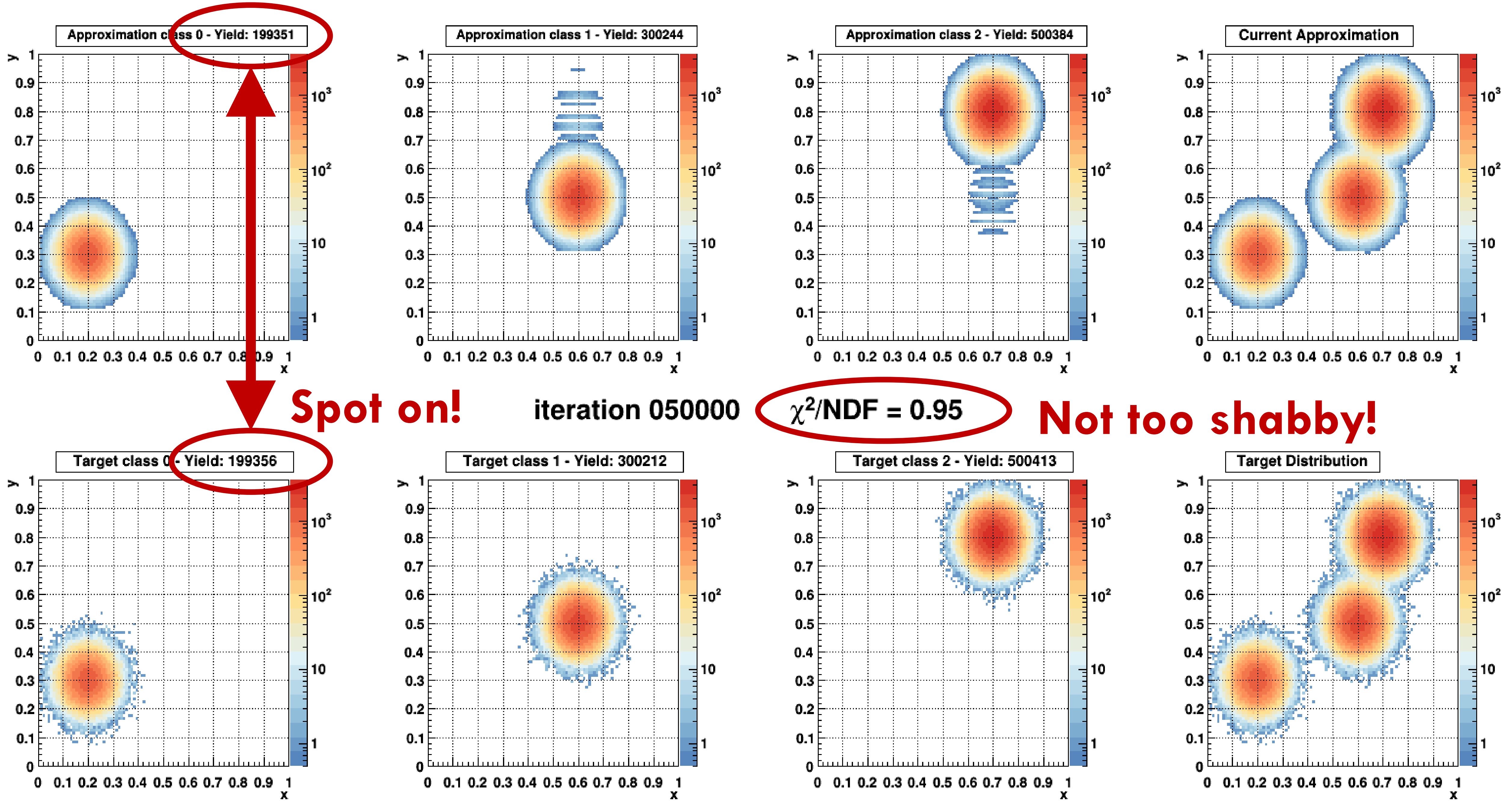
NOW LET'S SEE IT IN ACTION

We'll start with random distributions and see how well they converge…



Approximation class 0 - Yield: 636968
Approximation class 1 - Yield: 1075045
Approximation class 2 - Yield: 257538
Current Approximation
iteration 000000
χ²/NDF = 9488.33
Target class 0 - Yield: 199356
Target class 1 - Yield: 300212
Target class 2 - Yield: 500413
Target Distribution

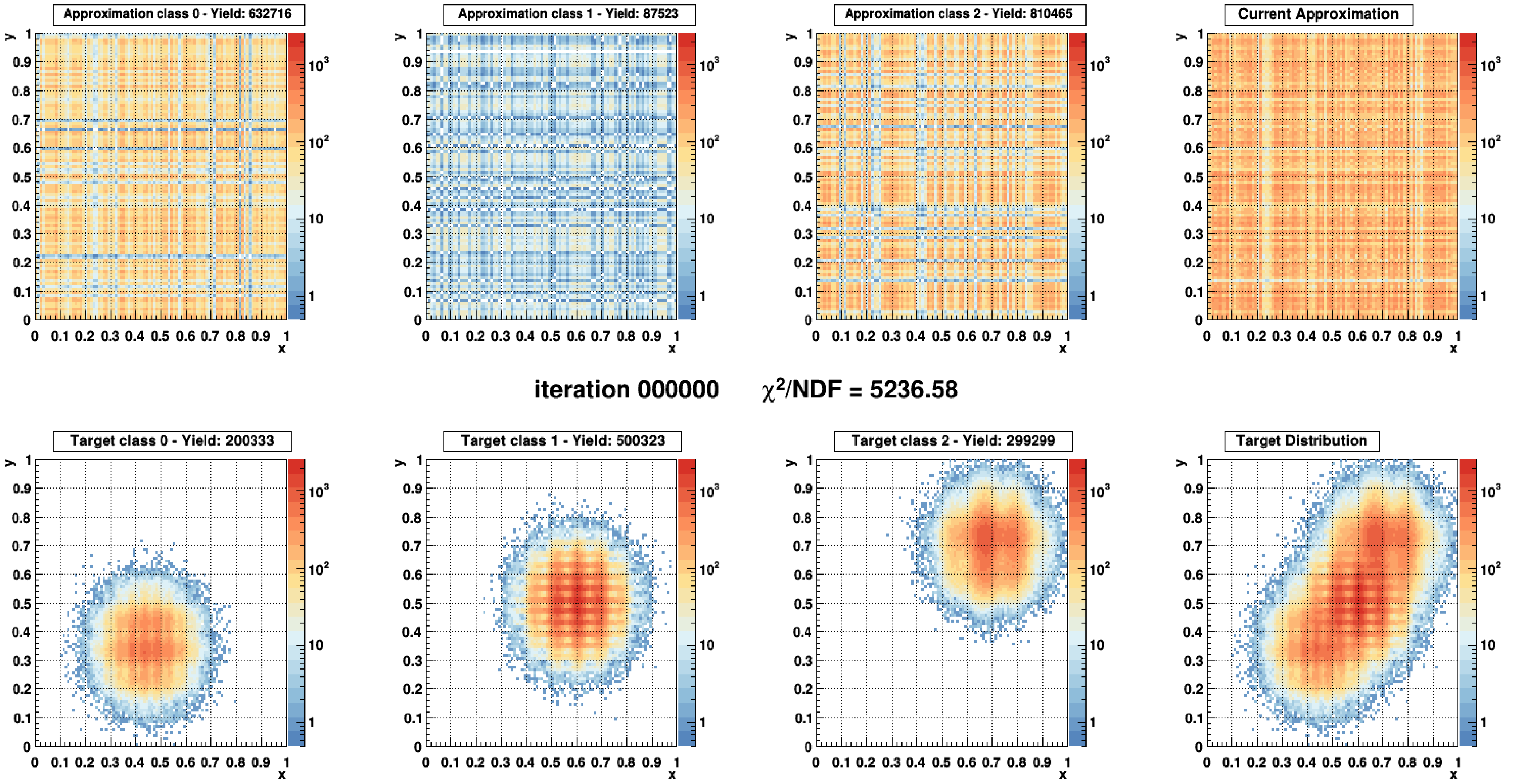

Approximation class 0 - Yield: 636968
Approximation class 1 - Yield: 1075045
Approximation class 2 - Yield: 257538
Current Approximation
iteration 000000
χ^2/NDF = 9488.33
Target class 0 - Yield: 199356
Target class 1 - Yield: 300212
Target class 2 - Yield: 500413
Target Distribution



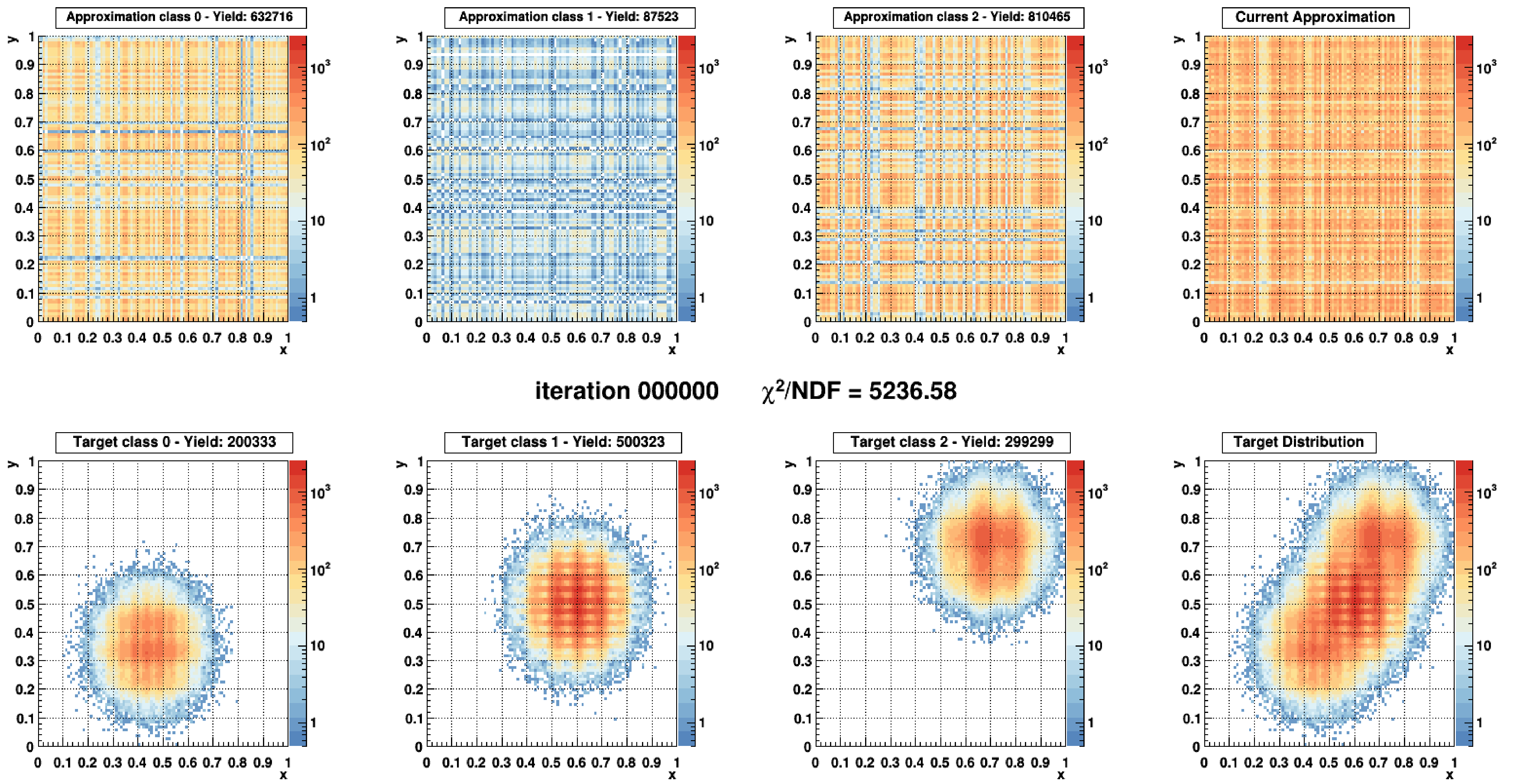
Approximation class 0 - Yield: 199351
Approximation class 1 - Yield: 300244
Approximation class 2 - Yield: 500384
Current Approximation
Spot on!
iteration 050000
χ^2/NDF = 0.95
Not too shabby!
Target class 0 - Yield: 199356
Target class 1 - Yield: 300212
Target class 2 - Yield: 500413
Target Distribution

THAT ONE WAS TOO EASY!

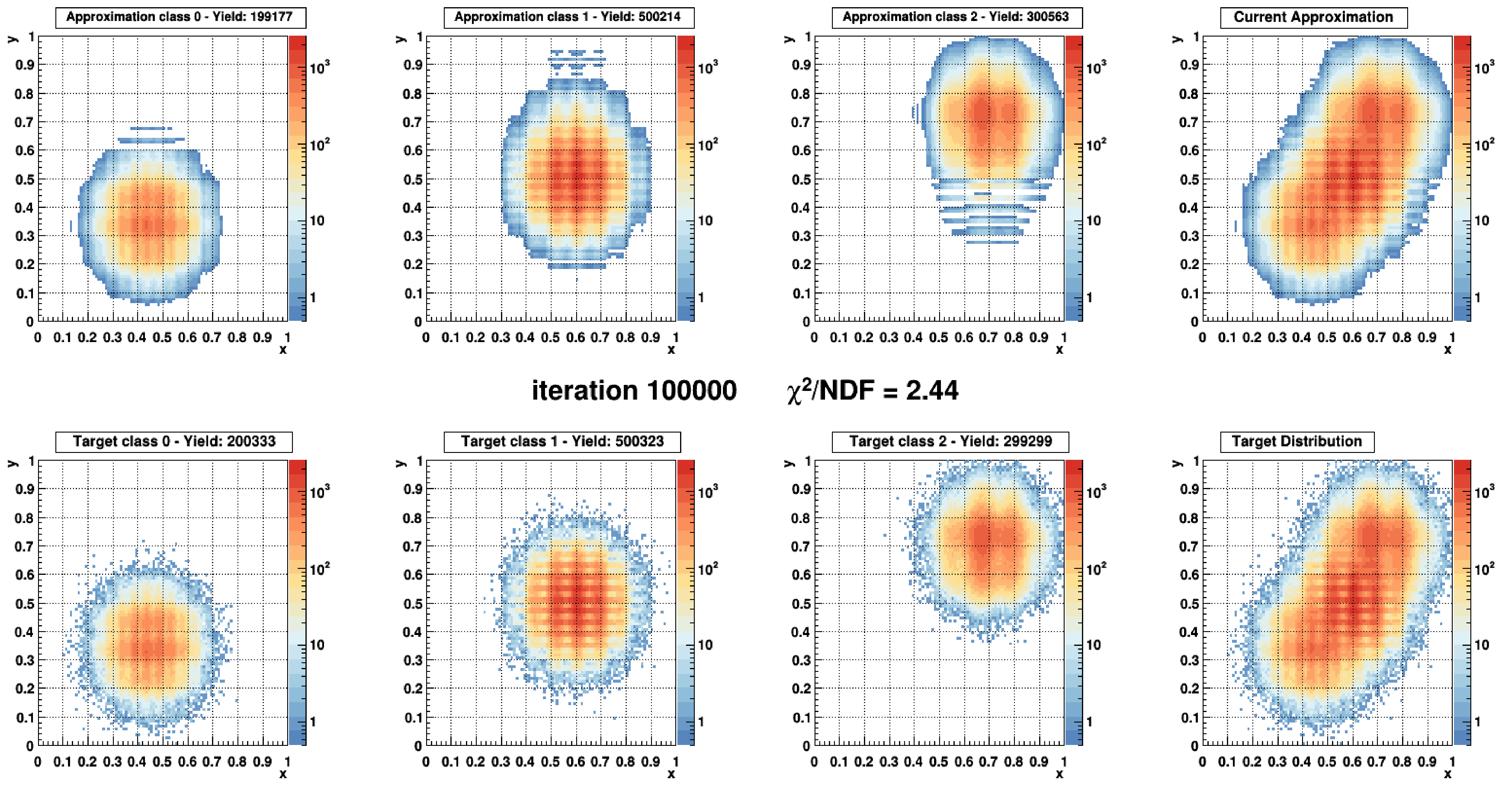
Modeling separated Gaussians is trivial,
what about more complicated distributions?



Approximation class 0 - Yield: 632716
Approximation class 1 - Yield: 87523
Approximation class 2 - Yield: 810465
Current Approximation
iteration 000000
χ^2/NDF = 5236.58
Target class 0 - Yield: 200333
Target class 1 - Yield: 500323
Target class 2 - Yield: 299299
Target Distribution
x
y



Approximation class 0 - Yield: 632716
Approximation class 1 - Yield: 87523
Approximation class 2 - Yield: 810465
Current Approximation
iteration 000000
χ^2/NDF = 5236.58
Target class 0 - Yield: 200333
Target class 1 - Yield: 500323
Target class 2 - Yield: 299299
Target Distribution
x
y

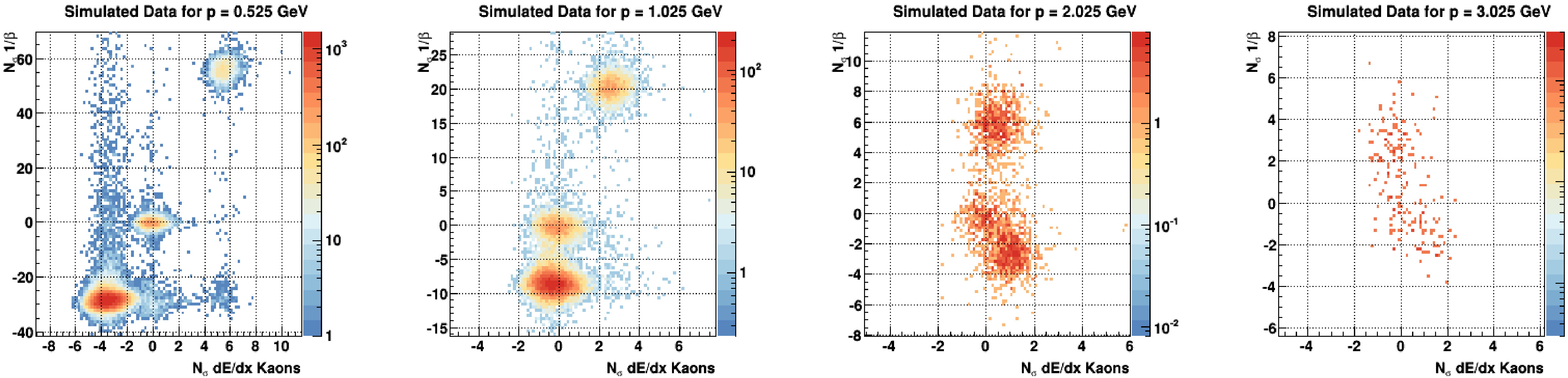

Approximation class 0 - Yield: 199177
Approximation class 1 - Yield: 500214
Approximation class 2 - Yield: 300563
Current Approximation
iteration 100000
χ^2/NDF = 2.44
Target class 0 - Yield: 200333
Target class 1 - Yield: 500323
Target class 2 - Yield: 299299
Target Distribution
x
y

BETTER… BUT STILL UNREALISTIC

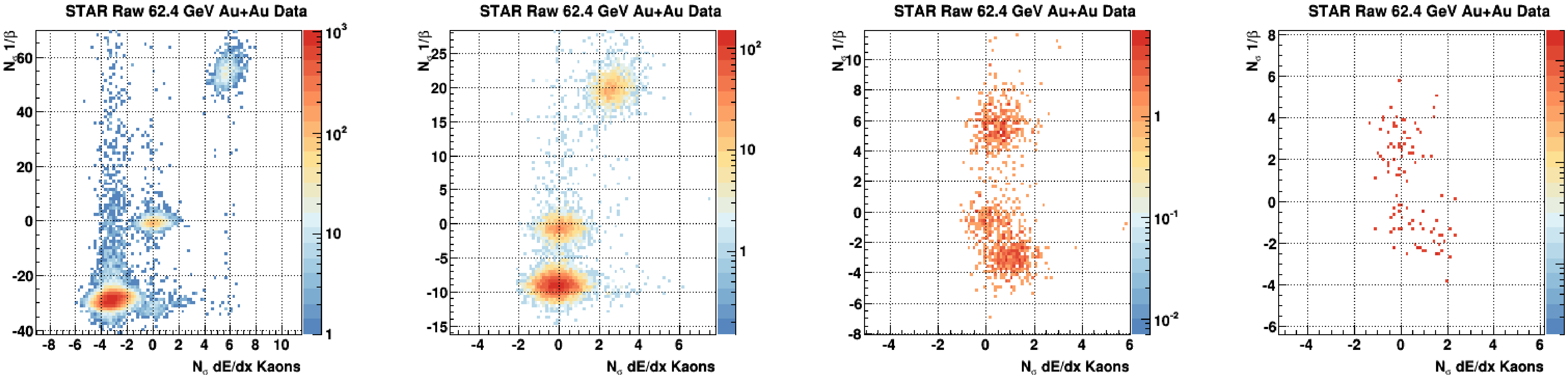
How does it perform with less contrived examples?

SIMULATED DATASET: 1 BILLION TRACKS

- π/K/p momentum spectra approximates those from Au+Au 62.4 GeV
- Time-of-Flight and dE/dx measurements are made for each particle
  - Non-Gaussian Shapes
  - Landau distribution for 15-45 dE/dx hits with highest 30% of measurements rejected for each track
  - Possibility of similar momentum tracks merging and having higher dE/dx measurements
  - Possibility of ToF measurement mistmatches between different particles
  - Student's t-distribution for ToF measurement resolution
- Momentum has limited resolution (2% on curvature)
- Finite momentum bin width size (50 MeV bins)
- **It's dirty, it's messy, and it's realistic**



Pretty similar, I'm not trying to brush anything under the rug here.

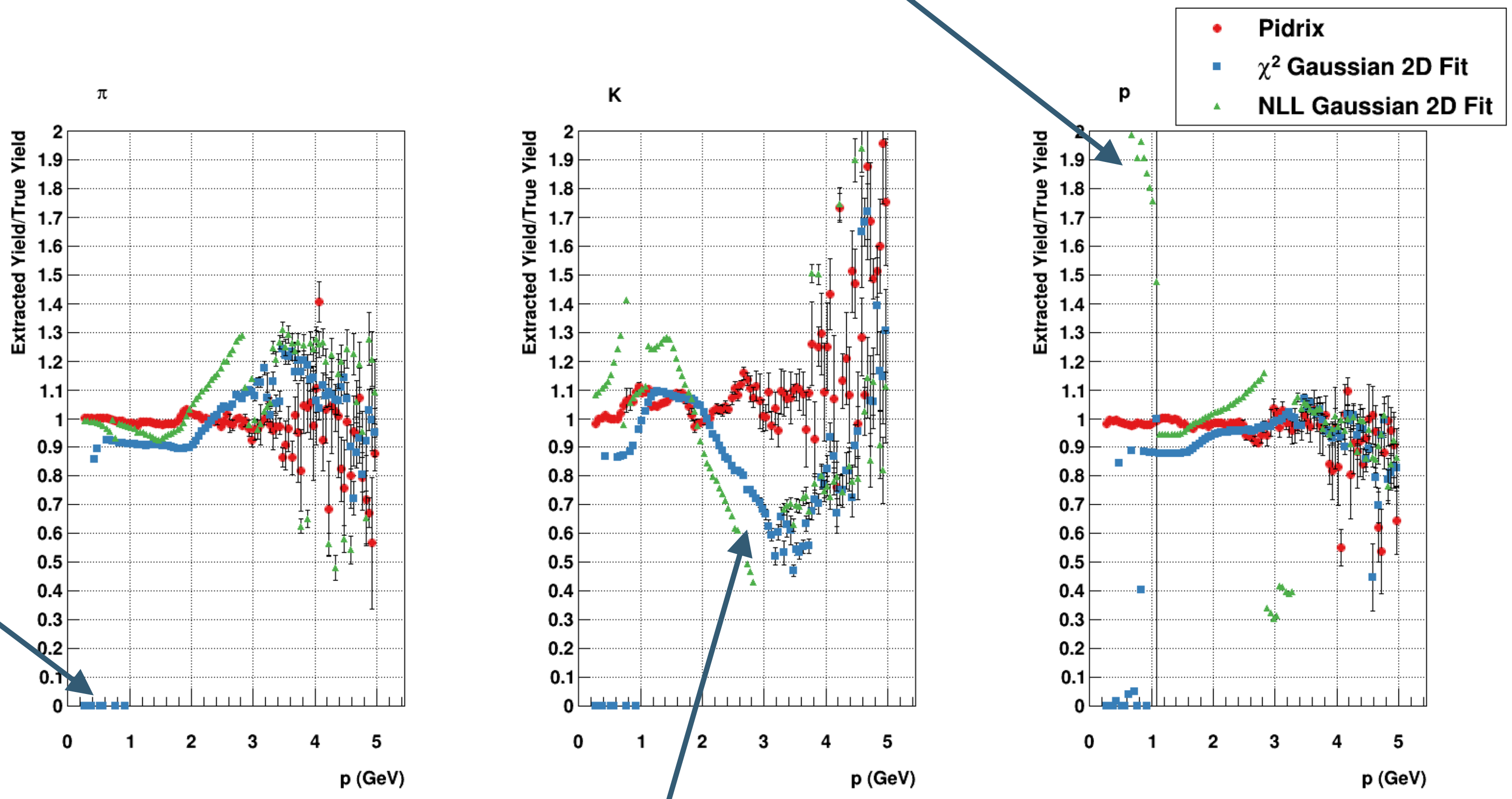


NOW WE'LL COMPARE...

We're about to look at the ratio of Gaussian model fits and Pidrix fits over the true yields from simulations.

Off the charts!

Yuck, failed fits. What do we have to fudge to get these to work?



Pidrix outperforms the model fits, especially at high p

We don’t realize when we make this mistake because the spectra still looks realistic.

WHAT ABOUT ERRORS?

- Perform Gibb's sampling on elements of U and V
  - Interpolate between neighboring elements to set scale of fluctuations (to enforce symmetry of the transition)
    - Add one to interpolation to avoid zeros from propagating
  - Add Gaussian value with width of 10% of the fluctuation scale
  - Reject negative values
- Use k-means clustering between samples to associate particles
  - Various norms ('means and yields', symmetrized Kullback-Leibler, etc)
- This results in likelihood distributed U and V matrices
  - Directly compute standard deviations of yields, means, single particle distributions, or whatever you want

HOW DO THEY COMPARE?

What factor would the yield error have to be multiplied by to be correct?

	Pidrix	Chi-squared	Log-likelihood
π	4.6	750000	77
K	7.2	230000	160
P	4.2	270000	360

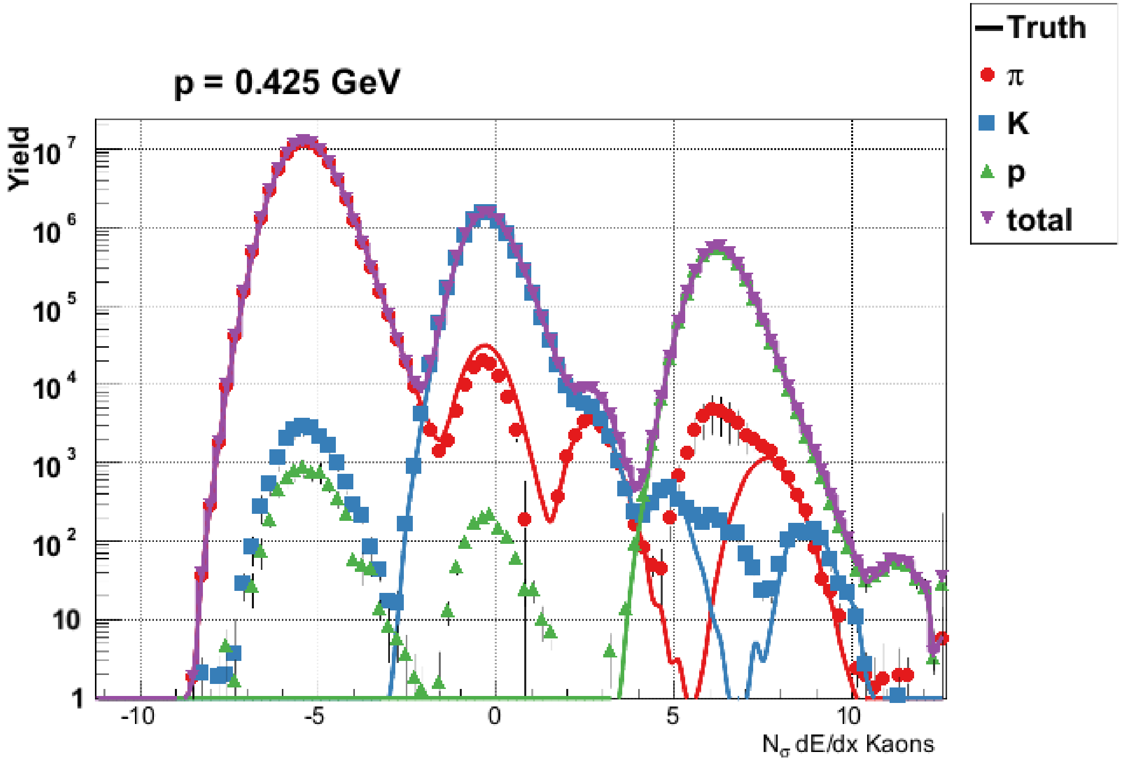
Minuit errors don't mean much when the model isn't exactly right.

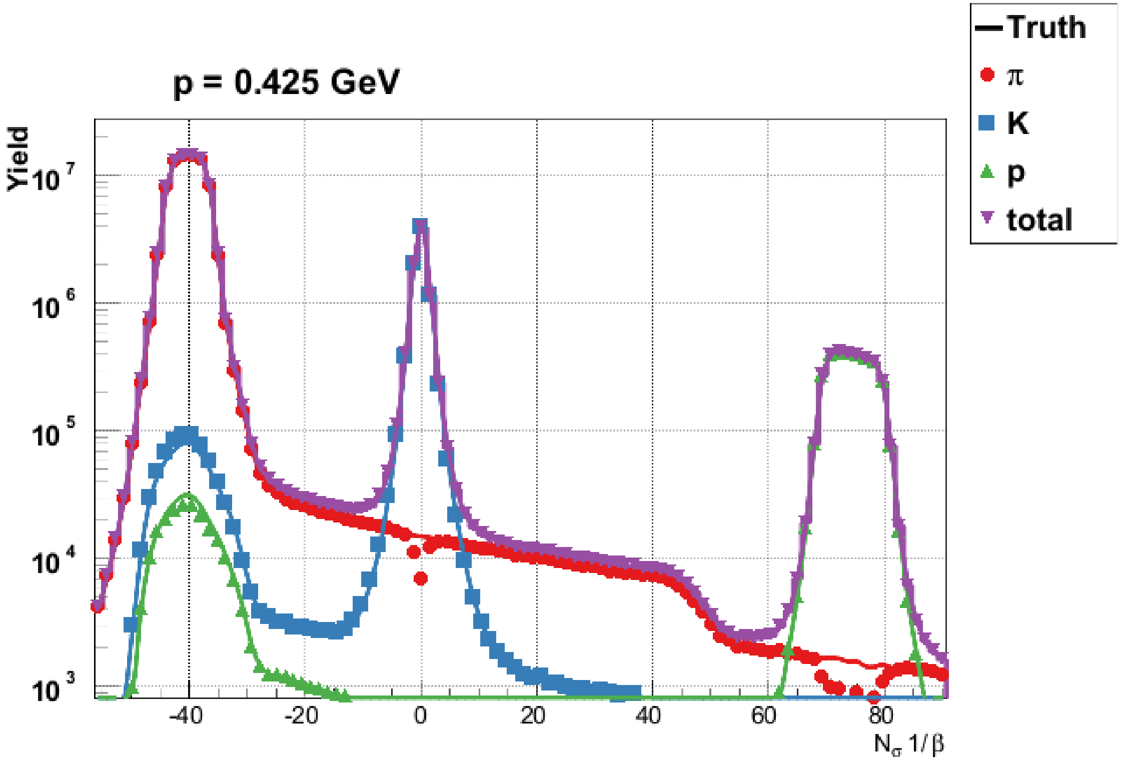
Pidrix gives *far more accurate errors* though still not great.

Small errors aren't better if they're wrong!

BACK TO U AND V…

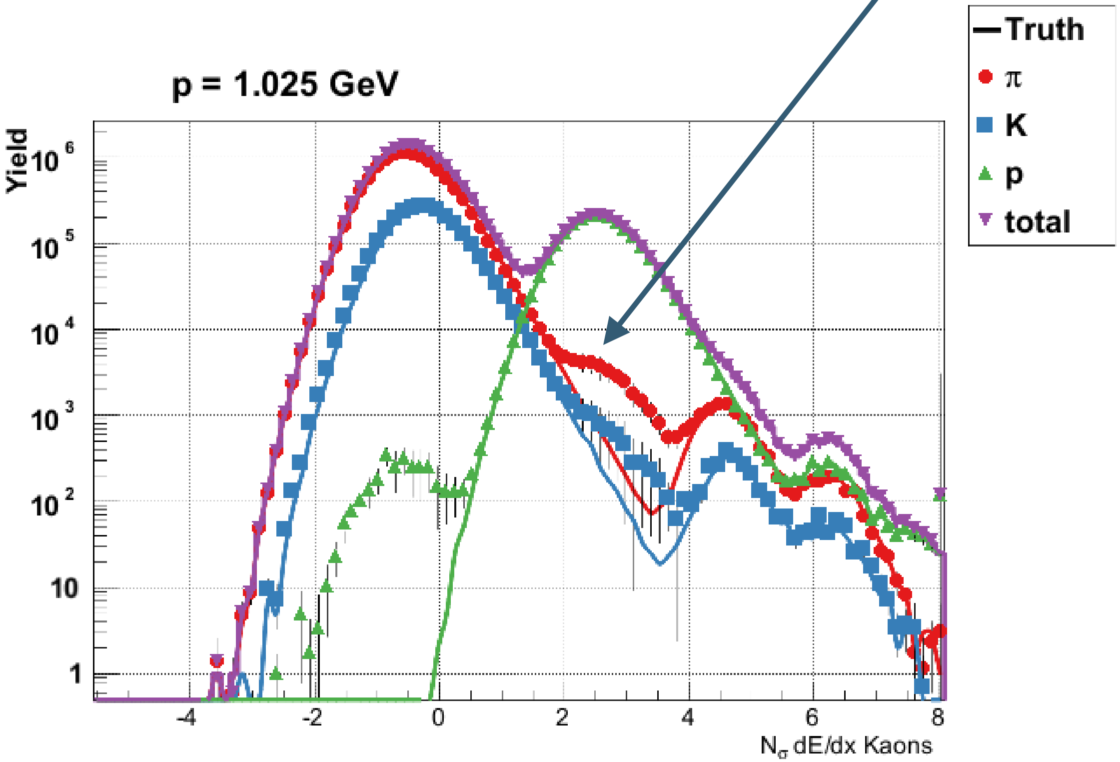
Their columns and rows have physical meaning.

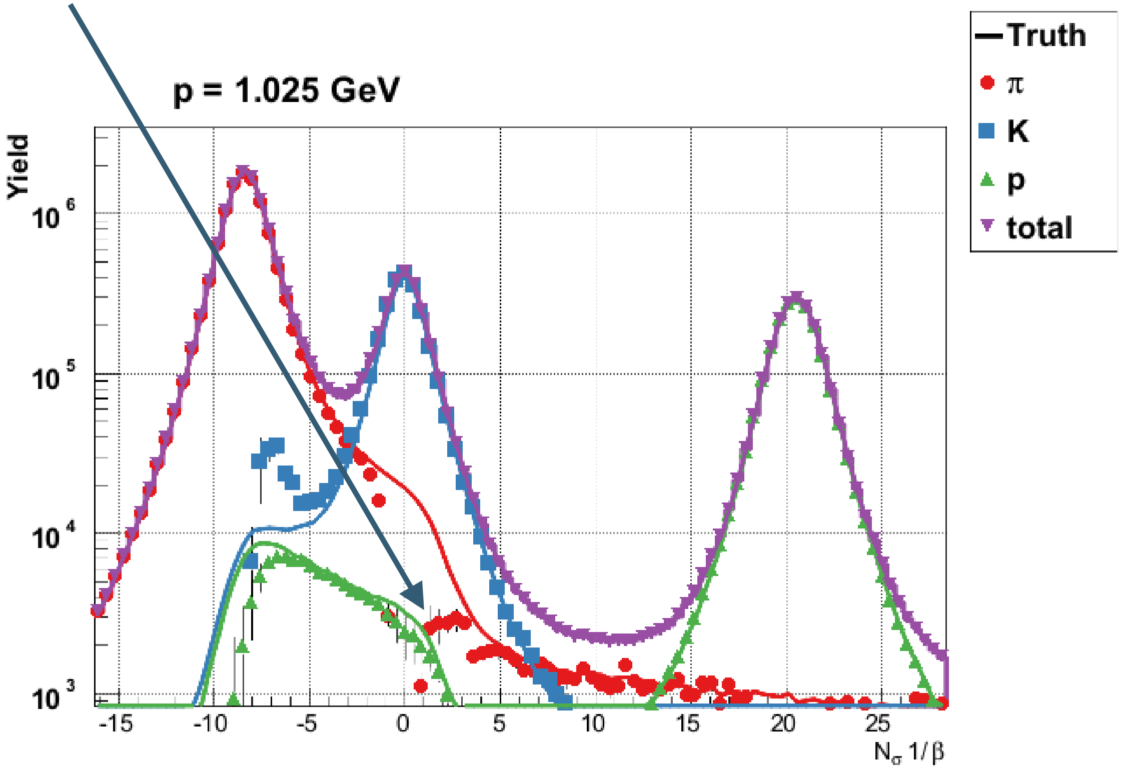




SOMETIMES MINOR MIXING ISSUES UNDER PEAKS

Is sampling not fully exploring the space?

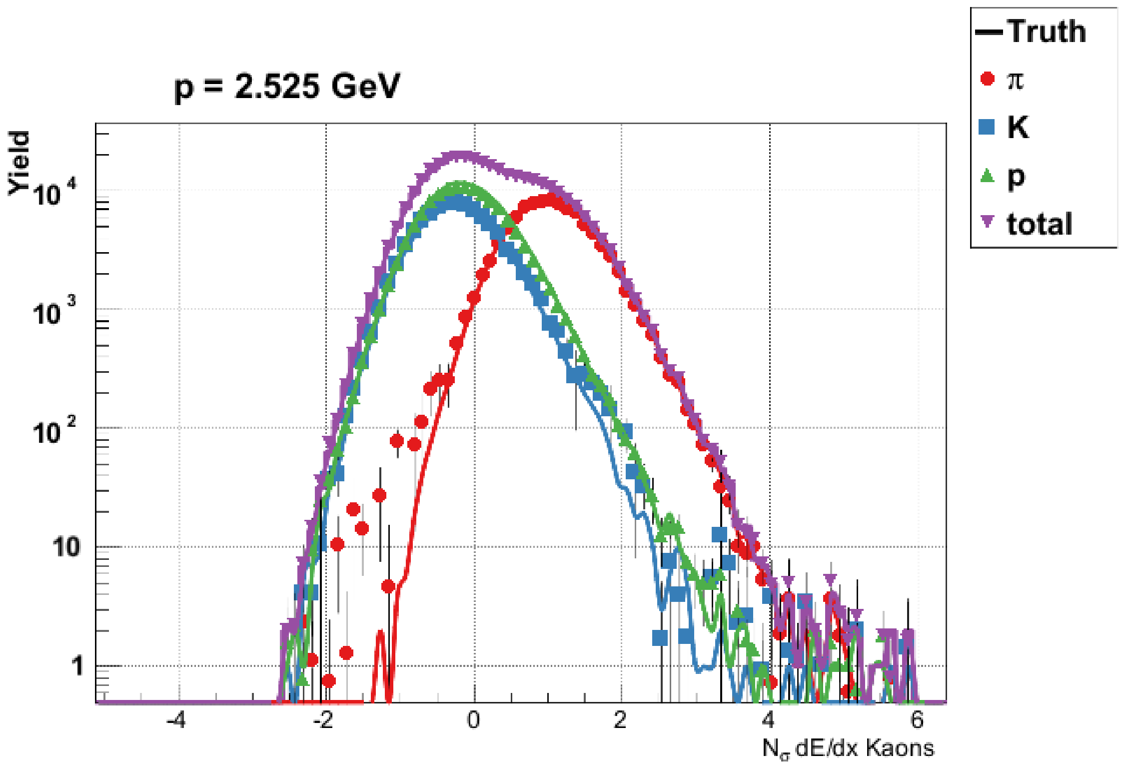


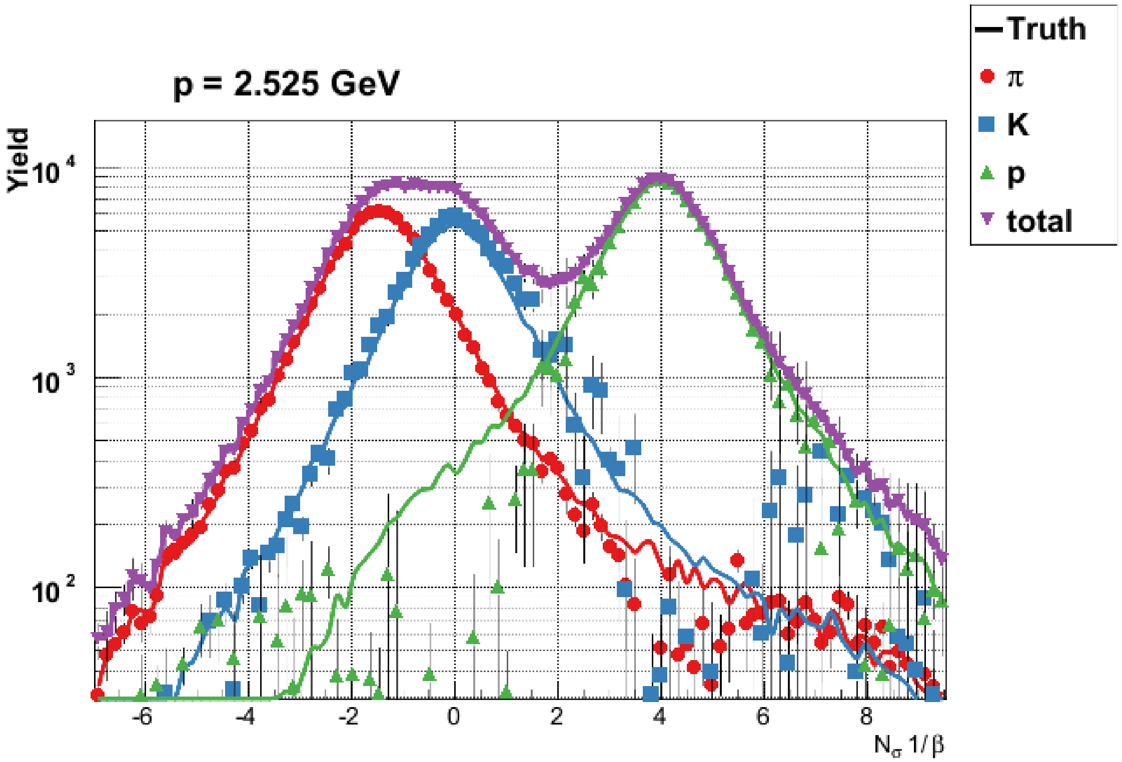


Or does momentum smearing drive this?

PIDRIX PARTICULARLY EXCELS AT HIGH P

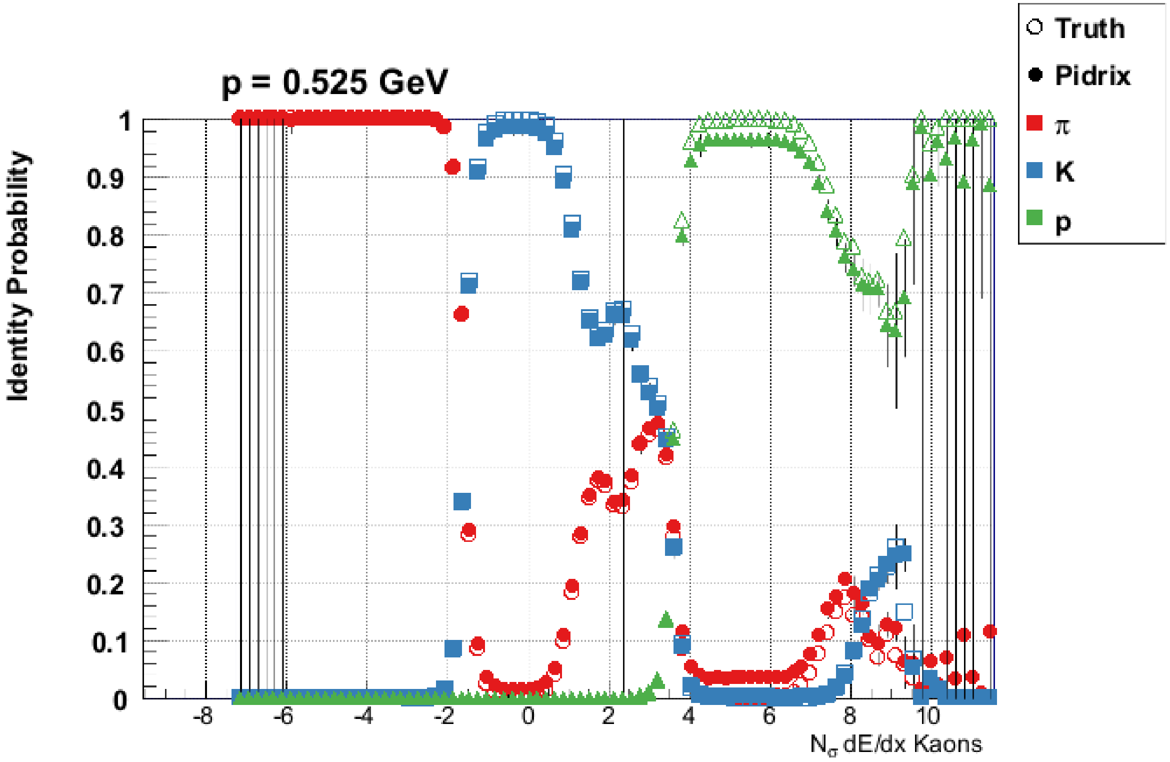
This region is extremely difficult to model.

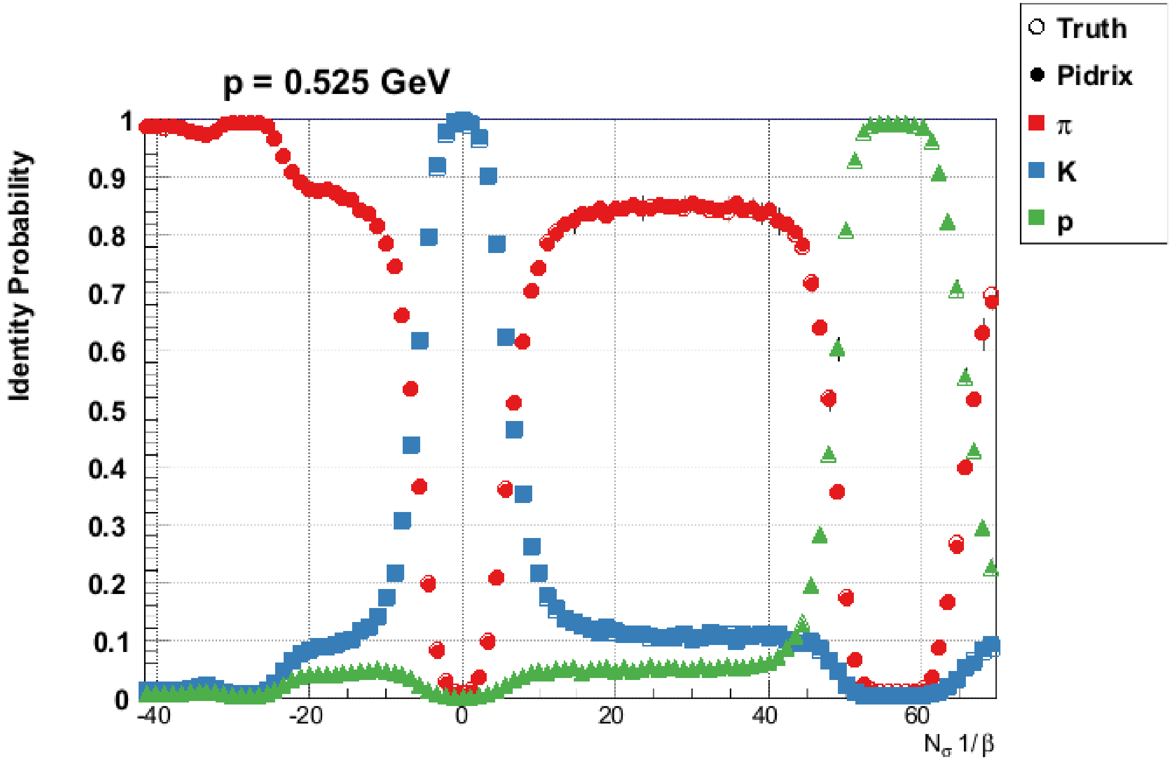




PURITY AND EFFICIENCY

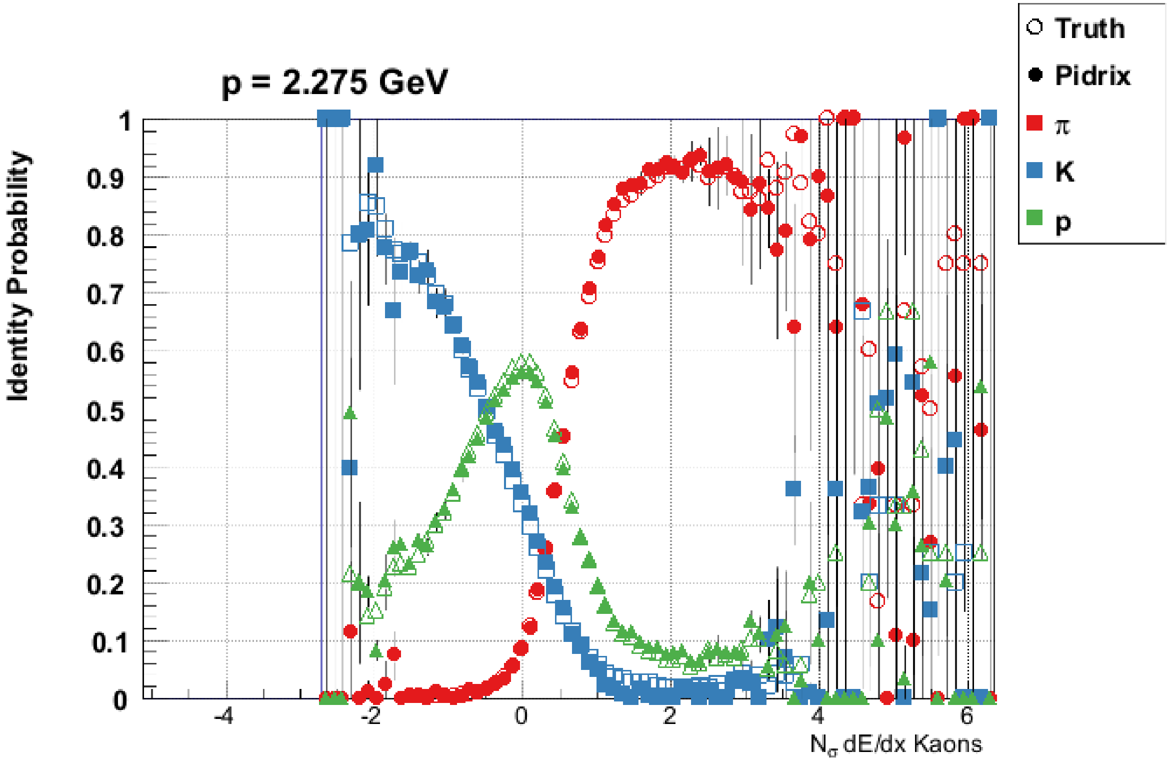
Want some help picking your cuts for high purity?

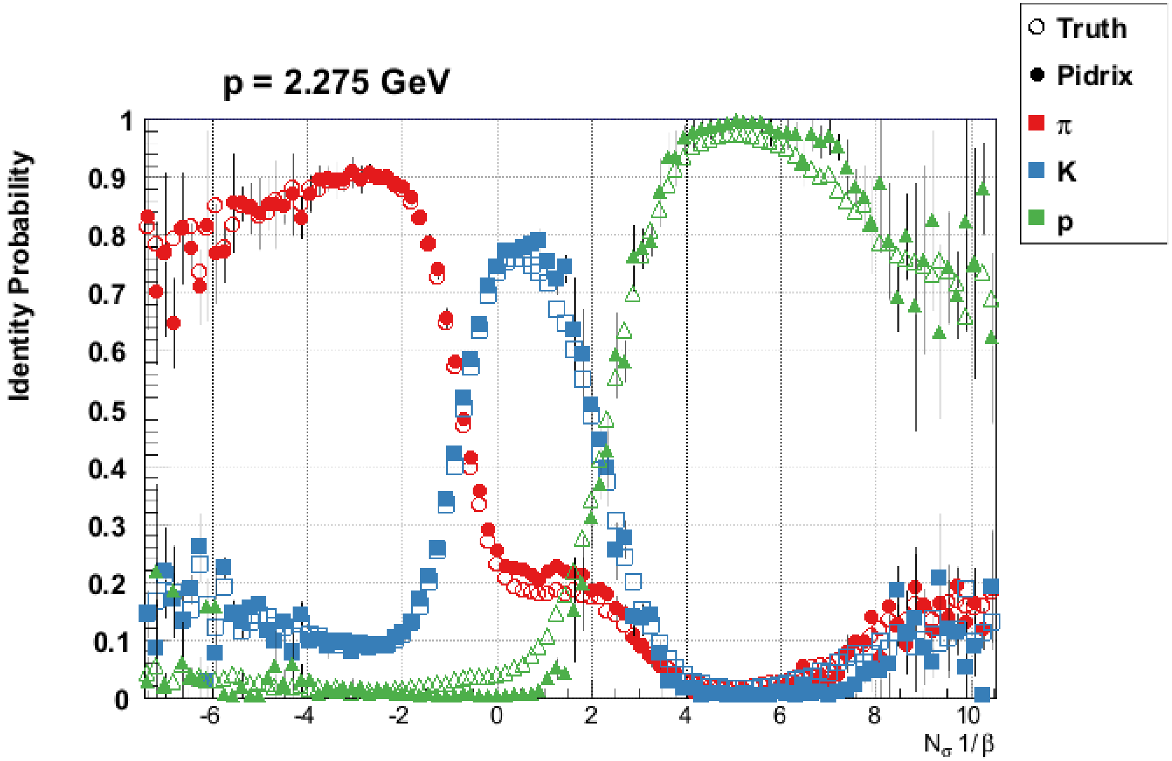




PURITY AND EFFICIENCY

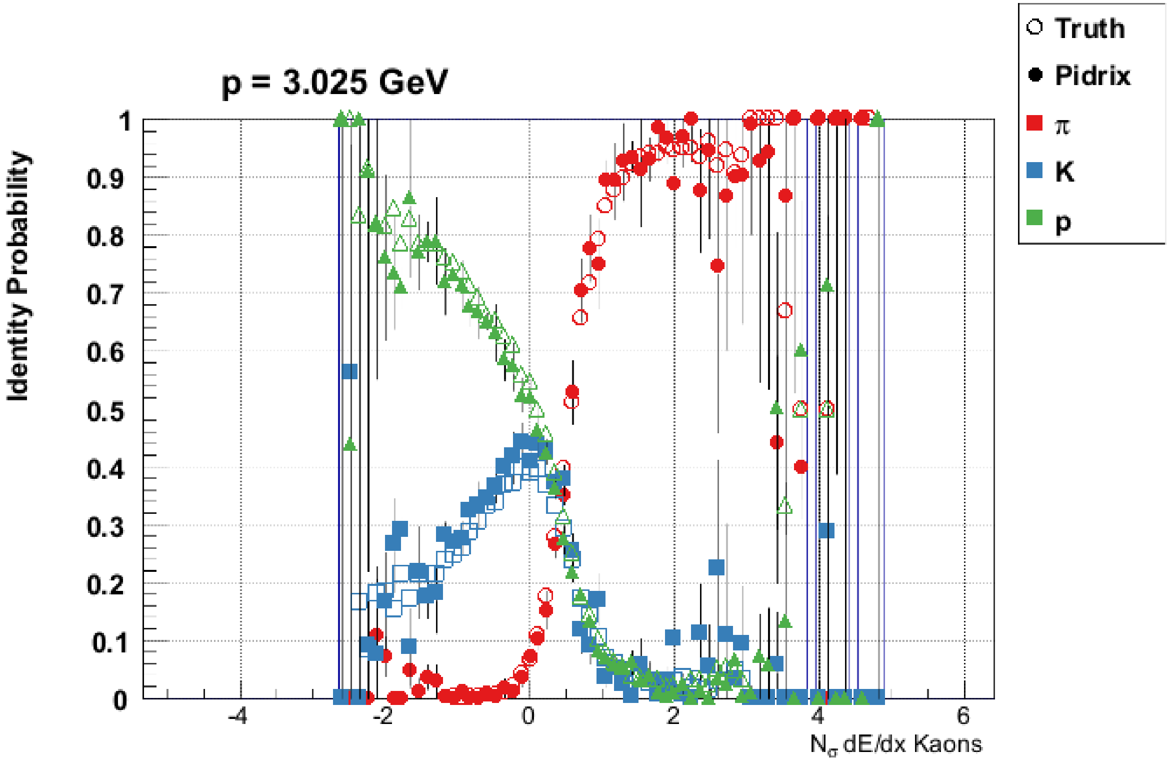
Or calculating what your cut efficiency is?

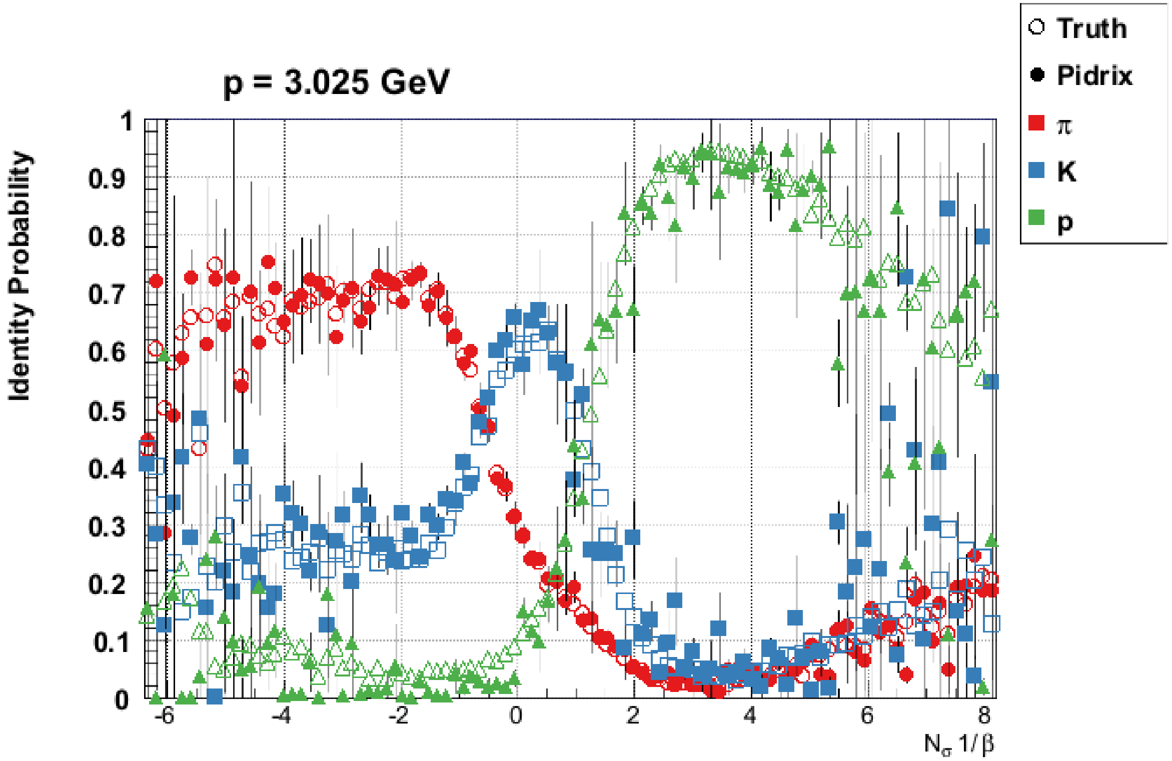




PURITY AND EFFICIENCY

These can be extracted directly from U and V.





REITERATION

- We're doing a negative log-likelihood minimization
  - Just not using a model and being clever about how we do this
- The worst assumption is that the measurements are uncorrelated
  - Even when this is broken by momentum resolution and finite bin width it still works
  - Progress is being made to accommodate breaking this assumption
- The same code applies to different cuts, binnings, and even experiments
  - No model means no model tuning
- It seems to work very well!

LESS IS MORE?

- A new method for particle identification has been described
  - Faster, easier, and less assumption oriented than model based fitting
  - Extremely promising accuracy observed so far
- FOSS library is available:
  - https://github.com/sangaline/pidrix
  - Documentation is lacking… examples coming soon
- Interested in using this method?
  - Please email esangaline@gmail.com
  - Especially if your experiment has >2 PID measurements for each track!