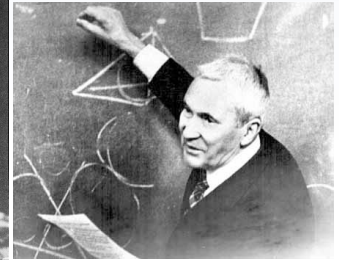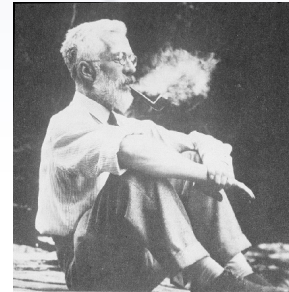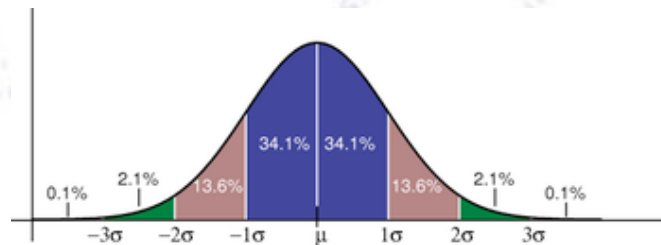# Applied Statistics

Advanced fitting

Troels C. Petersen (NBI)

"Statistics is merely a quantisation of common sense"

Defining the Chi-Square

<u>Problem Statement</u>: Given N data points (x,y), adjust the parameter(s) θ of a model, such that it fits data best.

The best way to do this, given uncertainties σ_i on y_i is by minimising:

$$\chi^2(\theta) = \sum_i^N \frac{(y_i - f(x_i, \theta))^2}{\sigma_i^2}$$

The power of this method is hard to overstate!

Not only does it provide a simple, elegant and unique way of fitting data, but more importantly it provides a **goodness-of-fit measure**.

This is the Chi-Square test!

Chi-Square probability interpretation

The Chi-Square probability can roughly be interpreted as follows:

- If χ^2 / **Ndof** ≃ **1** or more precisely if **0.01 < p(χ^2,Ndof) < 0.99**, then all is good.

- If χ^2 / **Ndof** ≫ **1** or more precisely if **p(χ^2,Ndof) < 0.01**, then your fit is bad, and your hypothesis is probably not correct.

- If χ^2 / **Ndof** ≪ **1** or more precisely if **0.99 < p(χ^2,Ndof)**, then your fit is TOO good and you probably overestimated the errors.

If the statistics behind the plot is VERY high (great than 10^6), then you might have a hard time finding a model, which truly describes all the features in the plot (as now tiny effects become visible), and one hardly ever gets a good Chi-Square probability.
However, in this case, one should not worry too much, unless very high precision is wanted.
Anyway, the Chi-Square still allows you to compare several models, and determine which one is the better.

Notes on the ChiSquare method

"It was formerly the custom, and is still so in works on the theory of observations, to derive the method of least squares from certain theoretical considerations, the assumed normality of the errors of the observations being one such.
*It is however, more than doubtful whether the conditions for the theoretical validity of the method are realised in statistical practice, and the student would do well to regard the method as recommended chiefly by its comparative simplicity and by the fact that it has **stood the test of experience"**.*

[G.U. Yule and M.G. Kendall 1958]

Example of Chi-Square
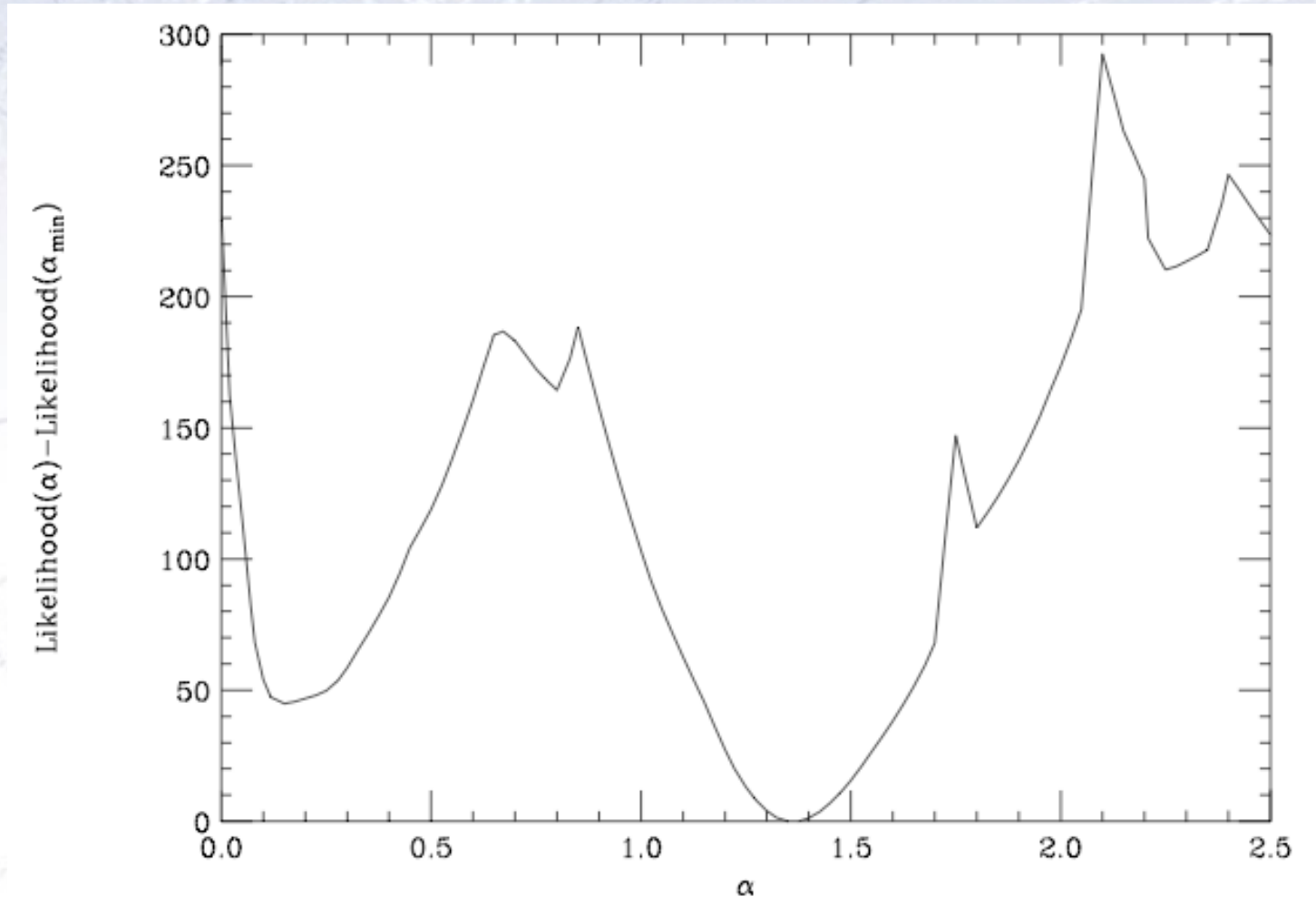
The fact that there are several minima makes fitting difficult/uncertain!
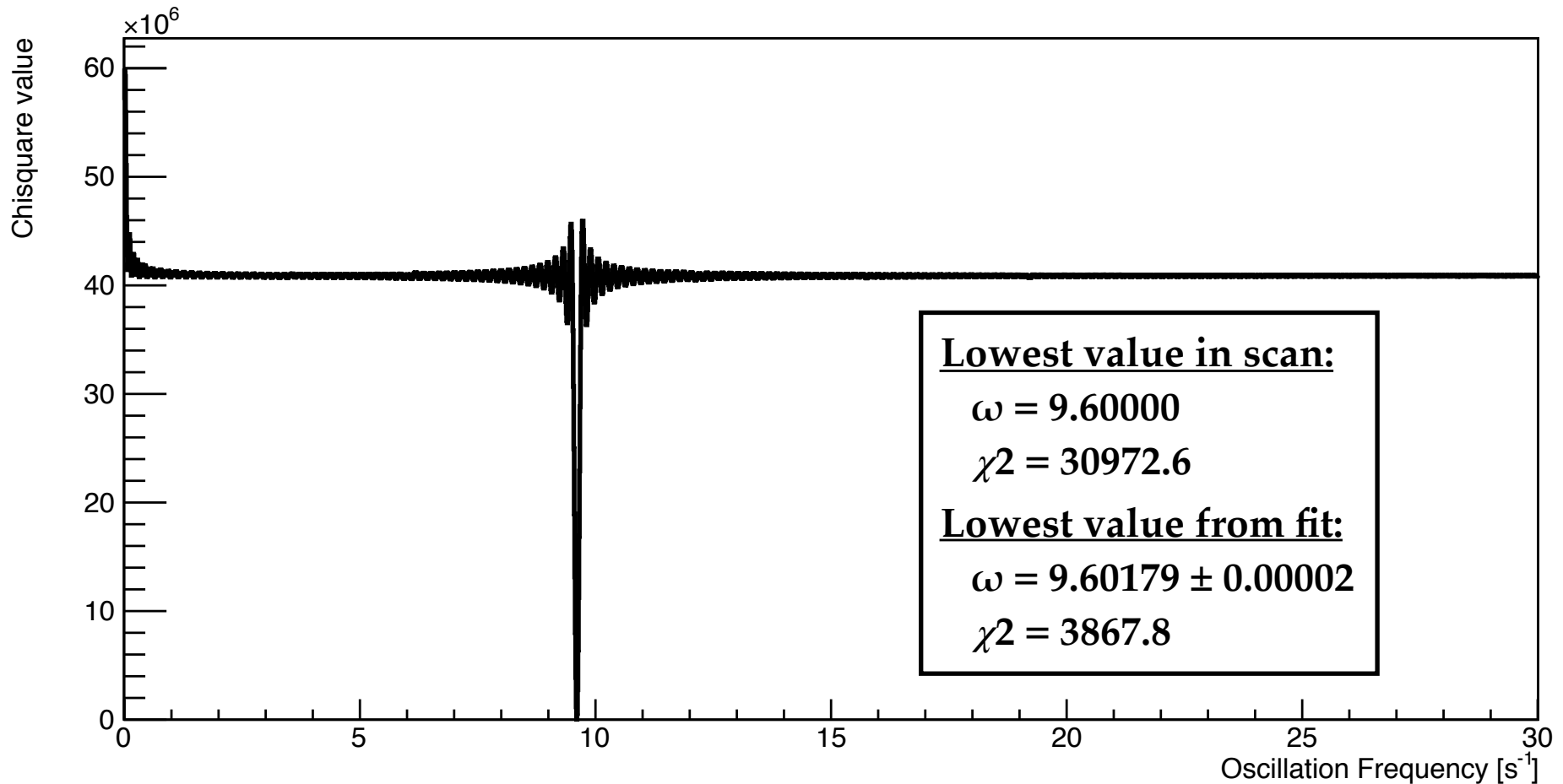*Always give good starting values!!!*

5

Example of Chi-Square

Especially fitting oscillatory data requires a good starting value for omega.
Even a small offset may result in a Chi2, which is not sensitive to omega!

χ² / ndf	3868 / 3796
Prob	0.2044
p0	−0.104 ± 8.111e−05
p1	0.5928 ± 0.0002452
p2	0.007427 ± 2.032e−05
p3	9.602 ± 2.035e−05
p4	2.819 ± 0.0004148

Example of Chi-Square

Especially fitting oscillatory data requires a good starting value for omega.
Even a small offset may result in a Chi2, which is not sensitive to omega!

Lowest value in scan:
  $\omega = 9.60000$
  $\chi2 = 30972.6$
Lowest value from fit:
  $\omega = 9.60179 \pm 0.00002$
  $\chi2 = 3867.8$
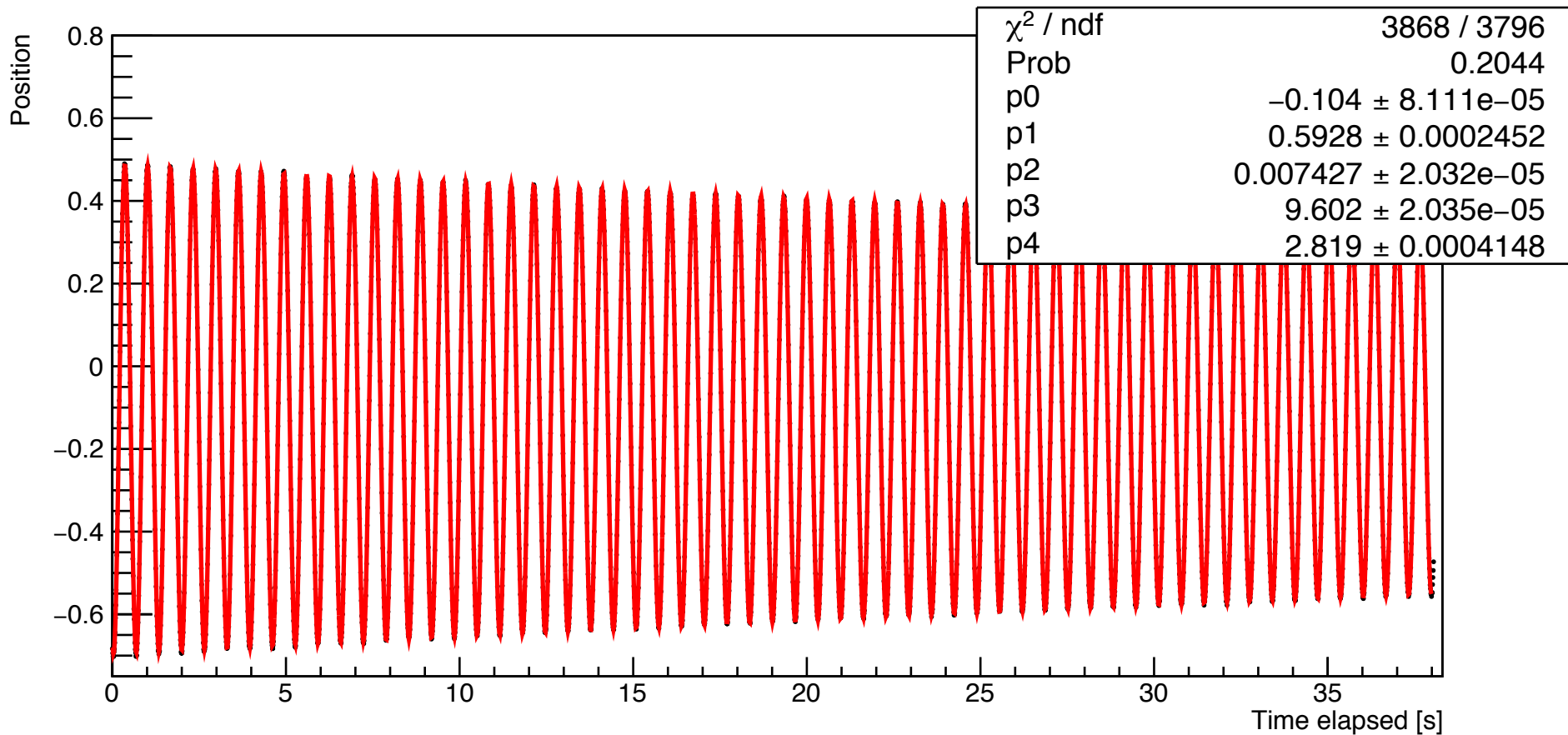
Very advanced fitting

Sometimes, one has several samples (signal, control, validation, background, etc.) to fit, and one would like to include the full information in all of these.

If one (input) sample dominates in size, then there is no reason not to fit this sample separately, and then just fix the parameters to the result of this fit.

However, if they are of similar size, then a simultaneous fit of the samples is the optimal solution. This can grow quite large…

In 2002 the BaBar collaborations fit for CP-violation ($\sin 2\beta$) included **98 floating parameters** applied to four datasets.

In 2005 the BaBar collaborations fit for mixing in the D0-system included **120 floating parameters** applied to six datasets.

But that was nothing compared to…

The ATLAS Higgs discovery fit
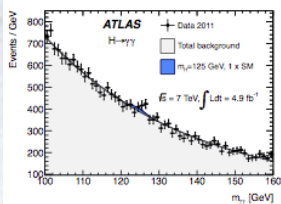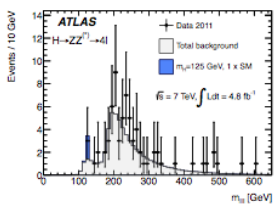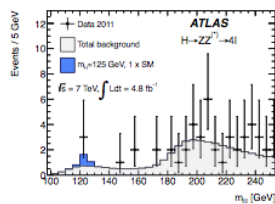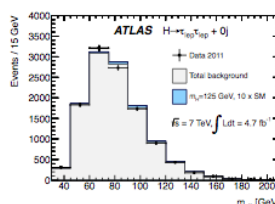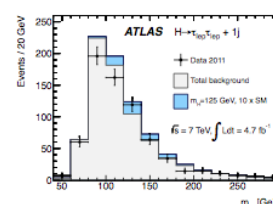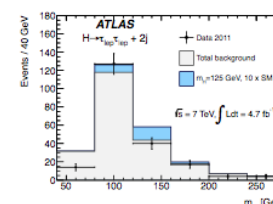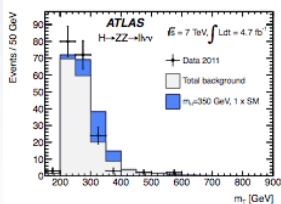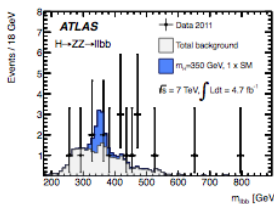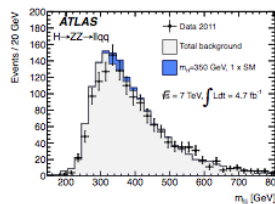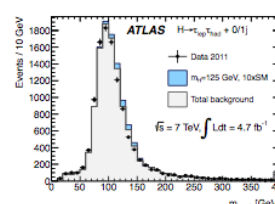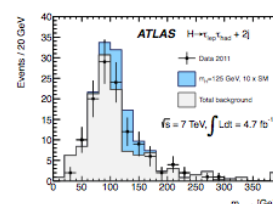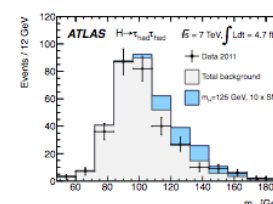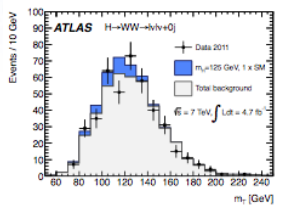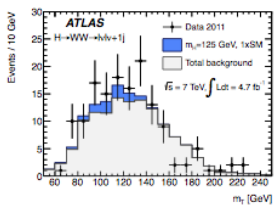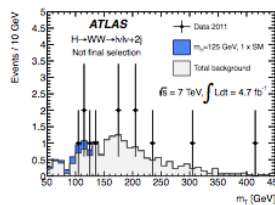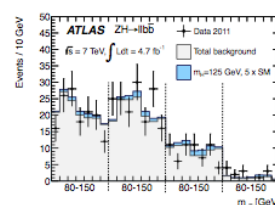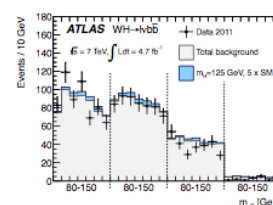
Fitting tips & tricks

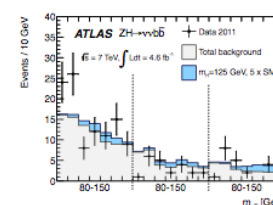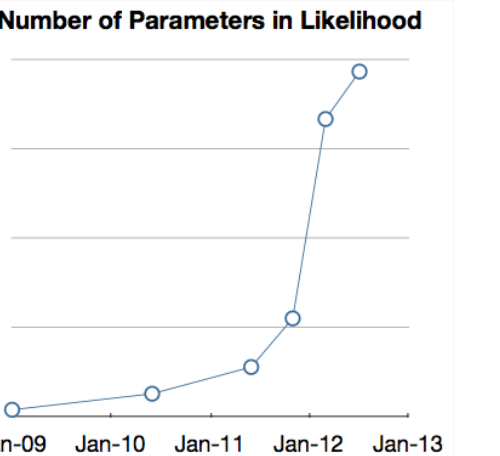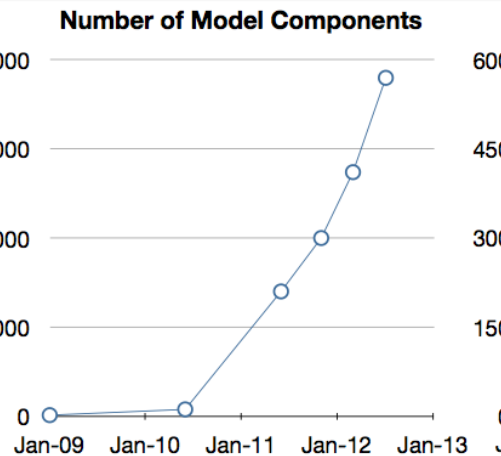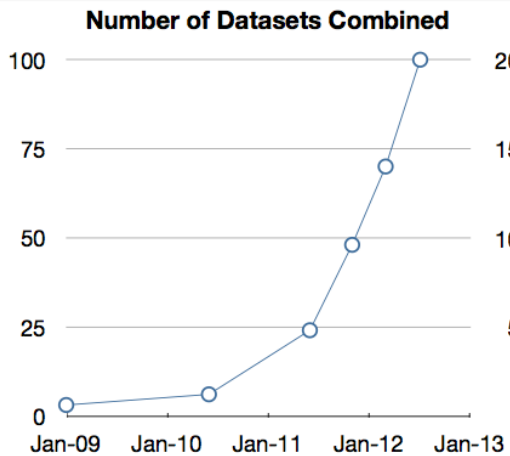There are a few tips & tricks that will make your (fitting) life a bit easier:
- Always give good initial values!!!
- Never start with an advanced fit - make a simple one work and expand!
- Try to make your parameters as little correlated as possible.
- Let the parameters represent the quantities of interest.
- Start with a ChiSquare fit, as these usually has better convergence.

When a fit refuses to work, try the following:
- Draw function on top of data to check formula and quality of initial values.
- Check the correlations between the parameters.
- Try to fix one or more parameter to a value you find reasonable.

Even with all of this advice, there is no guaranty that your fit will work.
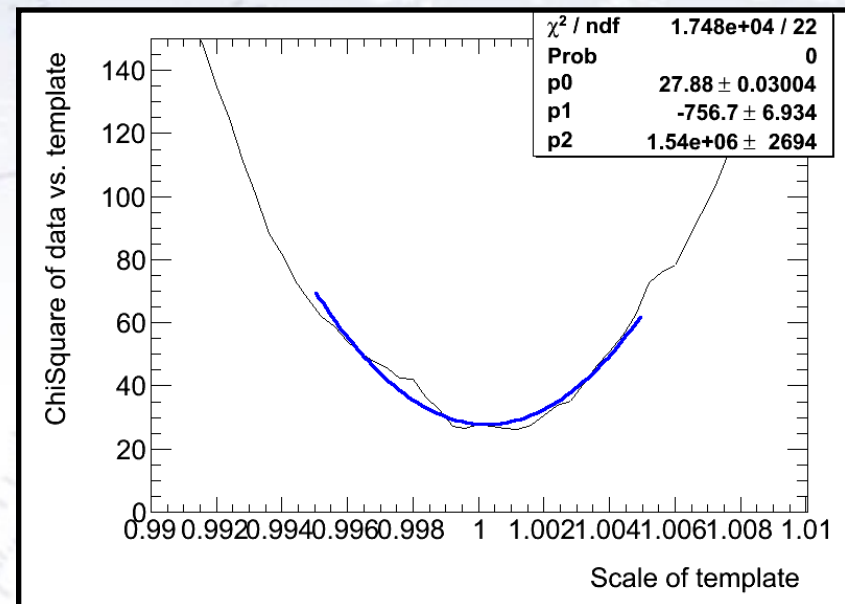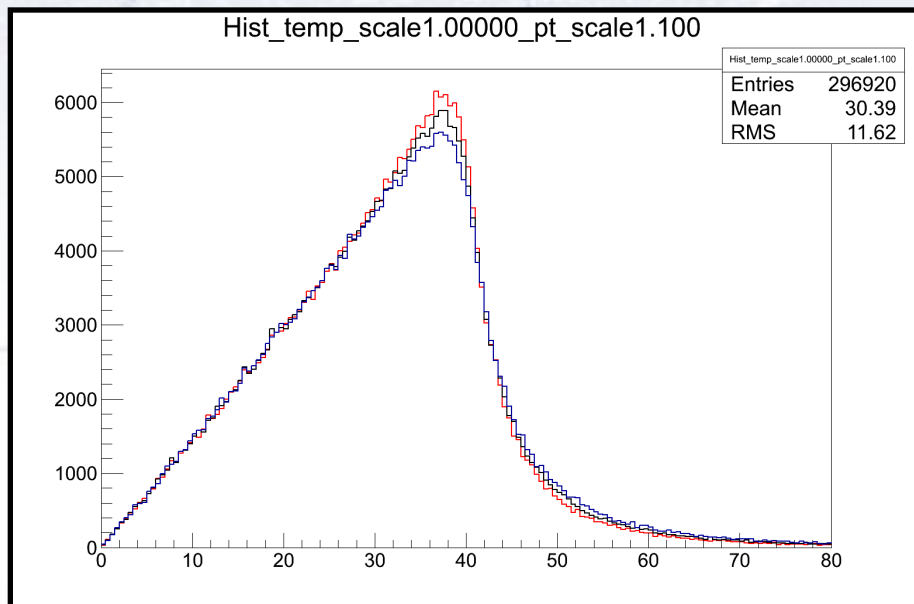It is after all a bit of an art….

Fitting equations/constraints Examples…

Fitting with templates

Sometimes, the shape to be fitted can not be expressed as a function, but obtained through a histogram from simulation/data.


For each template, one calculates the ChiSquare between the data and the template. You then repeat it for all templates, and subsequently obtain a parabola with a minimum (central value) and a curvature (uncertainty).

Setting the momentum scale

A well known type of event in particle physics (at he LHC accelerator) is the Z ➔ μ+ μ- decay, which has two muons flying out through the detector.

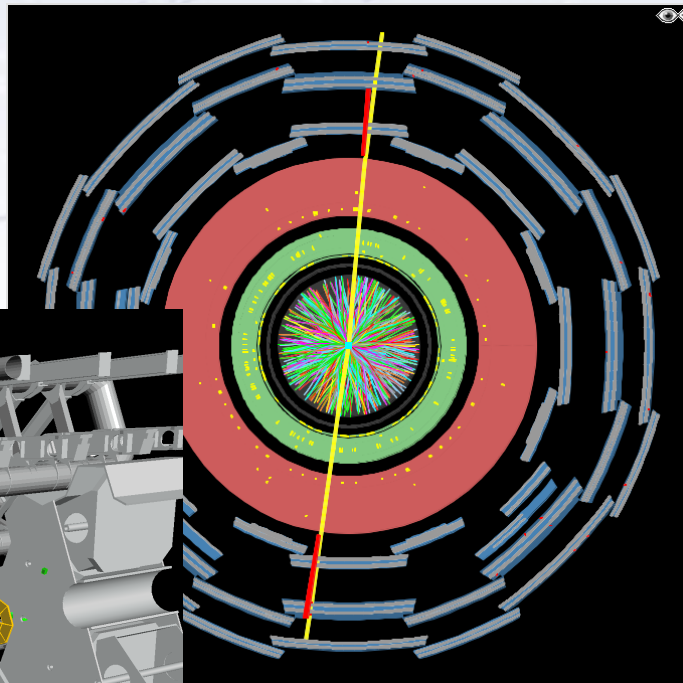As the Z-mass is well known, these events can be used to check and correct the reconstruction.

ATLAS EXPERIMENT
Run Number: 201289, Event Number: 24151616
Date: 2012-04-15 16:52:58 CEST

ATLAS EXPERIMENT
Run: 154822, Event: 14321500
Date: 2010-05-10 02:07:22 CEST
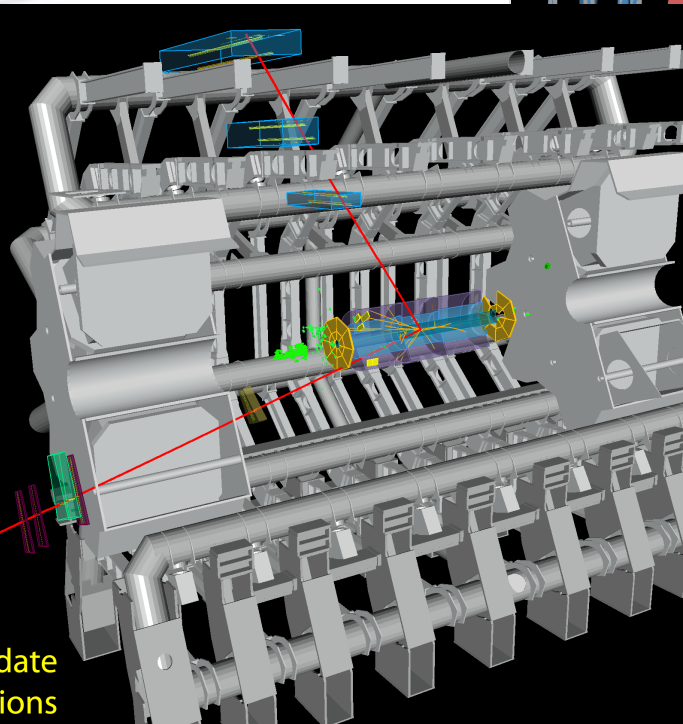
$p_T(\mu^-) = 27$ GeV $\eta(\mu^-) = 0.7$
$p_T(\mu^+) = 45$ GeV $\eta(\mu^+) = 2.2$
$M_{\mu\mu} = 87$ GeV

Z➔μμ candidate
in 7 TeV collisions
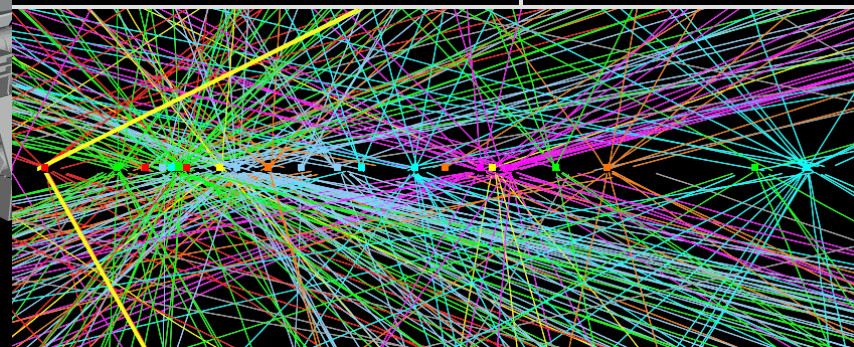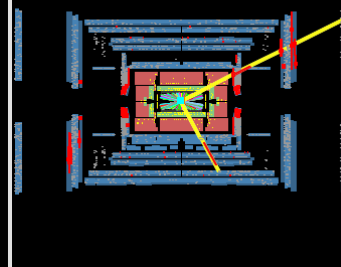
Setting the momentum scale

There are many ways in which this reconstruction can be biased due to the detector not being aligned correctly.

	ΔR	Δφ	ΔZ
R	**Radial Expansion** (distance scale)	**Curl** (Charge asymmetry)	**Telescope** (COM boost)
φ	**Elliptical** (vertex mass)	**Clamshell** (vertex displacement)	**Skew** (COM energy)
Z	**Bowing** (COM energy)	**Twist** (CP violation)	**Z expansion** (distance scale)

Charge dependent fit

The approach follows ATLAS-CONF-2012-141:

There is a charge independent (radial):

$$p \; \rightarrow \; p\left(1 + \delta_{\text{radial}}\right)$$

The charge and p_T dependent (sagitta):

$$q/p \; \rightarrow \; q/p\left(1 + qp_T\delta_{\text{sagitta}}\right)$$

The improvement is that the charge dependent should take the momentum into account, as the effect changes with momentum.

I chose the eta binning of the fit to have 24 or 51 bins.
The fit thus provides 24/51 values for δr and 24/51 for δs.

Introduction to idea

In order to use all Z ➤ ll events, the idea is to:

- Divide leptons into bins in eta, pt, phi, charge.

  Bin definition should match variables we are interested in [here: 51 in eta, 2 in charge]

- For lepton pair of bins **ij**, plot Z mass for data and MC (i.e. N^2 plots).

  The N^2 plots limits number of bins to about 100. Not all **ij**-values are filled, which is OK.

- For each **ij** data-MC pair, determine value **α_{ij}**:

$$\alpha_{ij} = m_Z^{MC} / m_Z^{data}$$

  The problem is: "from which lepton i or j" does the bias come from?

- Non-unit values are from lepton bins i and j. Obtain them, by minimizing:

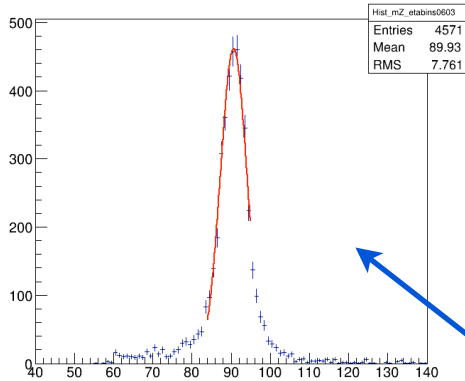$$\chi^2(\beta_i) = \sum_{ij} \left(\frac{\alpha_{ij} - \sqrt{\beta_i \beta_j}}{\sigma(\alpha_{ij})} \right)^2$$

From this we obtain the momentum/energy scales β for each bin i.
There are N^2 α_{ij} values (perhaps a bit less) to constrain the N β values.

Outline of the method
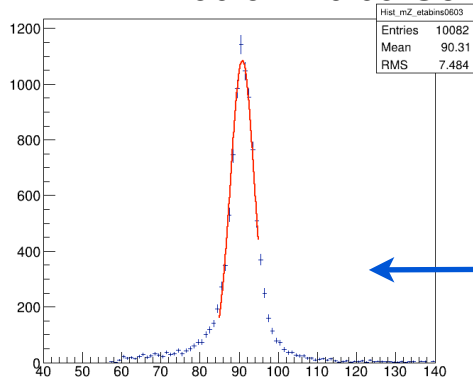
For each bin in η+η-,
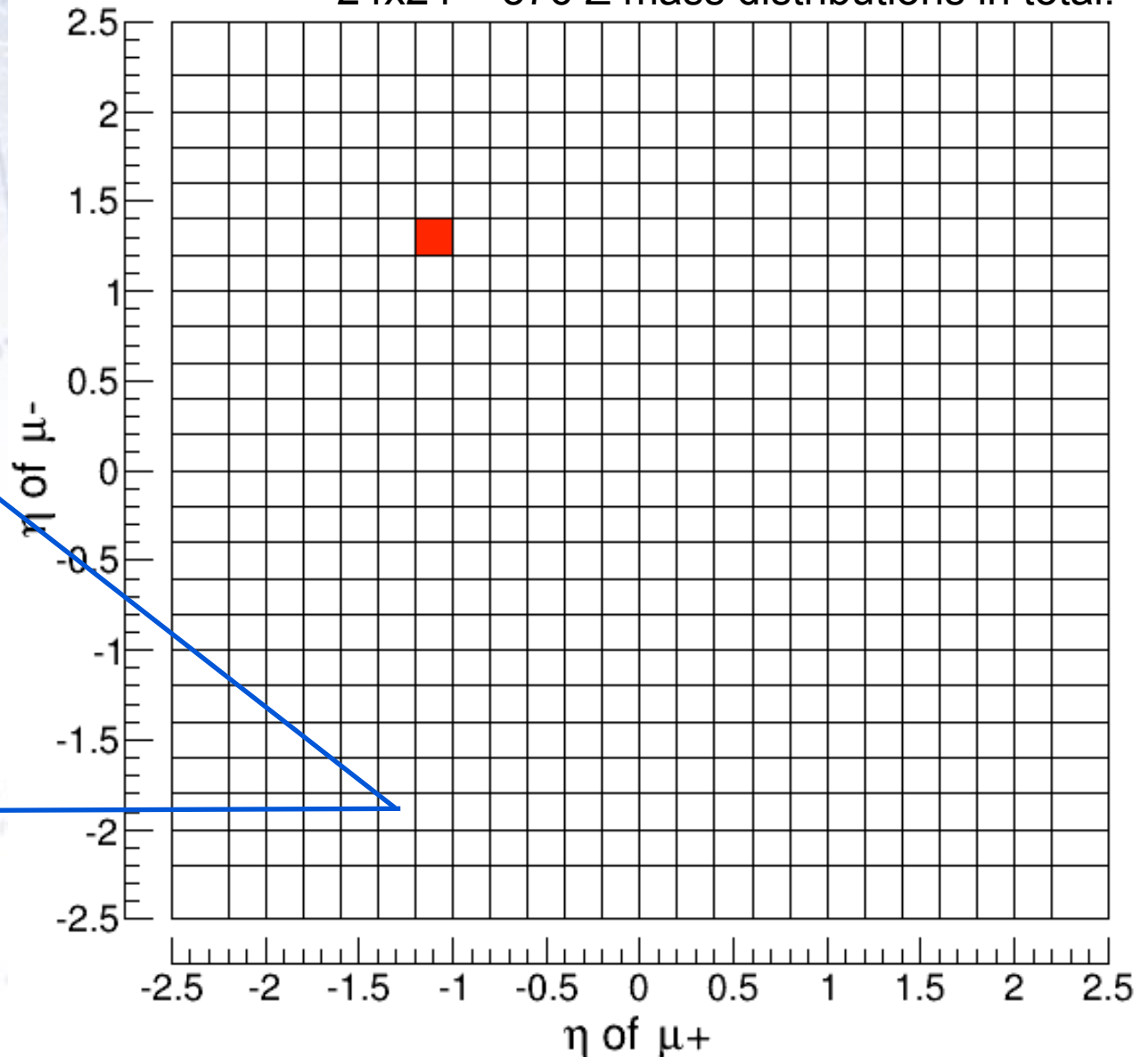fit the mass peak:

Data: mZ = 89.93 ± 0.13 GeV

MC: mZ = 90.31 ± 0.09 GeV

Ratio: = 0.9971 ± 0.0009 GeV

24x24 = 576 Z mass distributions in total!
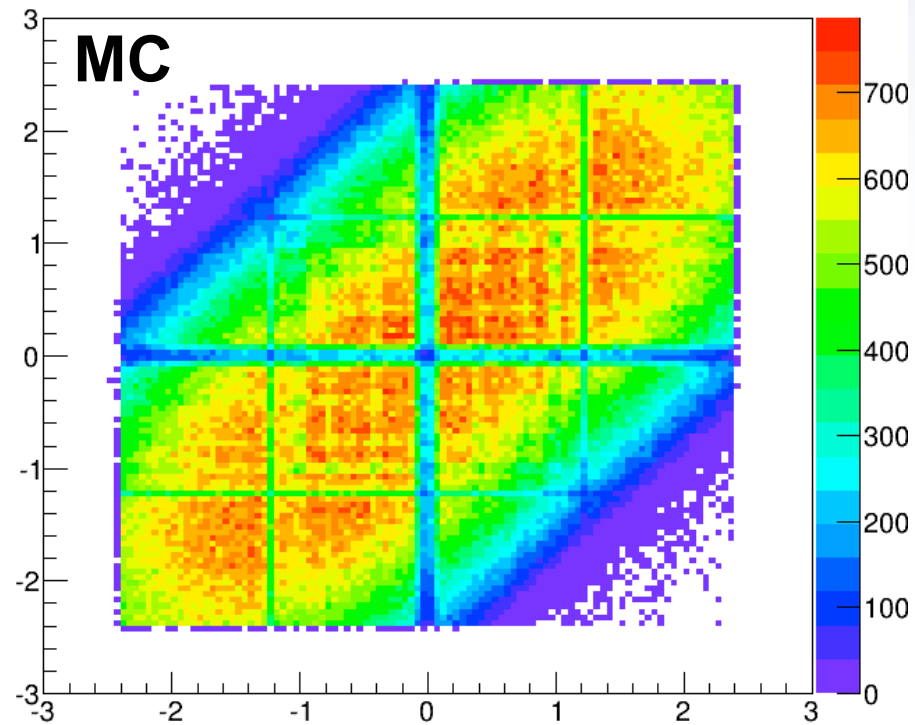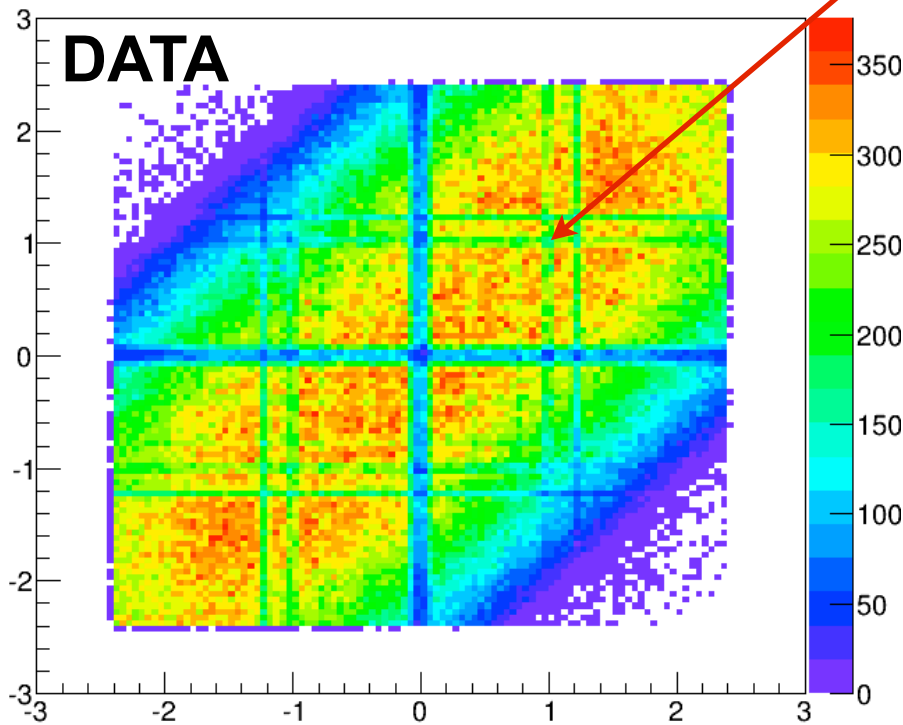

η of μ-

η of μ+

Inspection of eta distribution

Data: 1.64M events, MC: 4.7M events
At first sight, they seem pretty consistent. However, a closer look reveals differences. What is causing the difference at $|\eta| = 0.95$?

504 out of the 576 $\eta^+\eta^-$ has enough data (N > 25) and a good fit in both data and MC. These values are then used in the further fit.

Charge dependent fit

The 504 ratios of Z masses looks like this:

$$p \;\to\; p\,(1 + \delta_{\text{radial}})$$

$$q/p \;\to\; q/p\,(1 + qp_T\delta_{\text{sagitta}})$$

```
0  0  0.9936  0.0028
0  1  0.9980  0.0023
0  2  0.9974  0.0021
0  3  0.9940  0.0017
0  4  0.9946  0.0017
0  5  0.9931  0.0017
0  6  0.9957  0.0018
0  7  0.9977  0.0017
0  8  0.9980  0.0016
```

These values are used in a χ^2 as follows:

$$\chi^2 \;=\; \sum_{ij}\left(\frac{R(m_Z)_{ij} \;-\; \sqrt{(1+\delta r_i)(1+p_T\delta s_i)\,\times\,(1+\delta r_j)(1-p_T\delta s_j)}}{\sigma(R(m_Z)_{ij})}\right)^2$$

```
1   0  1.0006  0.0025
1   1  0.9979  0.0019
1   2  0.9963  0.0016
1   3  0.9968  0.0014
1   4  0.9960  0.0013
1   5  0.9969  0.0014
1   6  0.9971  0.0013
1   7  0.9962  0.0013
1   8  0.9959  0.0012
1   9  0.9965  0.0012
1  10  0.9980  0.0012
1  11  0.9986  0.0015
```

In short, the dr and ds values should minimise the expected difference between data and MC Z mass ratio.
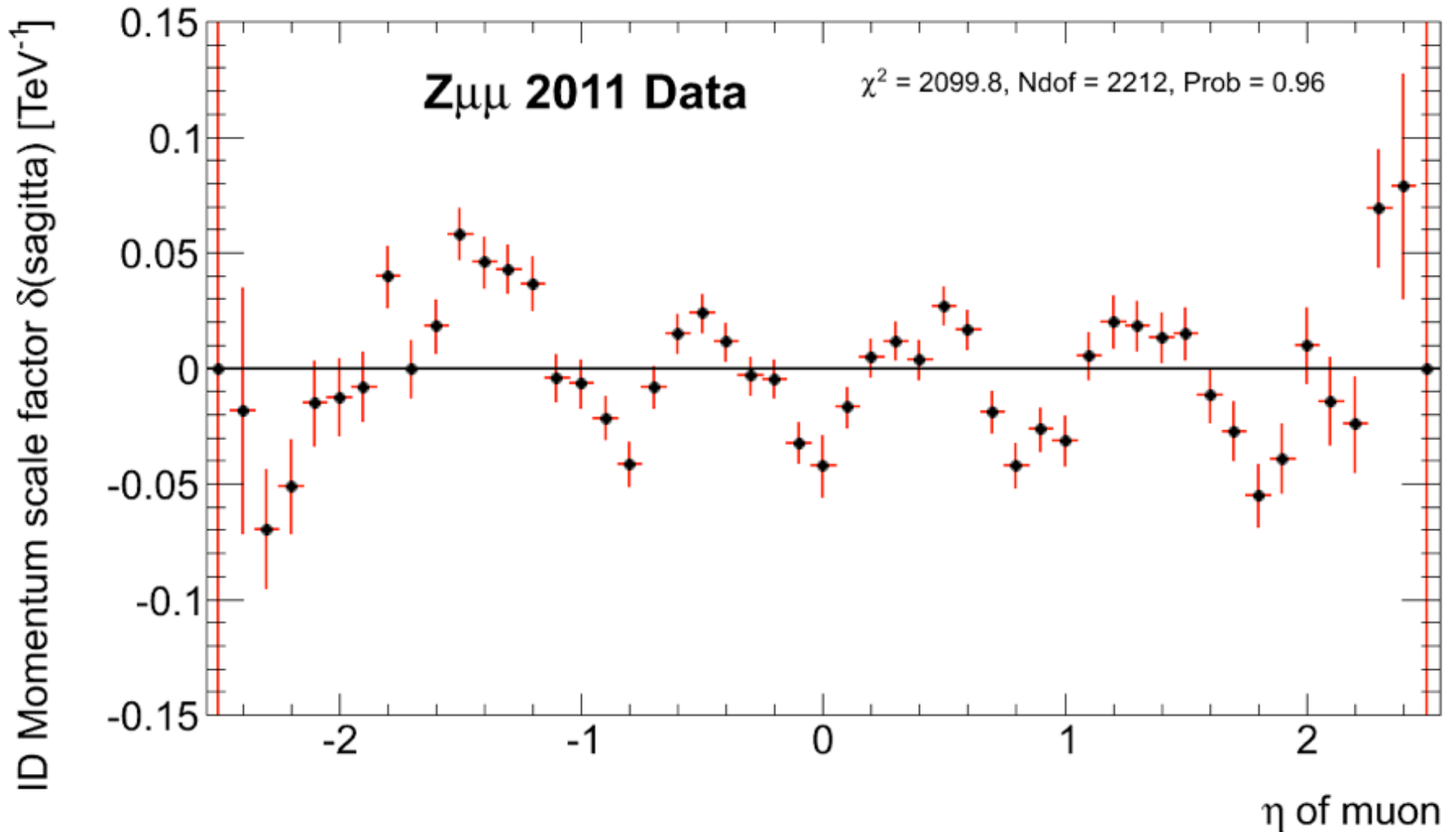
Result for Inner Detector

This is what out detector looks like for **Inner Detector Muons**:
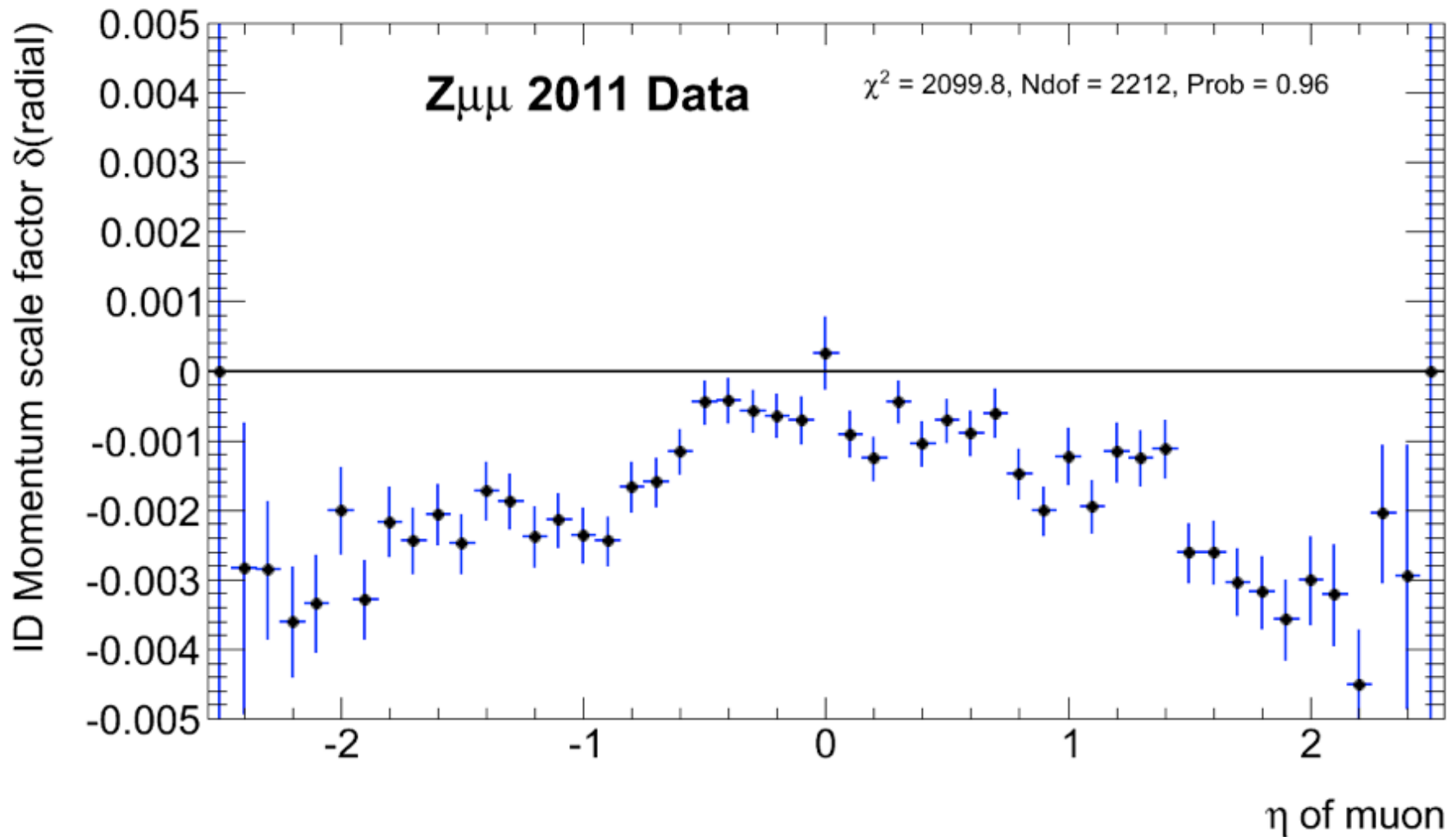
Result for Inner Detector

This is what out detector looks like for **Inner Detector Muons**:

Corrected result

Rerunning with the correction applied to MC, I get: