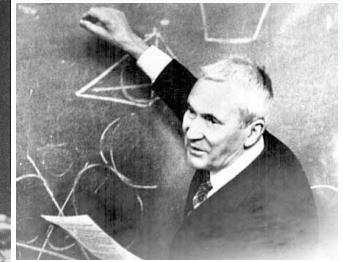
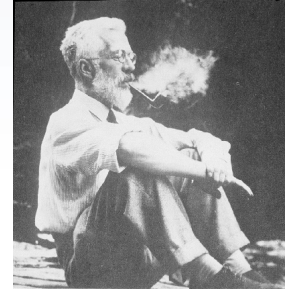
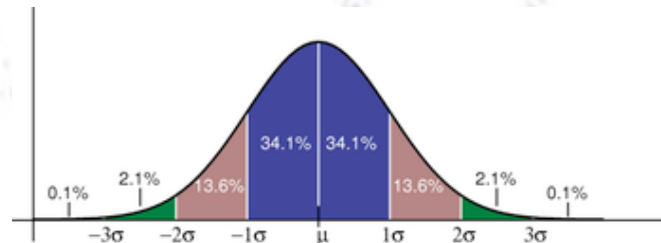


# Applied Statistics

## Advanced fitting



Troels C. Petersen (NBI)



*"Statistics is merely a quantisation of common sense"*

# Defining the Chi-Square

Problem Statement: Given  $N$  data points  $(x, y)$ , adjust the parameter(s)  $\theta$  of a model, such that it fits data best.

The best way to do this, given uncertainties  $\sigma_i$  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  $\chi^2 / \text{Ndof} \approx 1$  or more precisely if  $0.01 < p(\chi^2, \text{Ndof}) < 0.99$ , then all is good.
- If  $\chi^2 / \text{Ndof} \gg 1$  or more precisely if  $p(\chi^2, \text{Ndof}) < 0.01$ , then your fit is bad, and your hypothesis is probably not correct.
- If  $\chi^2 / \text{Ndof} \ll 1$  or more precisely if  $0.99 < p(\chi^2, \text{Ndof})$ , 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.

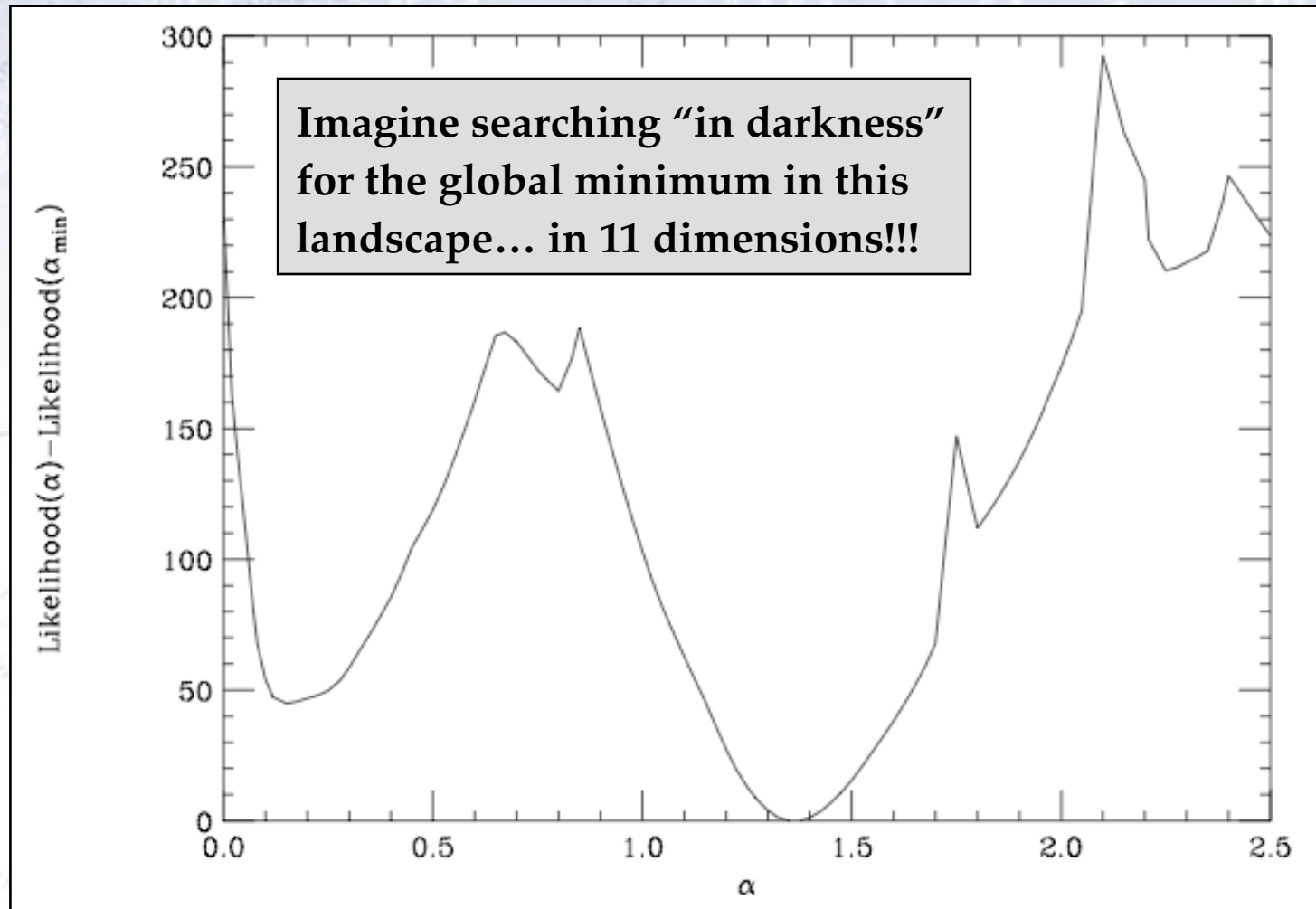
# Notes on the ChiSquare method

*“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 “landscape”

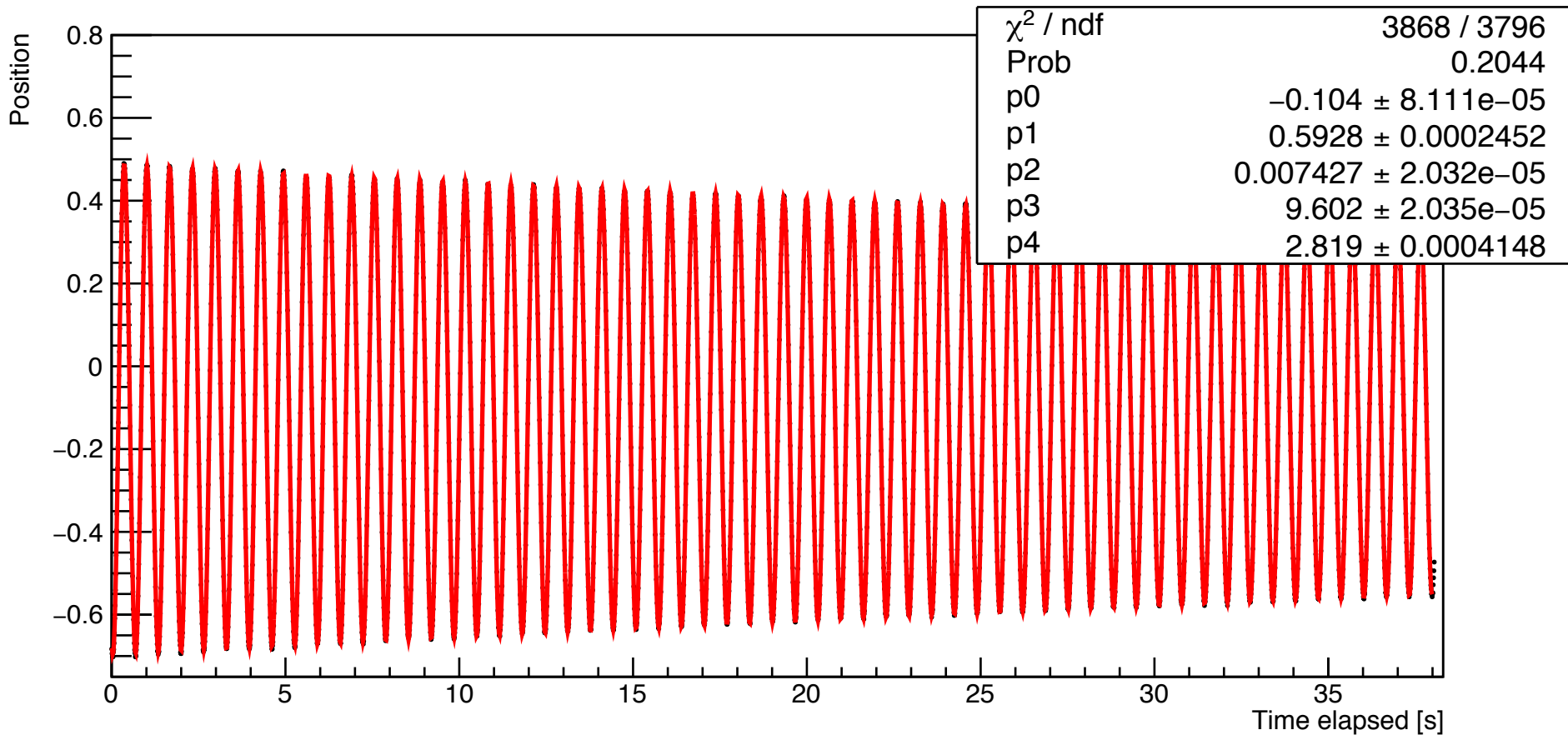


The fact that there are several minima makes fitting difficult/uncertain!

*Always give good starting values!!!*

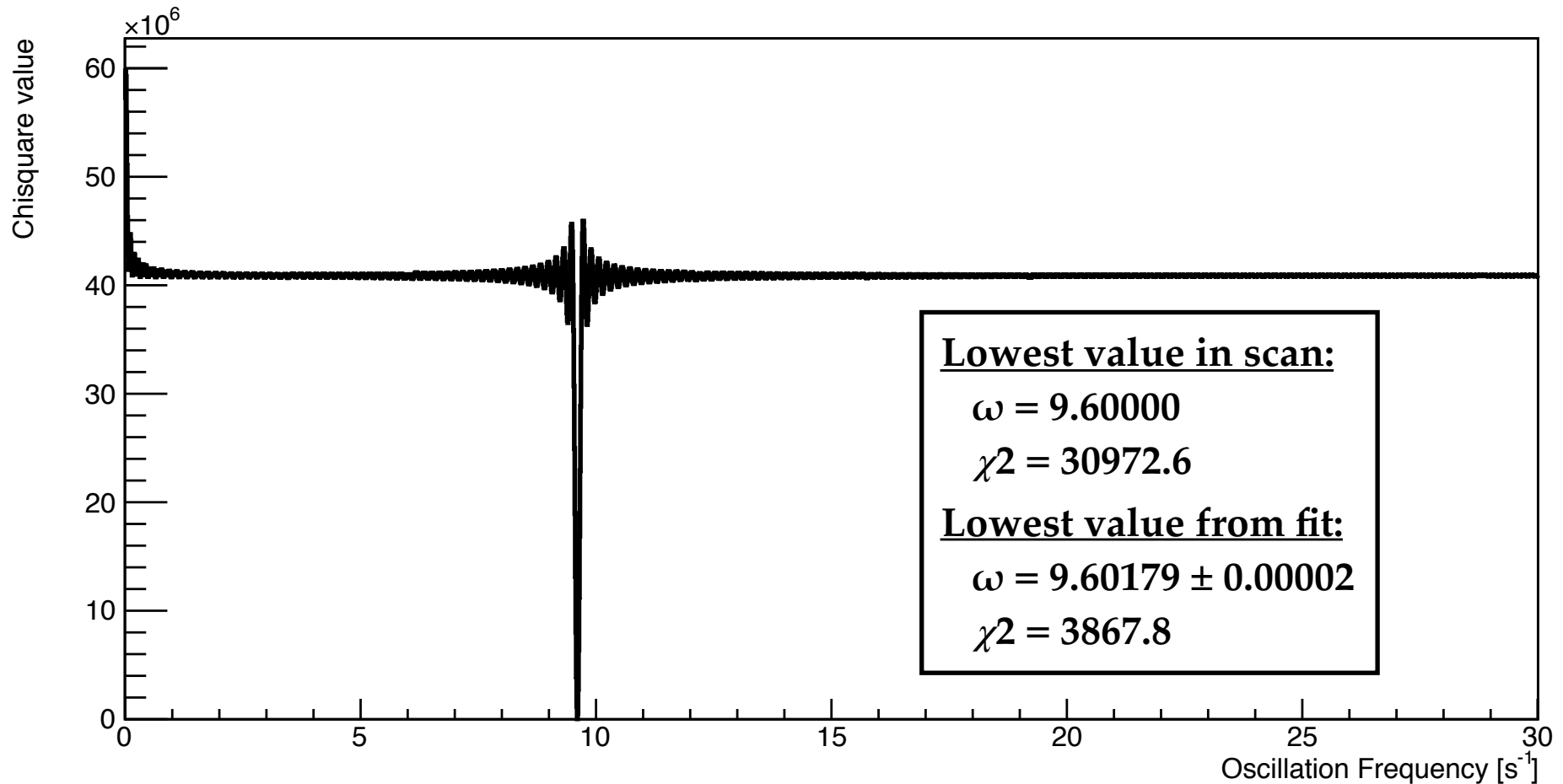
# Example of Chi-Square

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!



# 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!



# 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...

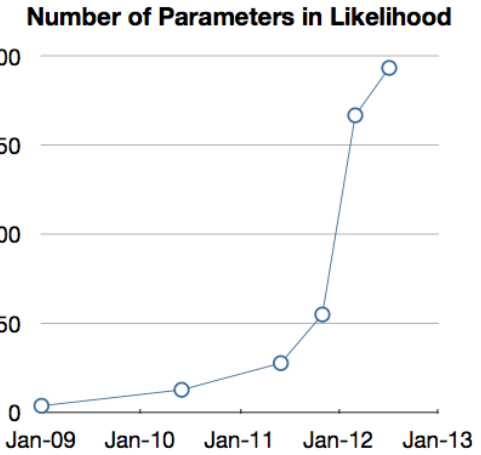
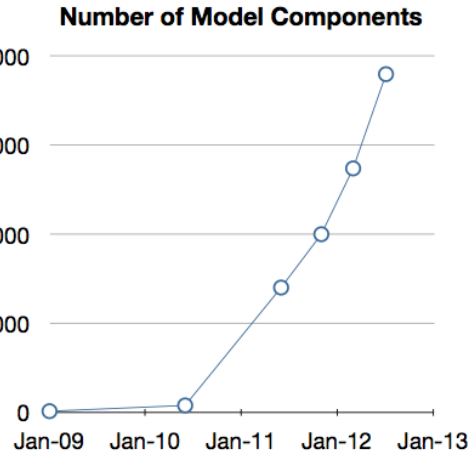
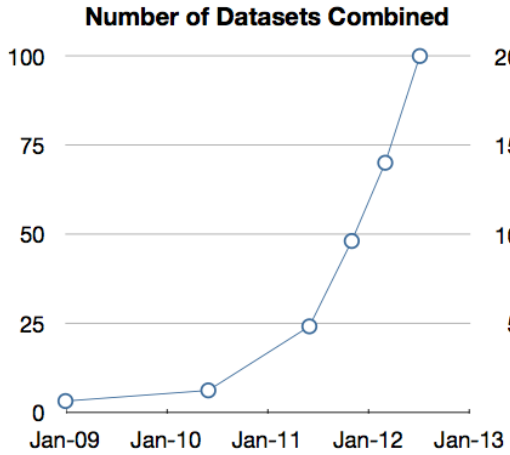
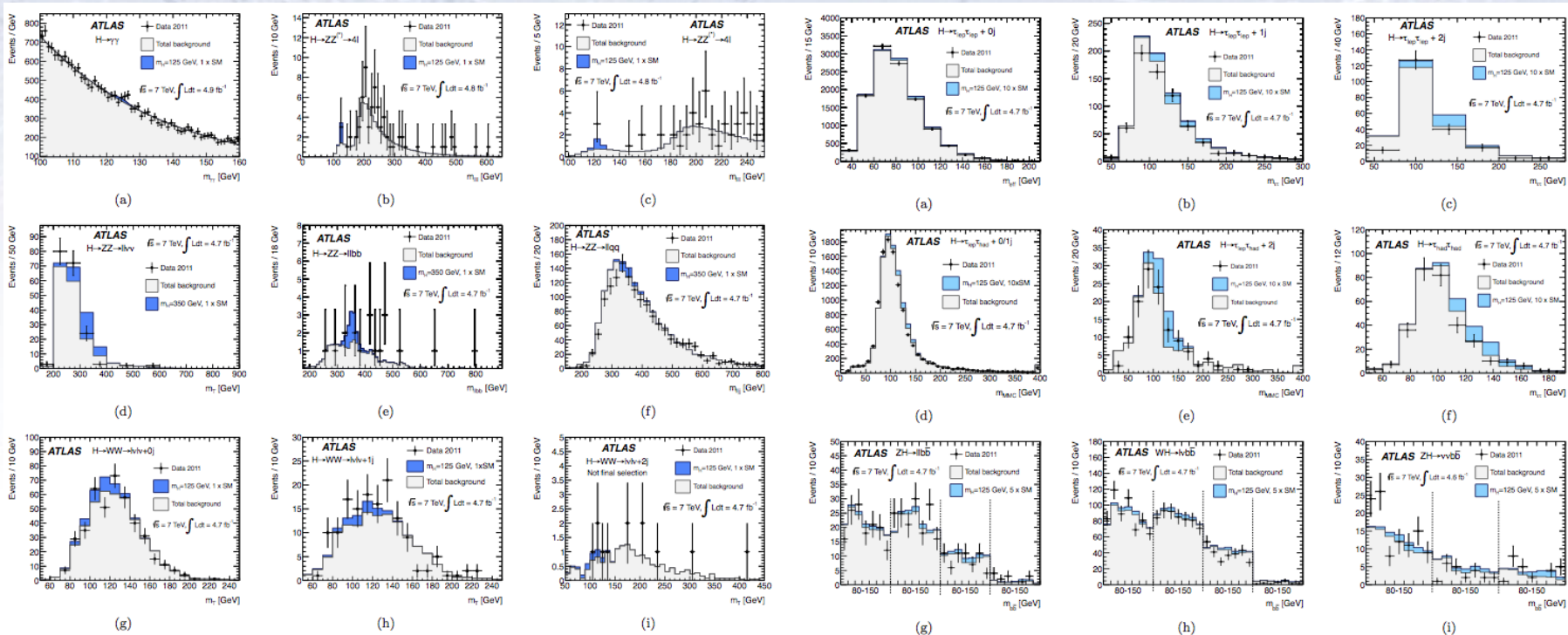
In 2002 the BaBar collaborations fit for CP-violation ( $\sin 2\beta$ ) included **98 floating parameters** applied to four datasets.

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 than LLH.

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...

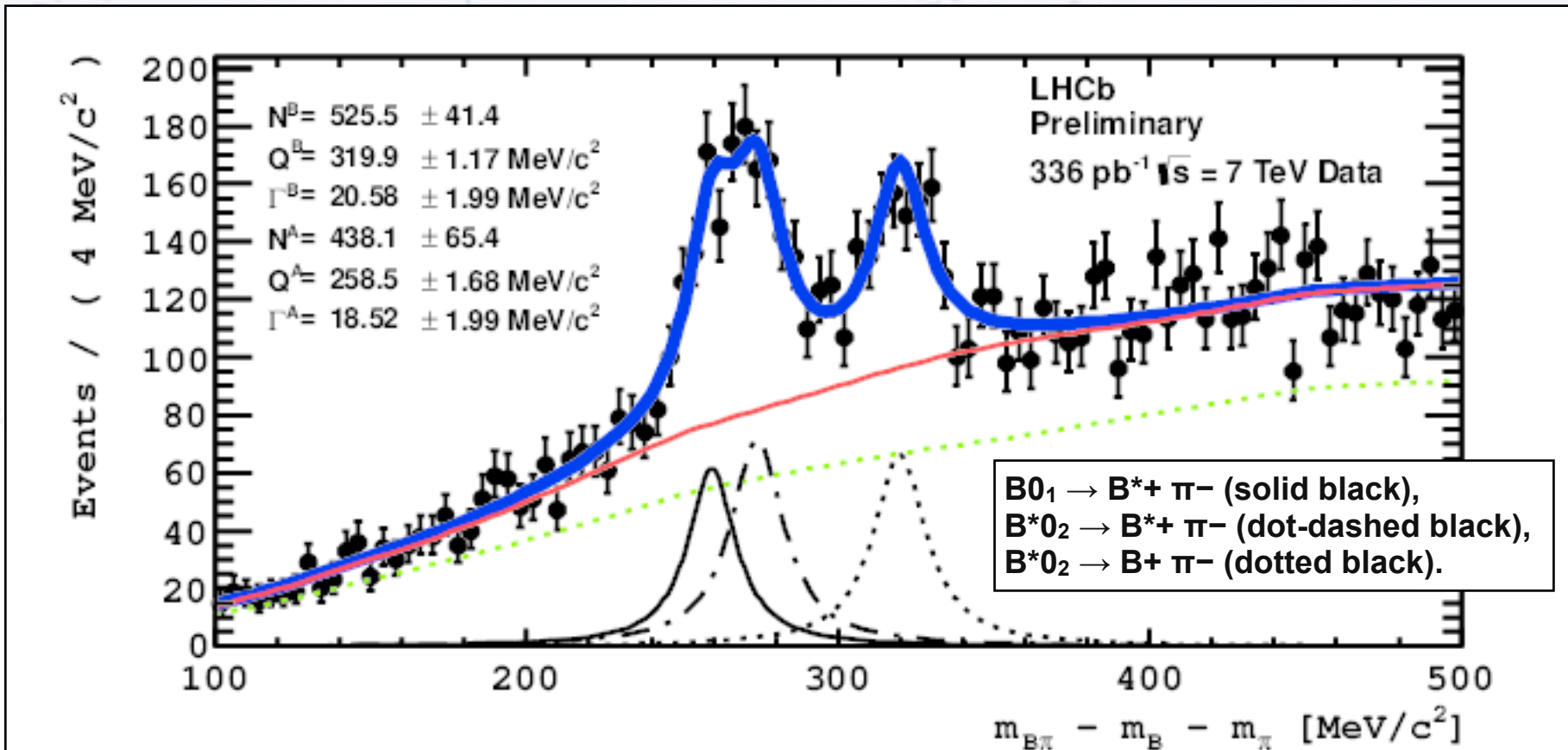
# Setting common parameters

For several reasons, one may decide to use common parameters:

- Parameters may be very correlated (e.g. multiple widths and lifetimes)
- Parameters may be expected to be identical (e.g. resolution or normalisation)

**A (well reasoned) parameter reduction is typically good!**

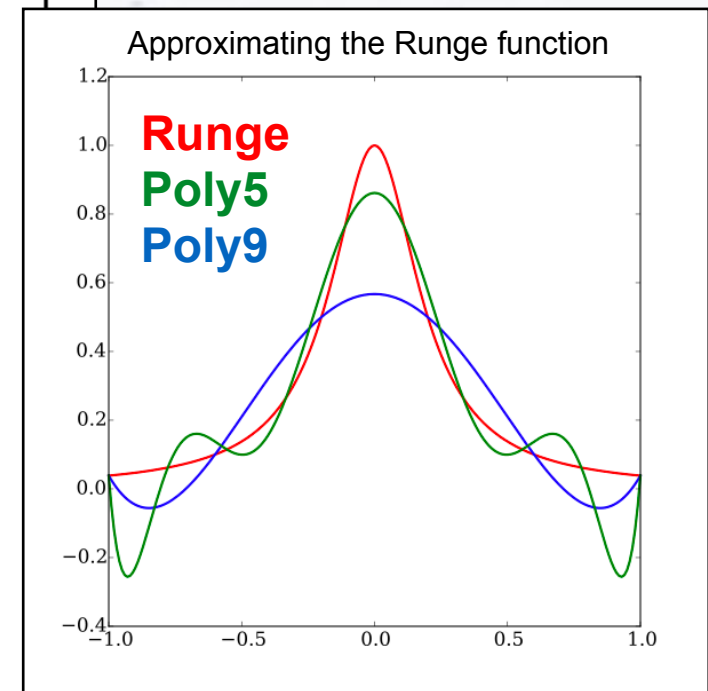
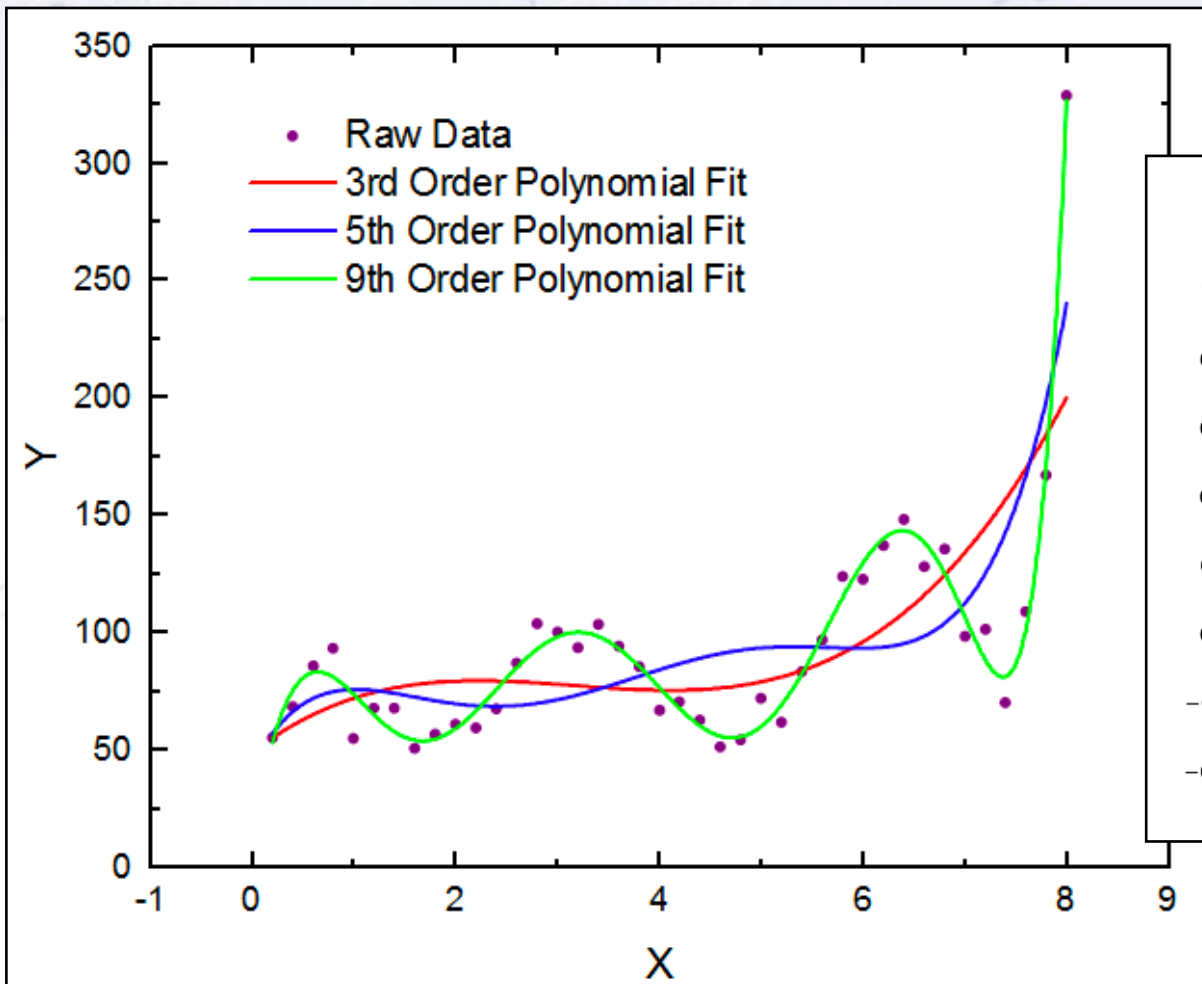
In the latter case, it may allow you to use a control/calibration channel in the fit.



# Fitting with polynomials

Polynomials can be good for fitting / approximating shapes, BUT...

- Coefficients can become heavily correlated (if not careful).
- Notoriously “wiggly” and with little / no resulting understanding.

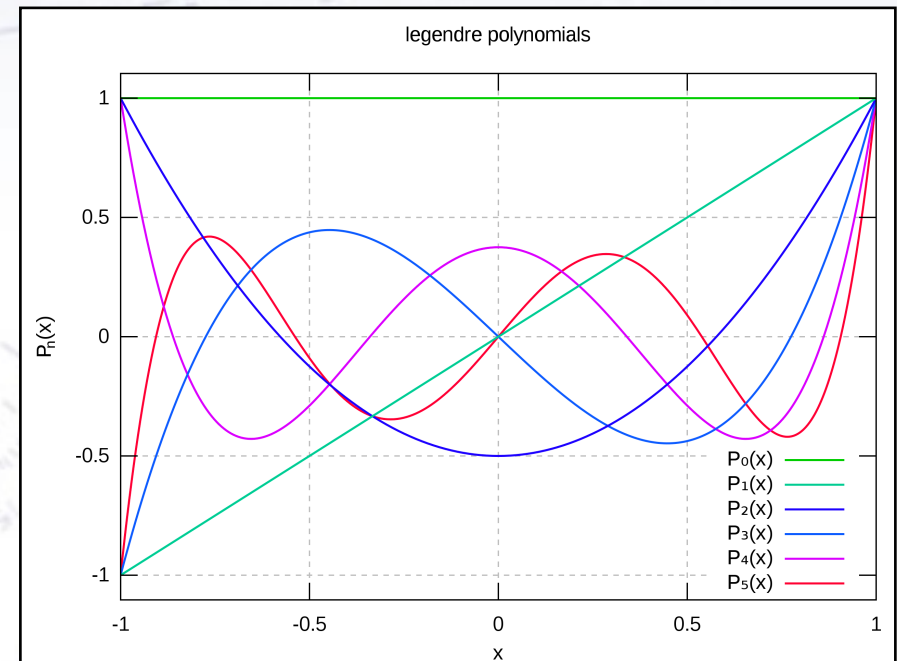


# Legendre & Chebyshev

To avoid correlations, it is very useful to “expand” polynomial around the relevant point/scale, e.g. fitting years (say 1990-2019).

It is very useful to use Legendre/Chebyshev polynomials for fitting, as these are designed to be orthogonal to each other, and hence minimise correlations:

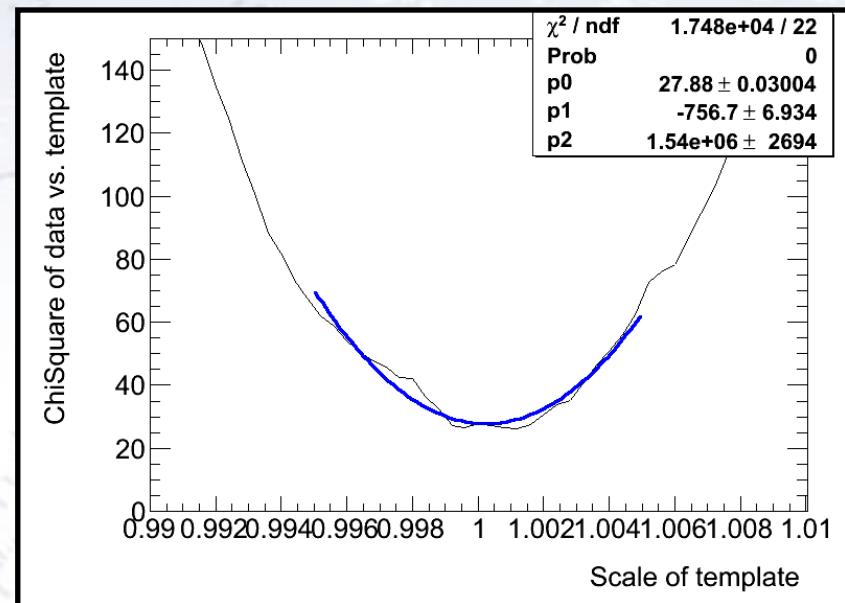
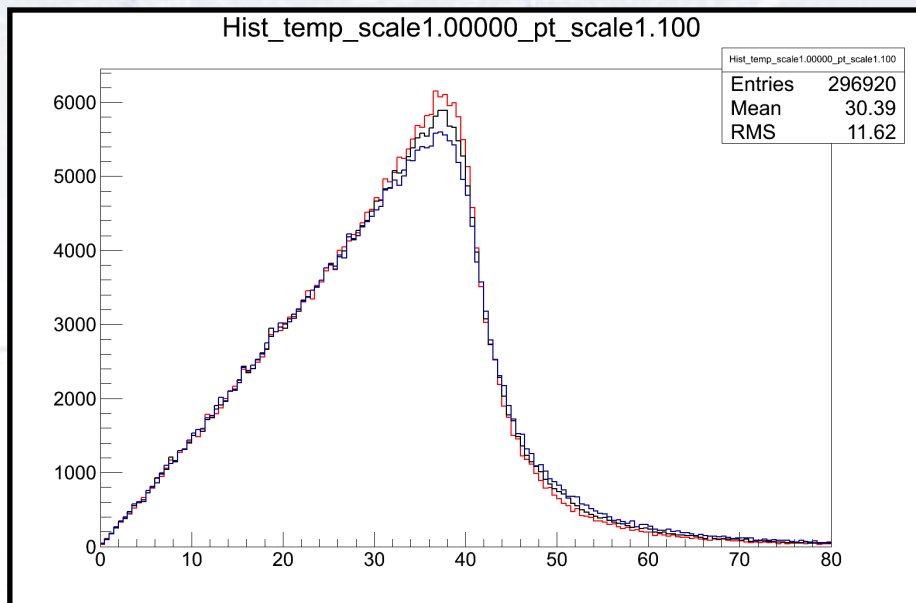
$n$	$P_n(x)$
0	1
1	$x$
2	$\frac{1}{2} (3x^2 - 1)$
3	$\frac{1}{2} (5x^3 - 3x)$
4	$\frac{1}{8} (35x^4 - 30x^2 + 3)$
5	$\frac{1}{8} (63x^5 - 70x^3 + 15x)$
6	$\frac{1}{16} (231x^6 - 315x^4 + 105x^2 - 5)$
7	$\frac{1}{16} (429x^7 - 693x^5 + 315x^3 - 35x)$
8	$\frac{1}{128} (6435x^8 - 12012x^6 + 6930x^4 - 1260x^2 + 35)$
9	$\frac{1}{128} (12155x^9 - 25740x^7 + 18018x^5 - 4620x^3 + 315x)$
10	$\frac{1}{256} (46189x^{10} - 109395x^8 + 90090x^6 - 30030x^4 + 3465x^2 - 63)$



An alternative is to use **Chebyshev polynomials**, which have many of the same “desired” properties, and minimise “Runge’s Phenomenon”, i.e. wigglyness.

# 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).



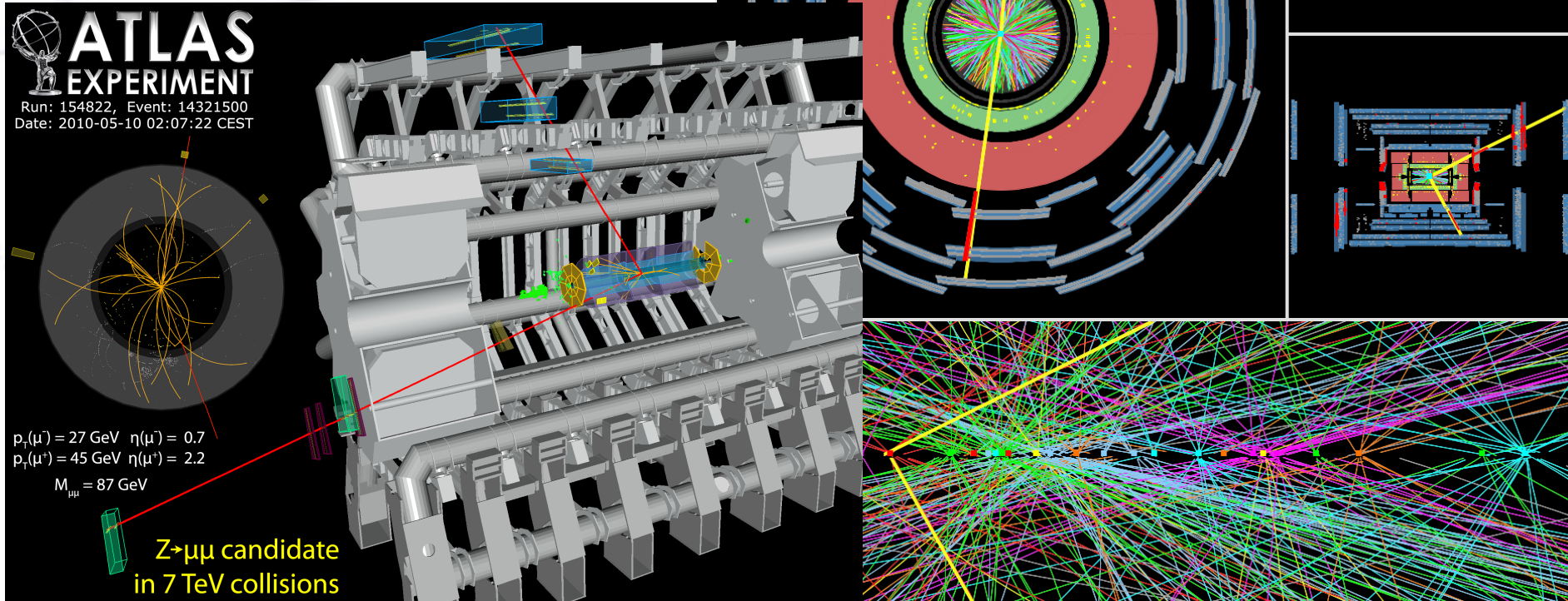
# Bonus Slides



# Setting the momentum scale

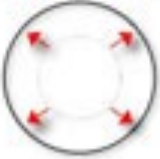







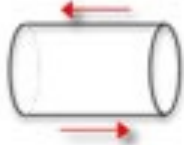
A well known type of event in particle physics (at the LHC accelerator) is the  $Z \rightarrow \mu^+ \mu^-$  decay, which has two muons flying out through the detector.

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



# Setting the momentum scale

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

	$\Delta R$	$\Delta\phi$	$\Delta Z$
R	<b>Radial Expansion</b> (distance scale) 	<b>Curl</b> (Charge asymmetry) 	<b>Telescope</b> (COM boost) 
$\phi$	<b>Elliptical</b> (vertex mass) 	<b>Clamshell</b> (vertex displacement) 	<b>Skew</b> (COM energy) 
Z	<b>Bowing</b> (COM energy) 	<b>Twist</b> (CP violation) 	<b>Z expansion</b> (distance scale) 

# Charge dependent fit

The approach follows ATLAS-CONF-2012-141:

There is a charge independent (radial):

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

The charge and  $p_T$  dependent (sagitta):

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

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  $\delta_r$  and 24/51 for  $\delta_s$ .

# Introduction to idea

In order to use all  $Z \rightarrow \ell\ell$  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  $\alpha_{ij}$ :

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

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:

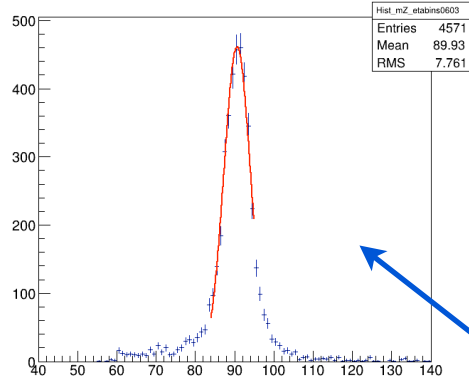
$$\chi^2(\beta_i) = \sum_{ij} \left( \frac{\alpha_{ij} - \sqrt{\beta_i \beta_j}}{\sigma(\alpha_{ij})} \right)^2$$

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

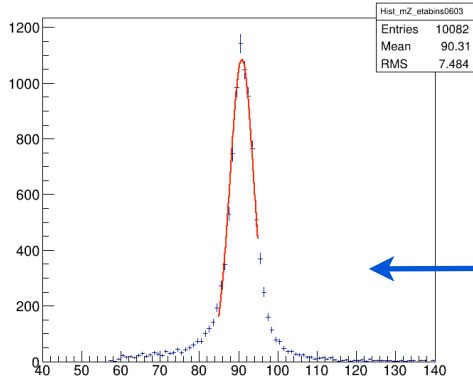
# Outline of the method

For each bin in  $\eta+\eta-$ ,  
fit the mass peak:

**Data:  $m_Z = 89.93 \pm 0.13$  GeV**

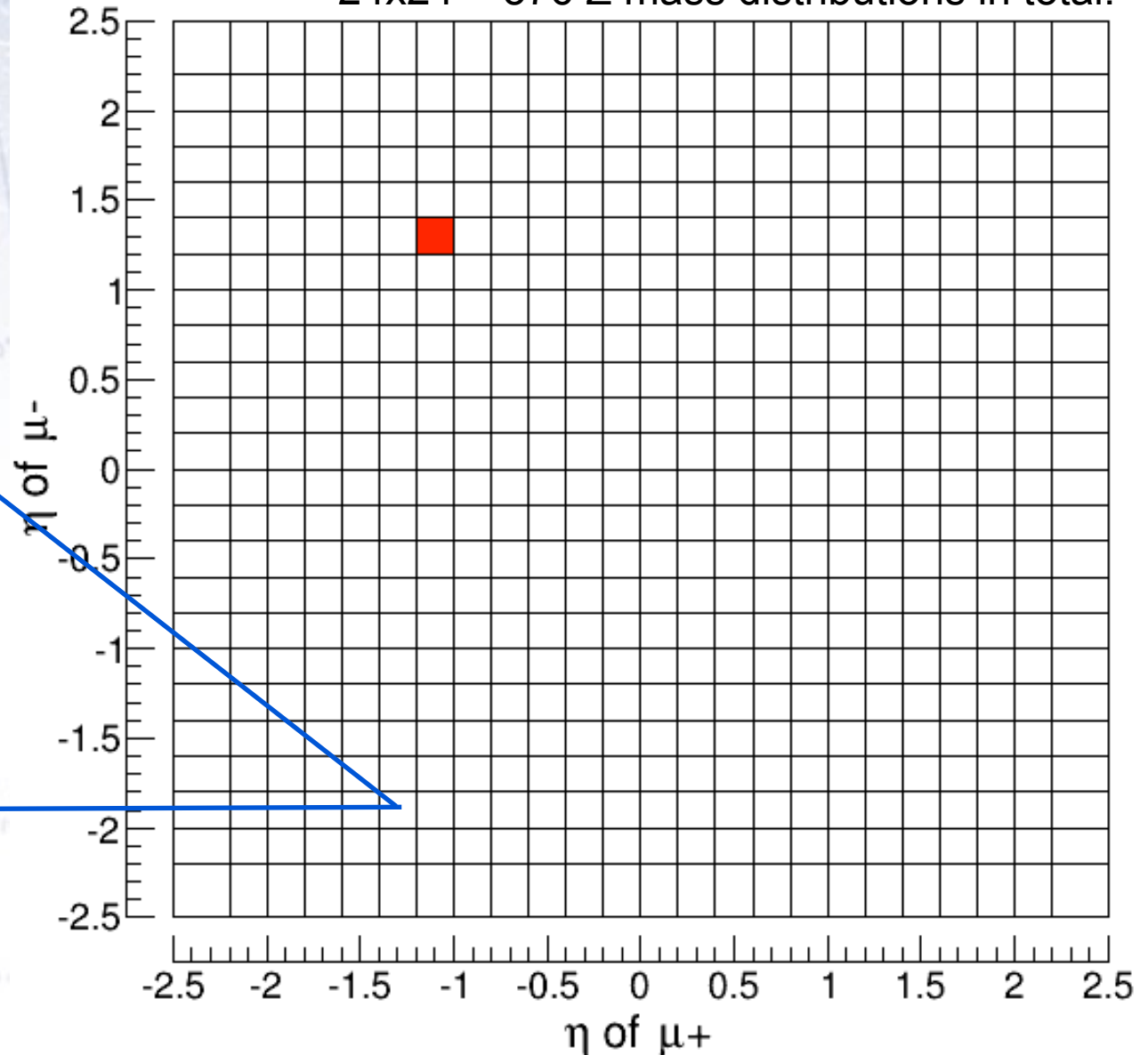


**MC:  $m_Z = 90.31 \pm 0.09$  GeV**



**Ratio: =  $0.9971 \pm 0.0009$  GeV**

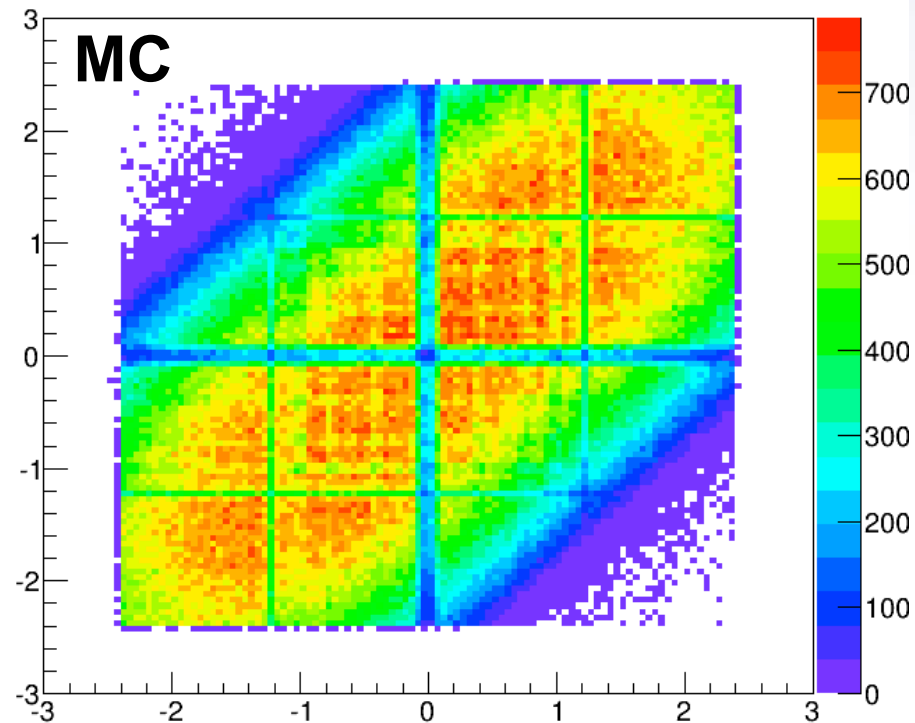
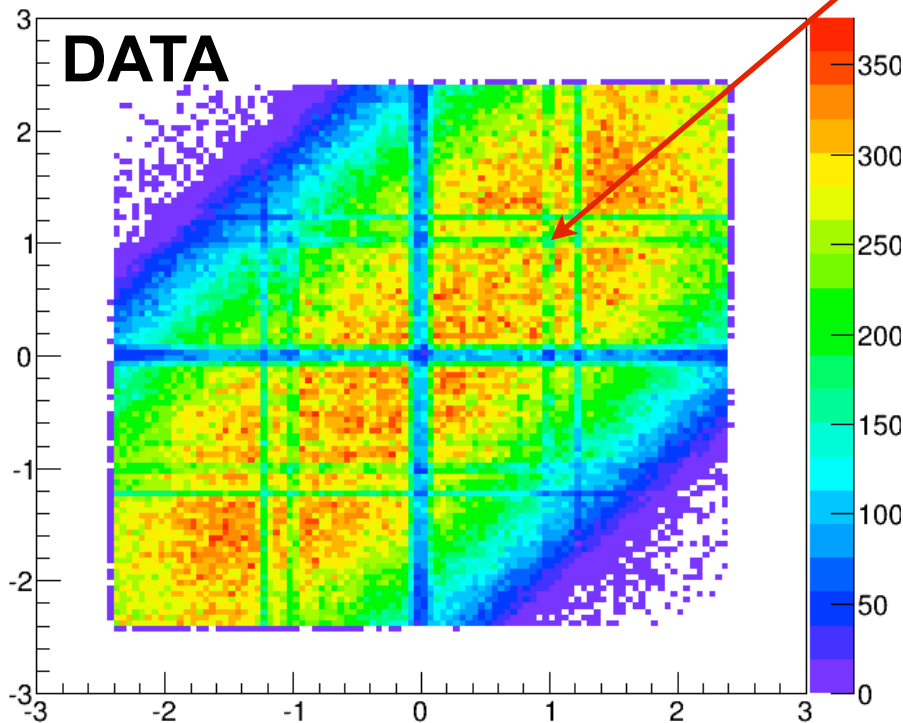
24x24 = 576 Z mass distributions in total!



# 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:

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

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

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

These values are used in a  $\chi^2$  as follows:

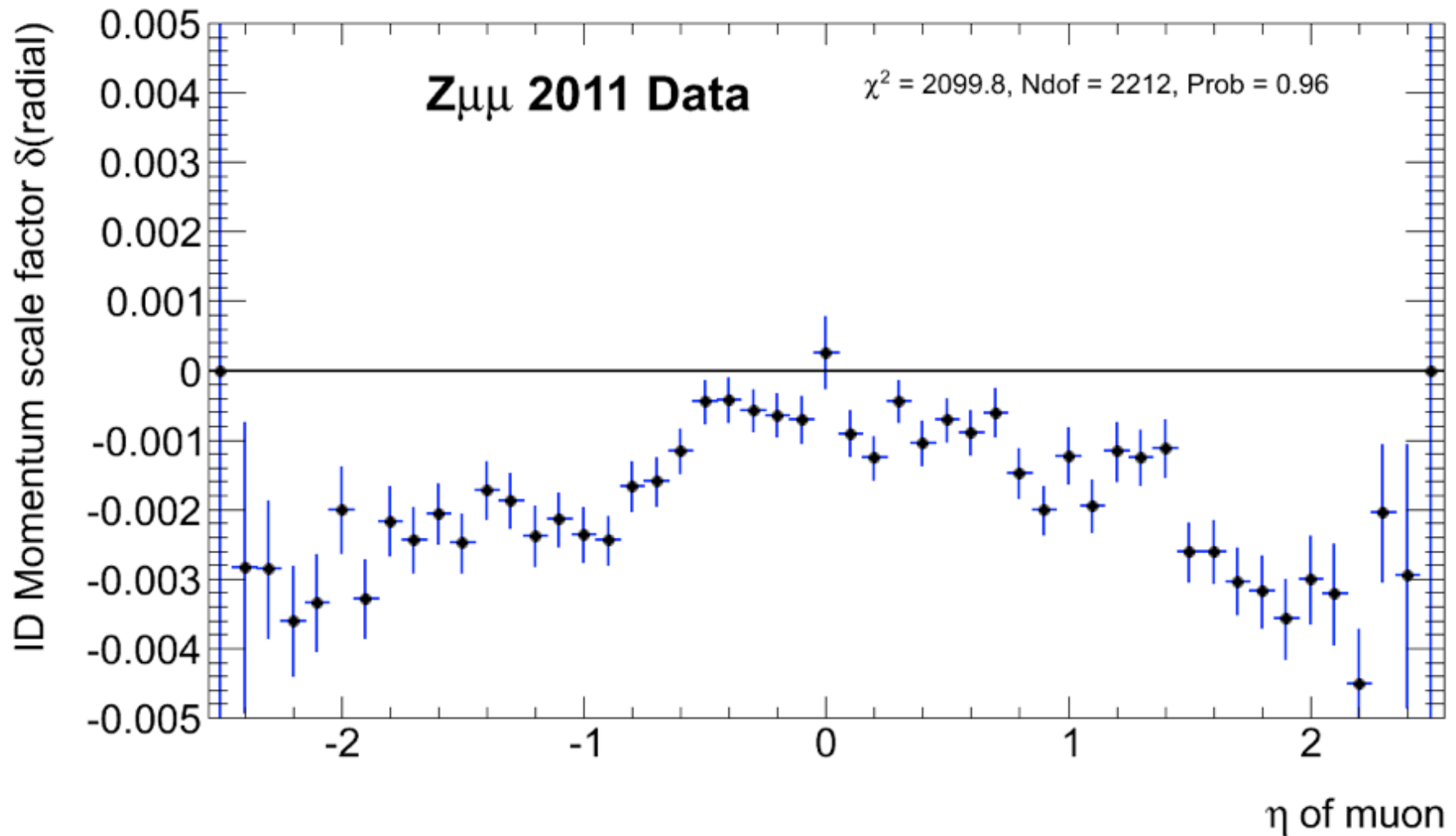
$$\chi^2 = \sum_{ij} \left( \frac{R(m_Z)_{ij} - \sqrt{(1 + \delta r_i)(1 + p_T \delta s_i)} \times (1 + \delta r_j)(1 - p_T \delta s_j)}{\sigma(R(m_Z)_{ij})} \right)^2$$

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

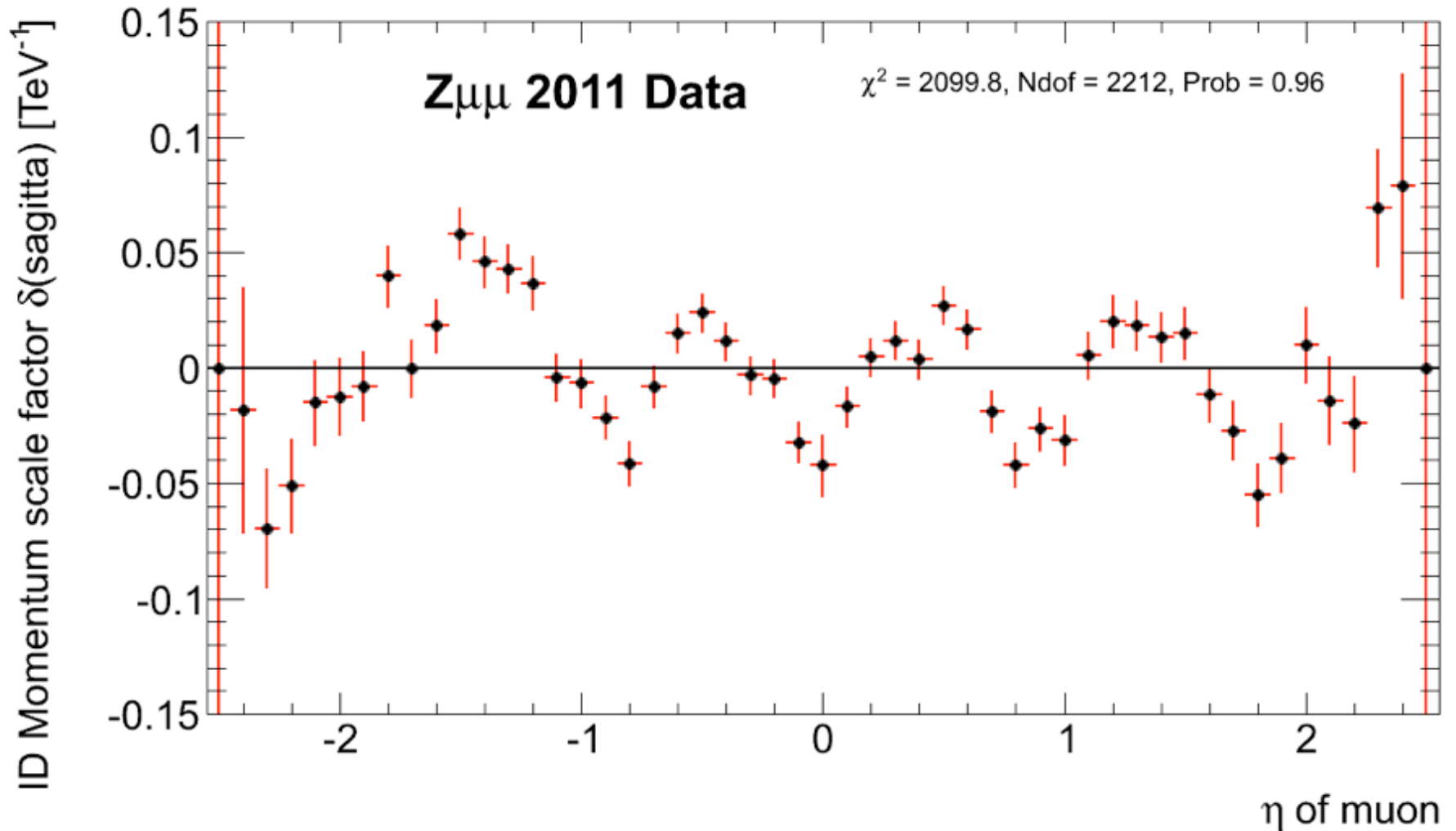
This is what out detector looks like for Inner Detector Muons:





# Result for Inner Detector

This is what our detector looks like for Inner Detector Muons:



# Corrected result

Rerunning with the correction applied to MC, I get:

