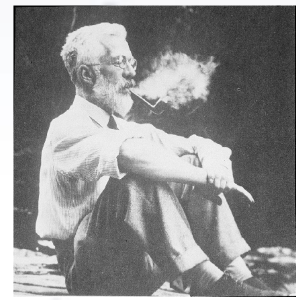
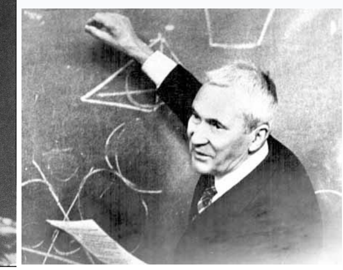
Applied Statistics

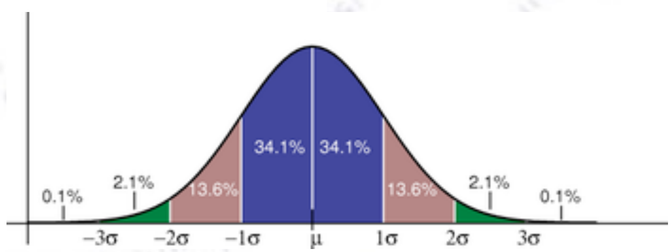
Measuring the length of a Table...





Troels C. Petersen (NBI)



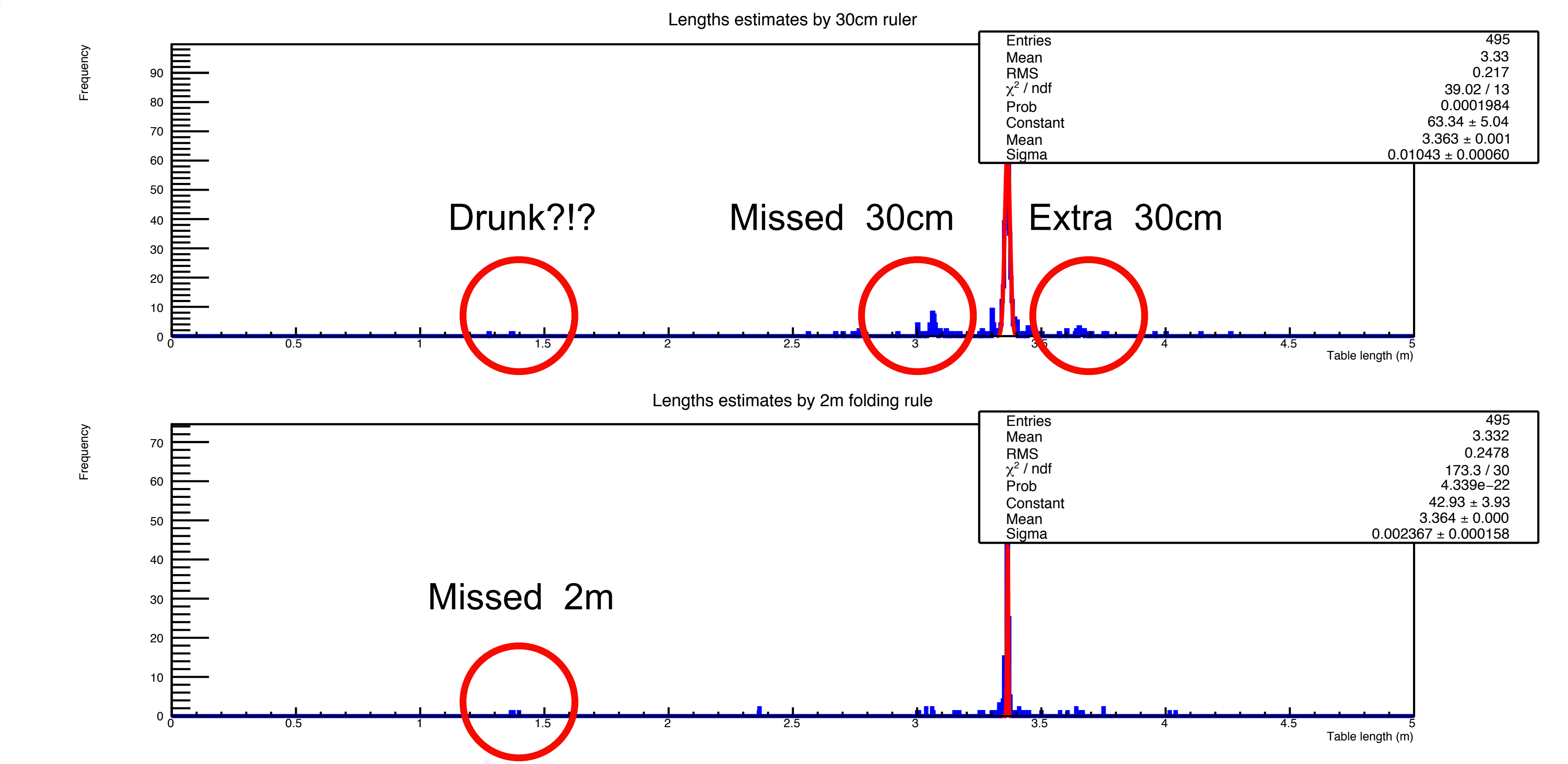
"Statistics is merely a quantisation of common sense"

My analysis

The table measurement data

The initial dataset contains (valid measurements):

- 30cm measurements: **495** Range: [0.0, 5.0] m
- 2m measurements: **493** Range: [0.0, 5.0] m



Raw ("Naive") results

<u>30cm:</u>

Mean = 3.3290 ± 0.0098 m

RMS = 0.22 m (N = 495)

<u>2m:</u>

Mean = 3.3320 ± 0.0112 m

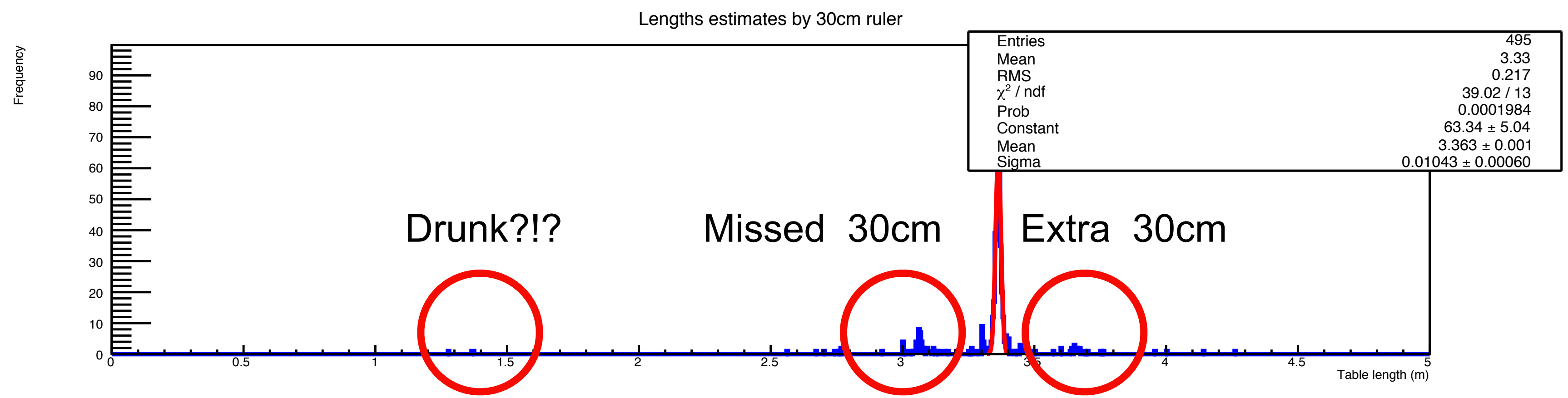
RMS = 0.25 m (N = 493)

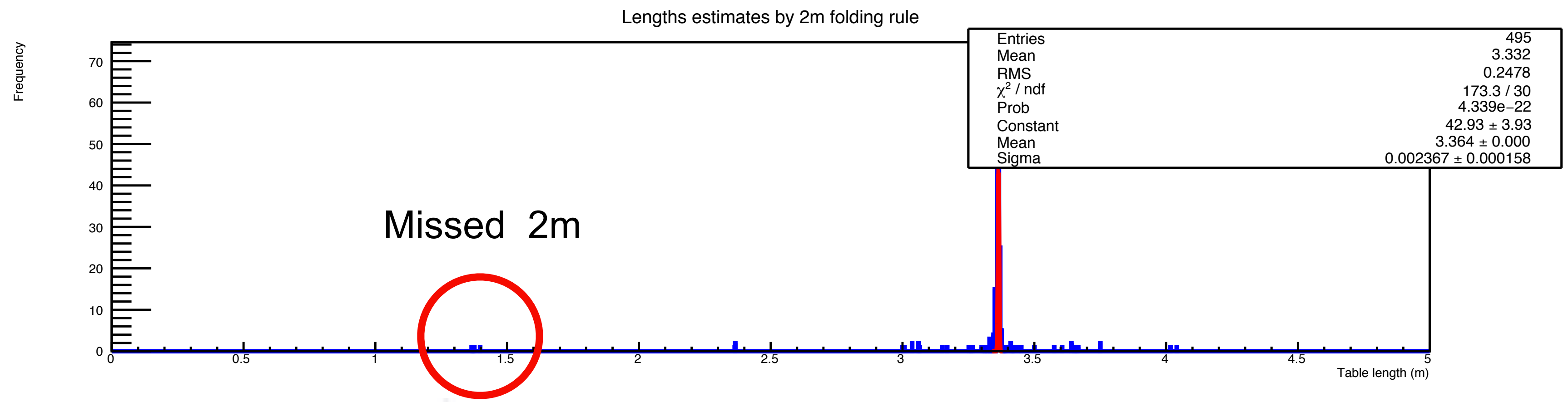
<u>Correspondence between 30cm and 2m measurement:</u>

Diff = -0.0030 ± 0.0148 m (-0.20 σ)

Include offsets?

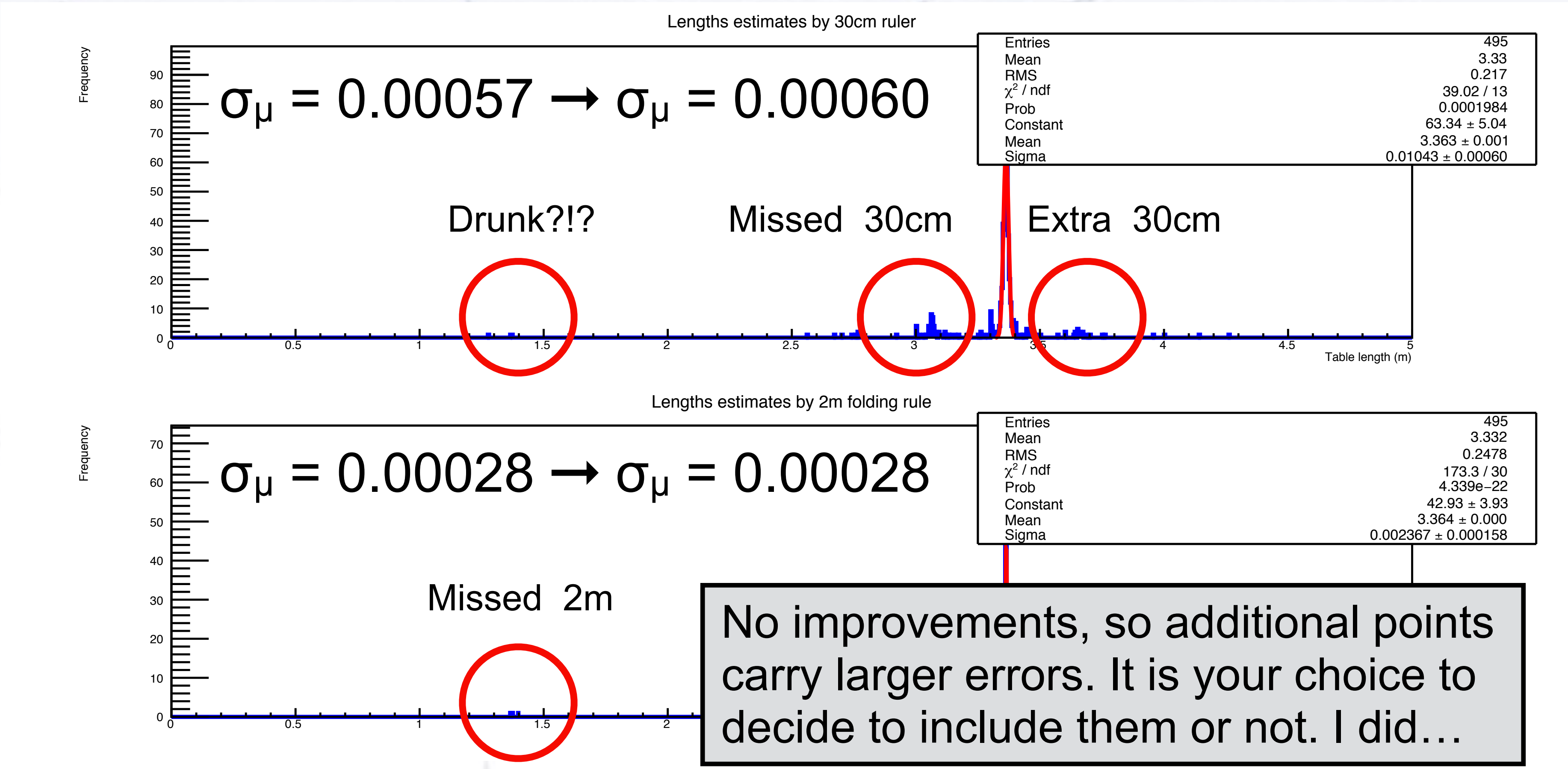
There are some clear and understandable mis-measurements.
Should one correct and include these?
Depends on resulting improvement, but decide without seeing the final result.





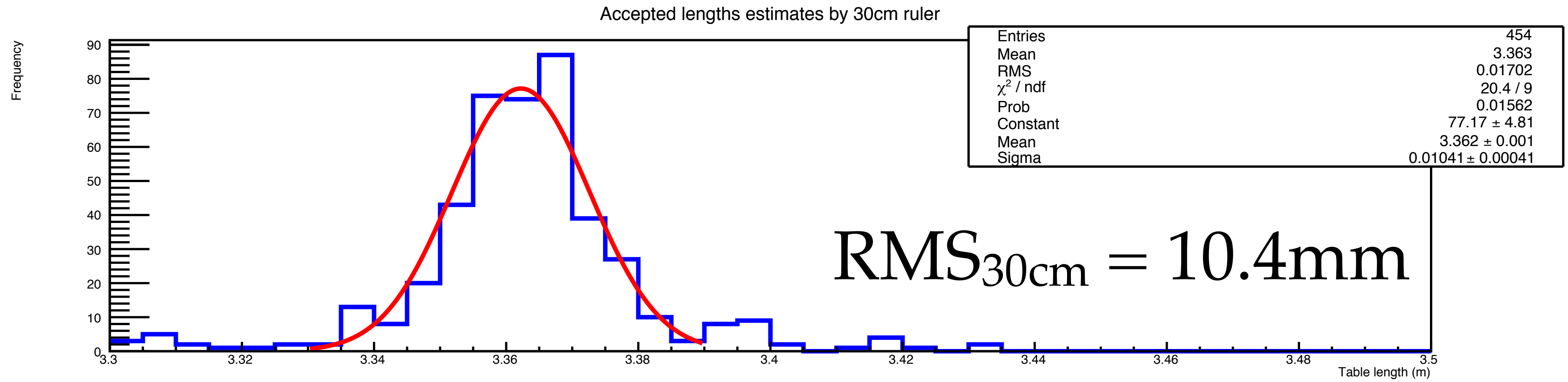
Include offsets?

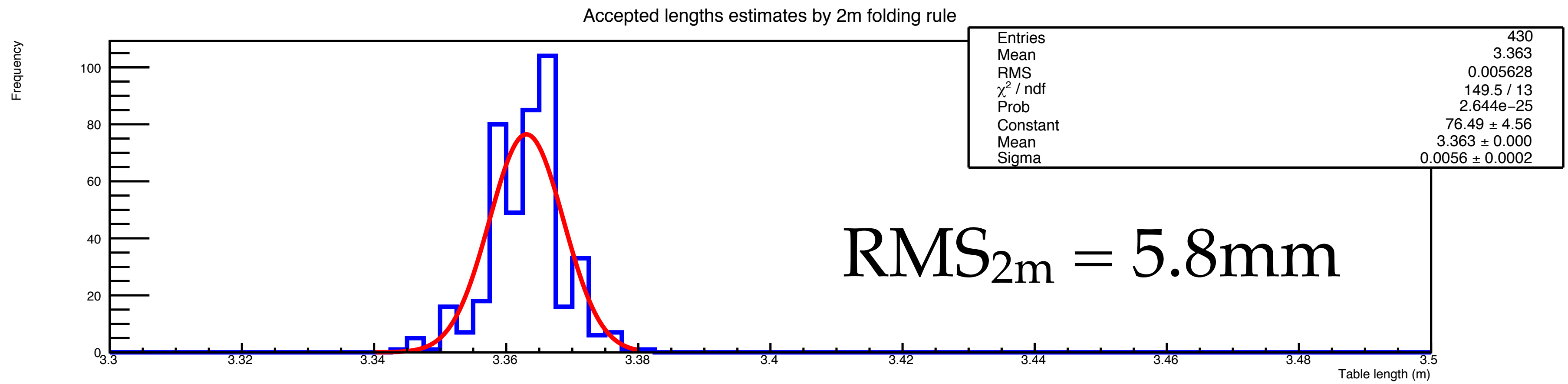
There are some clear and understandable mis-measurements.
Should one correct and include these?
Depends on resulting improvement, but decide without seeing the final result.



Inspecting the data

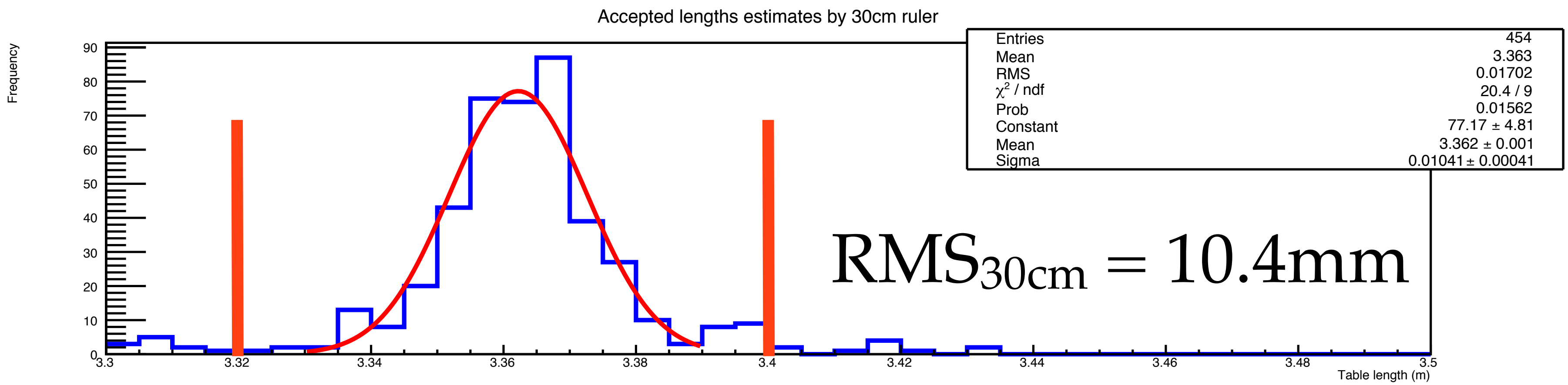
The 30cm peak seems Gaussian (p=2.4%) with binning 0.005 (smaller gives peaks).
The 2m peak does not seem Gaussian with any binning (here 0.0025), yet "collected".

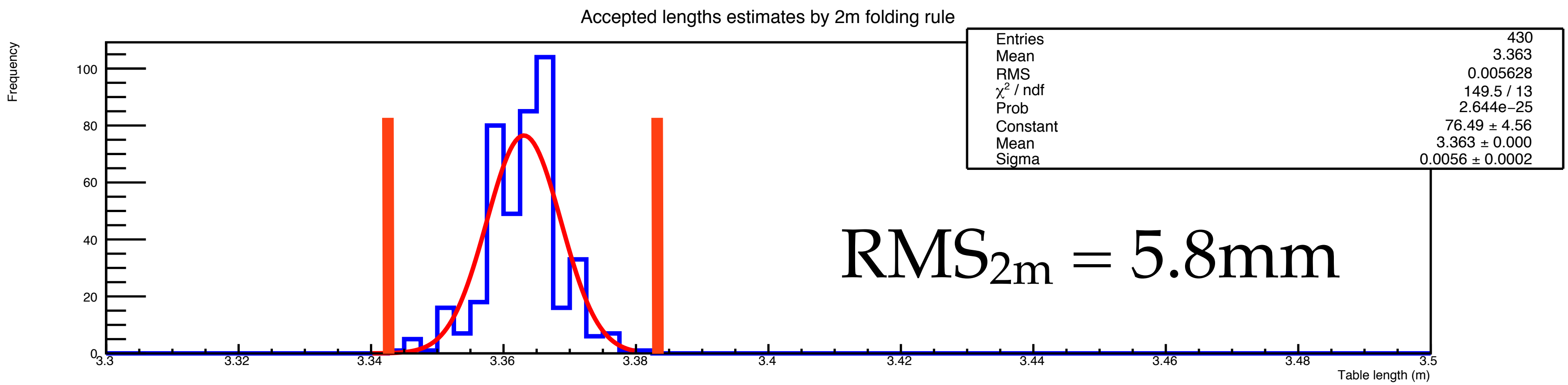




Inspecting the data

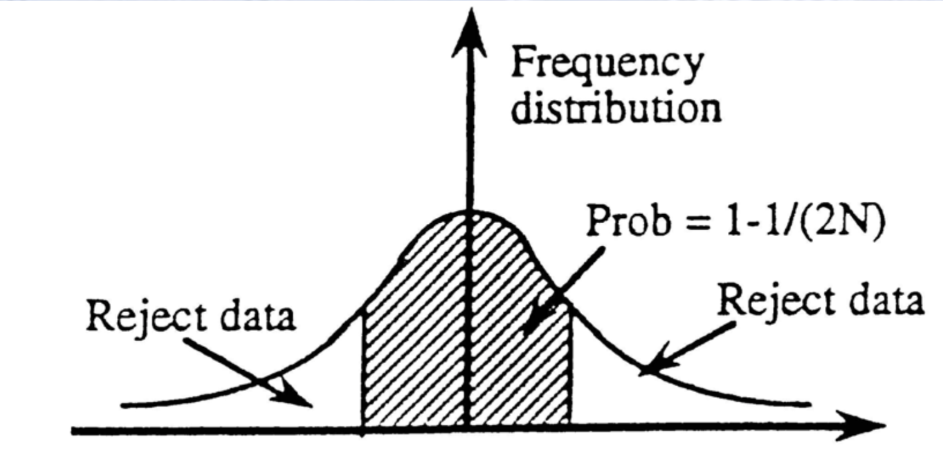
There are clearly some **mis-measurements**, which we would like to **exclude**. Using the fitted width, and accepting that this only includes the best measurements, I could **decide** to include measurements within **4 × RMS**:





Removing data points

Removing improbable data points is formalised in **Chauvenet's Criterion,** though many other methods exists (see Peirce, Grubbs, etc.)



The idea is to assume that the distribution is Gaussian, and ask what the probability of the farthest point is. If it is below some value, which is to be determined ahead of applying the criterion, then the point is removed, and the criterion is reapplied until no more points should be removed.

I choose to say, that if the outermost point in the Gaussian case has **less than 5% chance of being this far out** (taking the total number of points into account), then I reject it.

However, **ALWAYS keep a record of your original data**, as it may contain more effects than you originally thought.

…a fair hearing?

```
Rejected 30cm measurements:
 304: L=1.275  dL=2.071  Nsig=10.89   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=495  mean=3.3460  rms=0.1901   -> Rejected
 368: L=1.365  dL=1.985  Nsig=11.97   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=494  mean=3.3501  rms=0.1659   -> Rejected
  42: L=1.370  dL=1.984  Nsig=14.19   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=493  mean=3.3542  rms=0.1398   -> Rejected
  44: L=4.260  dL=0.902  Nsig= 8.39   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=492  mean=3.3582  rms=0.1075   -> Rejected
 146: L=2.563  dL=0.793  Nsig= 7.97   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=491  mean=3.3564  rms=0.0995   -> Rejected
 158: L=4.140  dL=0.782  Nsig= 8.41   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=490  mean=3.3580  rms=0.0929   -> Rejected
 143: L=2.670  dL=0.686  Nsig= 7.98   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=489  mean=3.3564  rms=0.0860   -> Rejected
 477: L=2.700  dL=0.658  Nsig= 8.19   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=488  mean=3.3578  rms=0.0803   -> Rejected
 308: L=4.004  dL=0.645  Nsig= 8.64   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=487  mean=3.3591  rms=0.0746   -> Rejected
 427: L=2.744  dL=0.614  Nsig= 8.93   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=486  mean=3.3578  rms=0.0687   -> Rejected
 407: L=2.759  dL=0.600  Nsig= 9.55   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=485  mean=3.3591  rms=0.0628   -> Rejected
 313: L=3.955  dL=0.595  Nsig=10.49   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=484  mean=3.3603  rms=0.0567   -> Rejected
 269: L=2.768  dL=0.591  Nsig=11.86   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=483  mean=3.3591  rms=0.0498   -> Rejected
 380: L=2.769  dL=0.591  Nsig=14.10   p1=0.00000000, p_all=0.00000000 >? pmin=0.050   N=482  mean=3.3603  rms=0.0419   -> Rejected
 133: L=3.180  dL=0.182  Nsig= 5.65   ... an=3.3616  rms=0.0321   -> Rejected
  71: L=3.200  dL=0.162  Nsig= 5.21   ... an=3.3619  rms=0.0311   -> Rejected
 320: L=3.205  dL=0.157  Nsig= 5.20   ... an=3.3623  rms=0.0302   -> Rejected
 319: L=3.215  dL=0.148  Nsig= 5.02   ... an=3.3626  rms=0.0294   -> Rejected
 194: L=3.221  dL=0.142  Nsig= 4.96   ... an=3.3629  rms=0.0286   -> Rejected
   1: L=3.250  dL=0.113  Nsig= 4.06   p ... n=3.3632  rms=0.0279   -> Rejected
 175: L=3.471  dL=0.108  Nsig= 3.92   ... an=3.3634  rms=0.0274   -> Rejected
 133: L=3.470  dL=0.107  Nsig= 3.95   ... an=3.3632  rms=0.0270   -> Rejected
  13: L=3.470  dL=0.107  Nsig= 4.02   p1=0.00002883, p_all=0.01354566 >? pmin=0.050   N=473  mean=3.3630  rms=0.0266   -> Rejected
 154: L=3.467  dL=0.104  Nsig= 3.98   p1=0.00003406, p_all=0.01594736 >? pmin=0.050   N=472  mean=3.3628  rms=0.0262   -> Rejected
 130: L=3.260  dL=0.103  Nsig= 3.98   p1=0.00003423, p_all=0.01599227 >? pmin=0.050   N=471  mean=3.3625  rms=0.0258   -> Rejected
 230: L=3.260  dL=0.103  Nsig= 4.05   p1=0.00002507, p_all=0.01171573 >? pmin=0.050   N=470  mean=3.3628  rms=0.0253   -> Rejected
 444: L=3.461  dL=0.098  Nsig= 3.93   p1=0.00004185, p_all=0.01943772 >? pmin=0.050   N=469  mean=3.3630  rms=0.0249   -> Rejected
  79: L=3.460  dL=0.097  Nsig= 3.96   p1=0.00003687, p_all=0.01710909 >? pmin=0.050   N=468  mean=3.3628  rms=0.0245   -> Rejected
  43: L=3.455  dL=0.092  Nsig= 3.83   p1=0.00006418, p_all=0.02952774 >? pmin=0.050   N=467  mean=3.3626  rms=0.0241   -> Rejected
 113: L=3.455  dL=0.093  Nsig= 3.90   p1=0.00004898, p_all=0.02256478 >? pmin=0.050   N=466  mean=3.3624  rms=0.0238   -> Rejected
 433: L=3.272  dL=0.090  Nsig= 3.85   p1=0.00005881, p_all=0.02697886 >? pmin=0.050   N=465  mean=3.3622  rms=0.0234   -> Rejected
  81: L=3.450  dL=0.088  Nsig= 3.80   p1=0.00007216, p_all=0.03292944 >? pmin=0.050   N=464  mean=3.3624  rms=0.0231   -> Rejected
 345: L=3.450  dL=0.088  Nsig= 3.87   p1=0.00005540, p_all=0.02532365 >? pmin=0.050   N=463  mean=3.3622  rms=0.0227   -> Rejected
 351: L=3.450  dL=0.088  Nsig= 3.93   p1=0.00004175, p_all=0.01910567 >? pmin=0.050   N=462  mean=3.3620  rms=0.0224   -> Rejected
 333: L=3.450  dL=0.088  Nsig= 4.01   p1=0.00003083, p_all=0.01411413 >? pmin=0.050   N=461  mean=3.3618  rms=0.0220   -> Rejected
 324: L=3.448  dL=0.086  Nsig= 3.99   p1=0.00003299, p_all=0.01506122 >? pmin=0.050   N=460  mean=3.3616  rms=0.0217   -> Rejected
 167: L=3.447  dL=0.086  Nsig= 4.02   p1=0.00002922, p_all=0.01332302 >? pmin=0.050   N=459  mean=3.3614  rms=0.0213   -> Rejected
 126: L=3.445  dL=0.084  Nsig= 4.00   p1=0.00003155, p_all=0.01434519 >? pmin=0.050   N=458  mean=3.3612  rms=0.0209   -> Rejected
 162: L=3.445  dL=0.084  Nsig= 4.08   p1=0.00002276, p_all=0.01034822 >? pmin=0.050   N=457  mean=3.3610  rms=0.0206   -> Rejected
 340: L=3.280  dL=0.081  Nsig= 4.00   p1=0.00003223, p_all=0.01458856 >? pmin=0.050   N=456  mean=3.3609  rms=0.0202   -> Rejected
 393: L=3.285  dL=0.076  Nsig= 3.82   p1=0.00006640, p_all=0.02975990 >? pmin=0.050   N=455  mean=3.3610  rms=0.0199   -> Rejected
 121: L=3.433  dL=0.072  Nsig= 3.66   p1=0.00012405, p_all=0.05476548 >? pmin=0.050   N=454  mean=3.3612  rms=0.0196   -> Accepted
```

I rejected:
41 data points from the 30cm sample,
63 data points from the 2m sample.
And I inspected each and every one!

Unweighted results

30cm:

Mean = 3.36120 ± 0.00092 m

RMS = 0.020 m (N = 454)

2m:

Mean = 3.36295 ± 0.00026 m

RMS = 0.005 m (N = 430)

Correspondence between 30cm and 2m measurement:

Diff = -0.00175 ± 0.00130 m (-1.35 σ)

With corrections and Chavenet's (p=0.05)

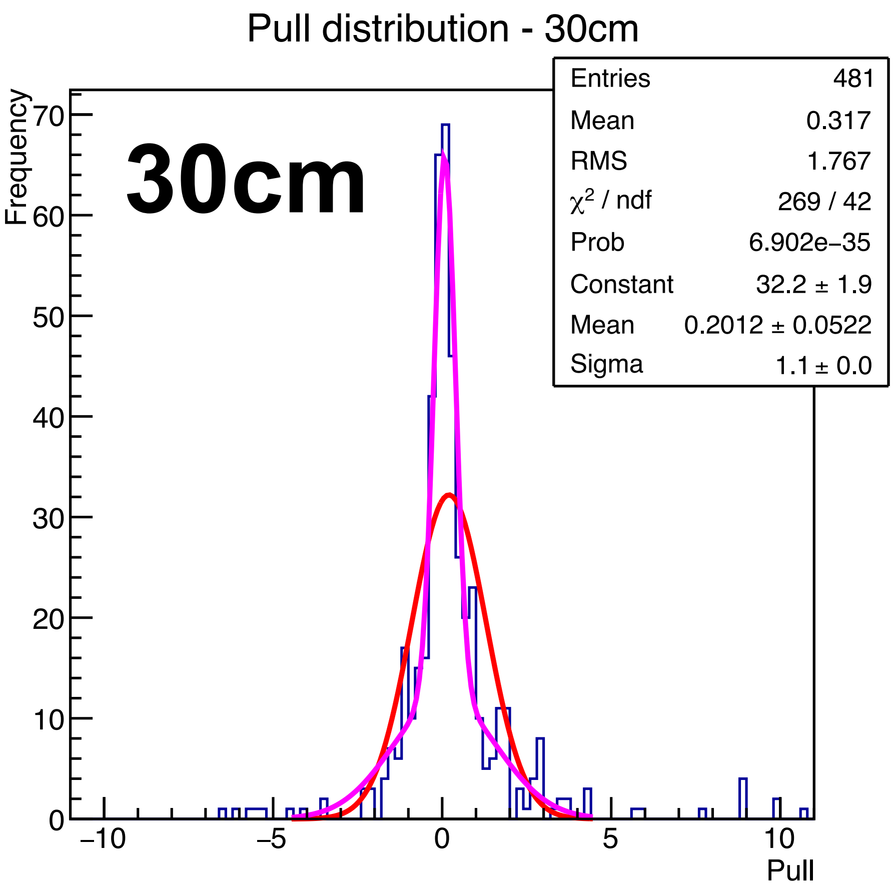
Improvement in error from naive to selected measurement 30cm: 10.6

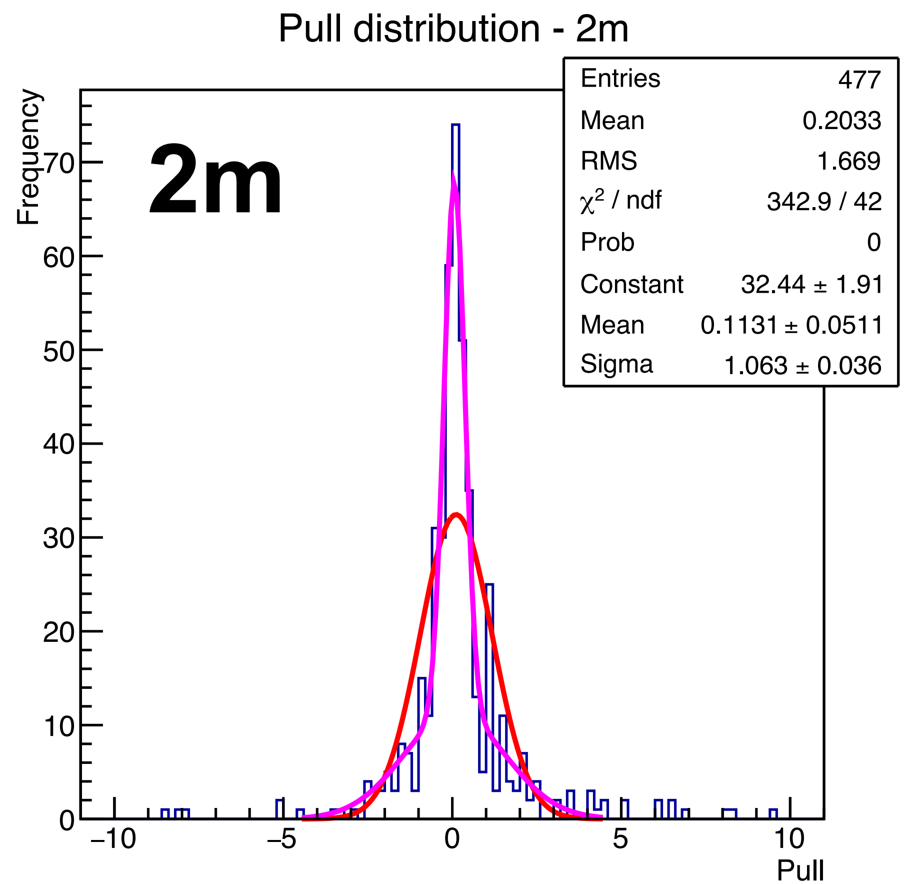
Improvement in error from naive to selected measurement 2m: 42.7

Weighted analysis

Considering the quoted uncertainties, we first need to evaluate their quality.

The plot to consider is a **PULL** plot, i.e. the distribution of: $z = \frac{x_i - \mu}{\sigma_i}$





The pulls should be unit Gaussian. I **decide** to exclude measurements beyond **4.5σ**.

Excluded data due to bad pull

30cm: Warning! Large pull: L = 3.4700 +- 0.0080 pull = 13.60
30cm: Warning! Large pull: L = 3.3700 +- 0.0010 pull = 8.80
30cm: Warning! Large pull: L = 1.3700 +- 0.0100 pull = -199.12
30cm: Warning! Large pull: L = 4.2600 +- 0.1000 pull = 8.99
30cm: Warning! Large pull: L = 3.3000 +- 0.0100 pull = -6.12
30cm: Warning! Large pull: L = 2.6700 +- 0.0100 pull = -69.12
30cm: Warning! Large pull: L = 2.5630 +- 0.0200 pull = -39.91
30cm: Warning! Large pull: L = 4.1400 +- 0.0400 pull = 19.47
30cm: Warning! Large pull: L = 3.4670 +- 0.0120 pull = 8.82
30cm: Warning! Large pull: L = 3.4710 +- 0.0050 pull = 21.96
30cm: Warning! Large pull: L = 3.4470 +- 0.0050 pull = 17.16
30cm: Warning! Large pull: L = 3.2210 +- 0.0310 pull = -4.52
30cm: Warning! Large pull: L = 3.3470 +- 0.0010 pull = -14.20
30cm: Warning! Large pull: L = 3.3650 +- 0.0005 pull = 7.60
30cm: Warning! Large pull: L = 3.3670 +- 0.0010 pull = 5.80
30cm: Warning! Large pull: L = 3.3290 +-
30cm: Warning! Large pull: L = 3.2600 +-
30cm: Warning! Large pull: L = 3.4240 +-
30cm: Warning! Large pull: L = 3.3050 +-
30cm: Warning! Large pull: L = 3.3000 +-
30cm: Warning! Large pull: L = 2.7680 +-
30cm: Warning! Large pull: L = 3.4200 +-
30cm: Warning! Large pull: L = 3.3900 +-
30cm: Warning! Large pull: L = 1.2750 +-
30cm: Warning! Large pull: L = 4.0040 +- 0.0300 pull = 21.43
30cm: Warning! Large pull: L = 3.9550 +- 0.0550 pull = 10.80
30cm: Warning! Large pull: L = 3.2150 +- 0.0110 pull = -13.29
30cm: Warning! Large pull: L = 1.3650 +- 0.0050 pull = -399.24
30cm: Warning! Large pull: L = 3.4110 +- 0.0050 pull = 9.96
30cm: Warning! Large pull: L = 3.4500 +- 0.0050 pull = 17.76
30cm: Warning! Large pull: L = 2.7690 +- 0.0050 pull = -118.44
30cm: Warning! Large pull: L = 3.3080 +- 0.0100 pull = -5.32
30cm: Warning! Large pull: L = 3.3700 +- 0.0010 pull = 8.80
30cm: Warning! Large pull: L = 2.7590 +- 0.0020 pull = -301.10
30cm: Warning! Large pull: L = 2.7440 +- 0.0500 pull = -12.34
30cm: Warning! Large pull: L = 3.3350 +- 0.0010 pull = -26.20
30cm: Warning! Large pull: L = 3.4610 +- 0.0100 pull = 9.98
30cm: Warning! Large pull: L = 2.7000 +- 0.0300 pull = -22.04

2m: Warning! Large pull: L = 3.0600 +- 0.0050 pull = -60.59
2m: Warning! Large pull: L = 3.7500 +- 0.0200 pull = 19.35
2m: Warning! Large pull: L = 3.3680 +- 0.0010 pull = 5.05
2m: Warning! Large pull: L = 3.6500 +- 0.0300 pull = 9.57
2m: Warning! Large pull: L = 3.3200 +- 0.0050 pull = -8.59
2m: Warning! Large pull: L = 3.6660 +- 0.0500 pull = 6.06
2m: Warning! Large pull: L = 3.5750 +- 0.0020 pull = 106.03
2m: Warning! Large pull: L = 3.0360 +- 0.0400 pull = -8.17
2m: Warning! Large pull: L = 3.3500 +- 0.0010 pull = -12.95
2m: Warning! Large pull: L = 3.0360 +- 0.0050 pull = -65.39
2m: Warning! Large pull: L = 2.3630 +- 0.0050 pull = -199.99
2m: Warning! Large pull: L = 3.3150 +- 0.0006 pull = -79.92
2m: Warning! Large pull: L = 3.4170 +- 0.0120 pull = 4.50
2m: Warning! Large pull: L = 3.6400 +- 0.0100 pull = 27.71
2m: Warning! Large pull: L = 4.0400 +- 0.0200 pull = 33.85
L = 4.0160 +- 0.1010 pull = 6.47
L = 3.0640 +- 0.0050 pull = -59.79
L = 3.6600 +- 0.0120 pull = 24.75
L = 3.3550 +- 0.0010 pull = -7.95
L = 3.7500 +- 0.0050 pull = 77.41
L = 3.3690 +- 0.0010 pull = 6.05
L = 3.0070 +- 0.0080 pull = -44.49
L = 3.1680 +- 0.0050 pull = -38.99
L = 3.0600 +- 0.0050 pull = -60.59
2m: Warning! Large pull: L = 3.3720 +- 0.0020 pull = 4.53
2m: Warning! Large pull: L = 3.1500 +- 0.0100 pull = -21.29
2m: Warning! Large pull: L = 2.3640 +- 0.0300 pull = -33.30
2m: Warning! Large pull: L = 3.4200 +- 0.0050 pull = 11.41
2m: Warning! Large pull: L = 3.3720 +- 0.0011 pull = 8.23
2m: Warning! Large pull: L = 3.3970 +- 0.0030 pull = 11.35
2m: Warning! Large pull: L = 3.6050 +- 0.0150 pull = 16.14
2m: Warning! Large pull: L = 3.3070 +- 0.0050 pull = -11.19
2m: Warning! Large pull: L = 3.3730 +- 0.0020 pull = 5.03
2m: Warning! Large pull: L = 3.3760 +- 0.0020 pull = 6.53
2m: Warning! Large pull: L = 2.3650 +- 0.0050 pull = -199.59
2m: Warning! Large pull: L = 2.3650 +- 0.0040 pull = -249.49
2m: Warning! Large pull: L = 3.0020 +- 0.0700 pull = -5.16
2m: Warning! Large pull: L = 3.2600 +- 0.0200 pull = -5.15
2m: Warning! Large pull: L = 3.2650 +- 0.0010 pull = -97.95
2m: Warning! Large pull: L = 3.5000 +- 0.0200 pull = 6.85
2m: Warning! Large pull: L = 3.1695 +- 0.0020 pull = 90.72
2m: Warning! Large pull: L = 3.3660 +- 0.0003 pull = 12.20
2m: Warning! Large pull: L = 3.6540 +- 0.0200 pull = 14.55
2m: Warning! Large pull: L = 3.3710 +- 0.0010 pull = 8.05
2m: Warning! Large pull: L = 3.0650 +- 0.0050 pull = -59.59

I rejected:
38 data points from the 30cm sample,
45 data points from the 2m sample.
And I inspected each and every one!

Weighted results

<u>30cm:</u>

Mean = 3.36376 ± 0.00030 m

RMS = undefined!   (N = 443)

<u>2m:</u>

Mean = 3.36385 ± 0.00014 m

RMS = undefined! (N = 432)

<u>Correspondence between 30cm and 2m measurement:</u>

Diff = -0.00009 ± 0.00033 m   (-0.27 σ)

30cm: $\sigma_\mu = 0.00092 \rightarrow \sigma_\mu = 0.00030$ (factor 3.1)

2m: $\sigma_\mu = 0.00026 \rightarrow \sigma_\mu = 0.00014$ (factor 1.8)

Weighted results

30cm:

Mean = 3.36376 ± 0.00030 m

Chi2 = 460.5, Ndof = 442, Prob = 0.26

2m:

Mean = 3.36385 ± 0.00014 m

Chi2 = 451.1, Ndof = 431, Prob = 0.24

Correspondence between 30cm and 2m measurement:

Diff = -0.00009 ± 0.00033 m (-0.27 σ)

30cm: $\sigma_\mu = 0.00092 \rightarrow \sigma_\mu = 0.00030$ (factor 3.1)

2m: $\sigma_\mu = 0.00026 \rightarrow \sigma_\mu = 0.00014$ (factor 1.8)

A problem?

Correspondence: unweighted vs. weighted 30cm measurements:

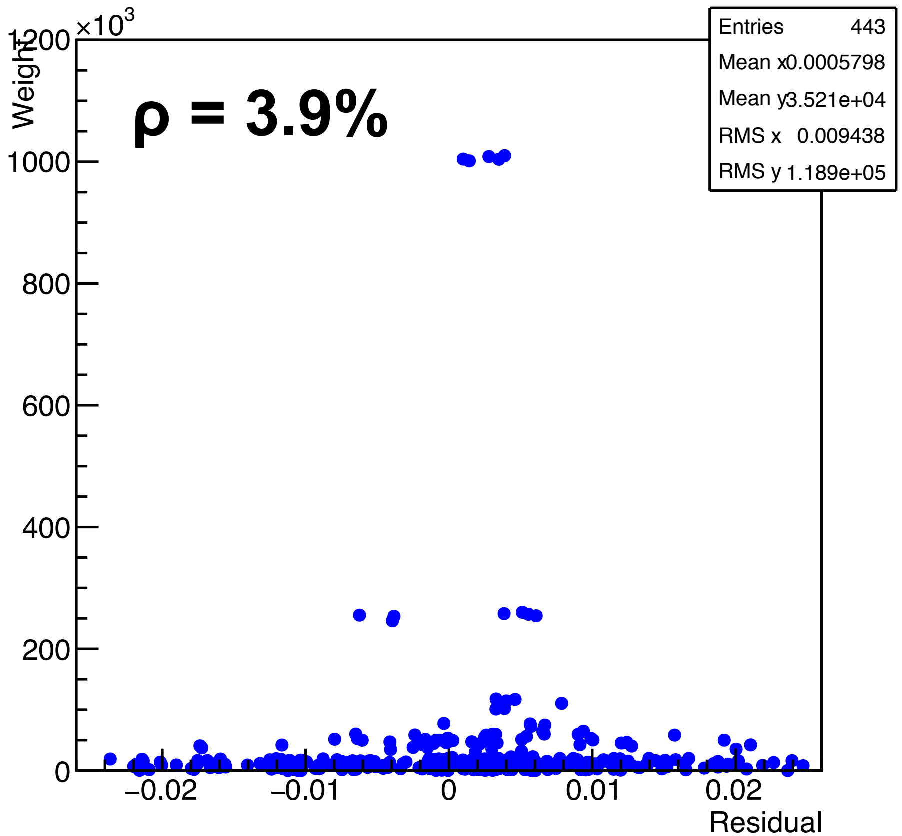
Diff = 0.00257 +- 0.00097   (2.65 sigma)

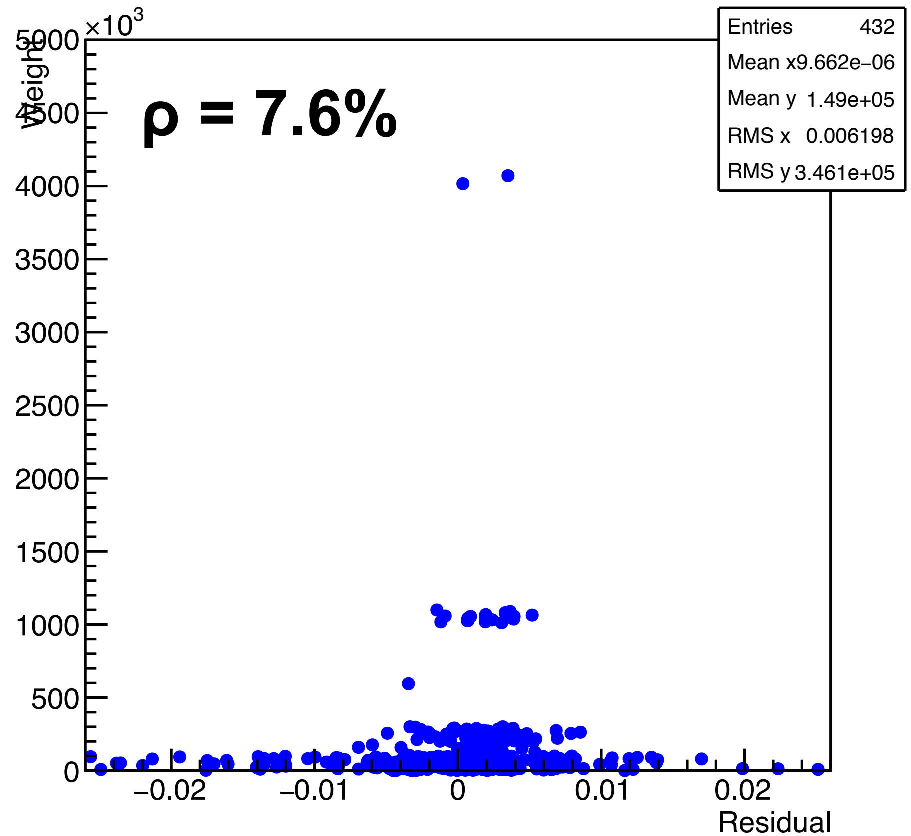
Correspondence: unweighted vs. weighted 2m measurements:

Diff = 0.00090 +- 0.00030   (3.04 sigma)

A correlation?

Plotted is the **measurement residuals** (measurement - mean) as a function of the **weight of the measurement** (i.e. 1/uncertainty^2).





There is a slight correlation between the lengths and uncertainties quoted!
"Longer measurements have smaller errors!". Why? I don't know. But data shows it!

Fitting for a result

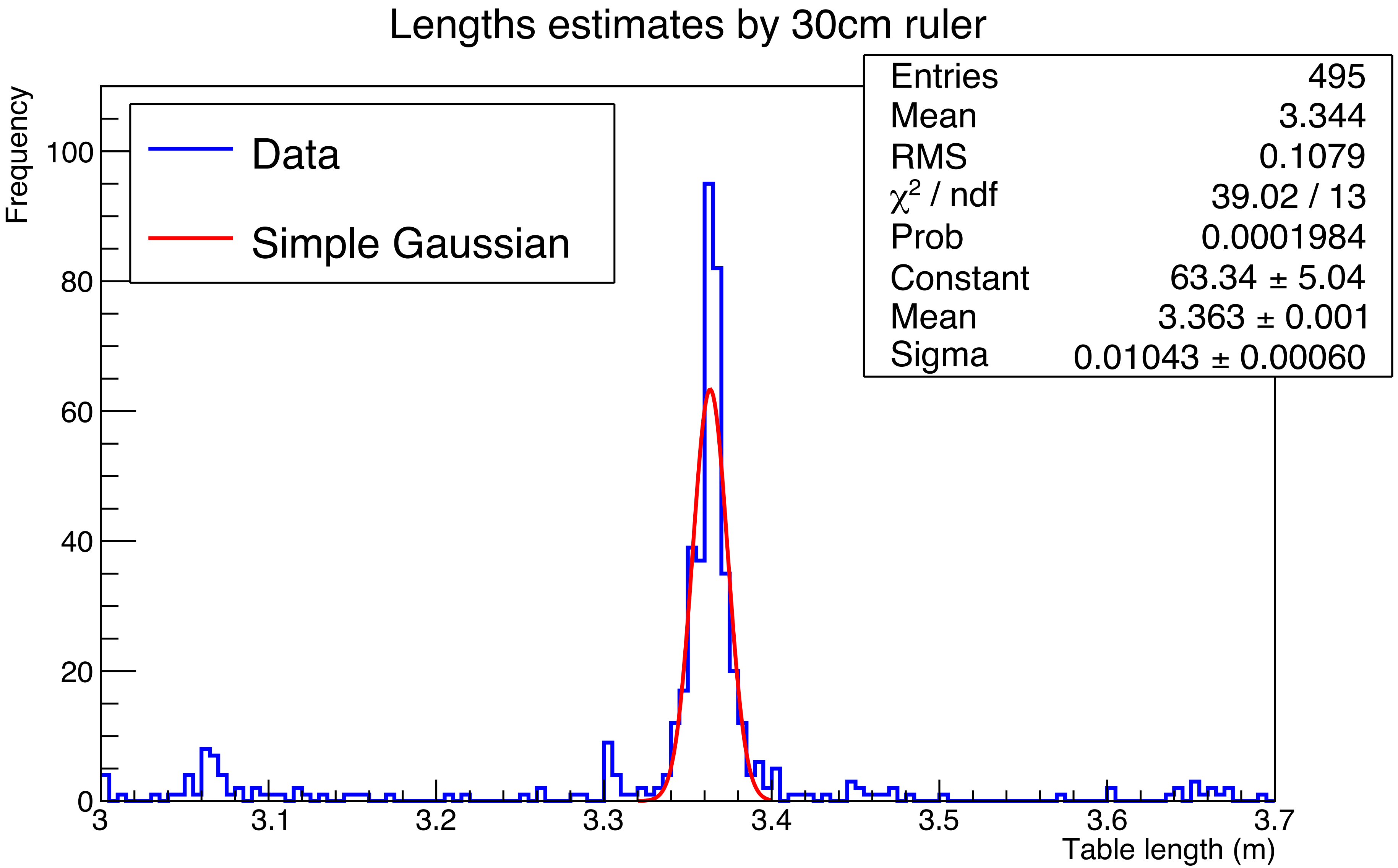
A completely different approach is to fit the RAW data, hence describing all data points instead of excluding some.

This approach is philosophically more clean, but certainly not easy!

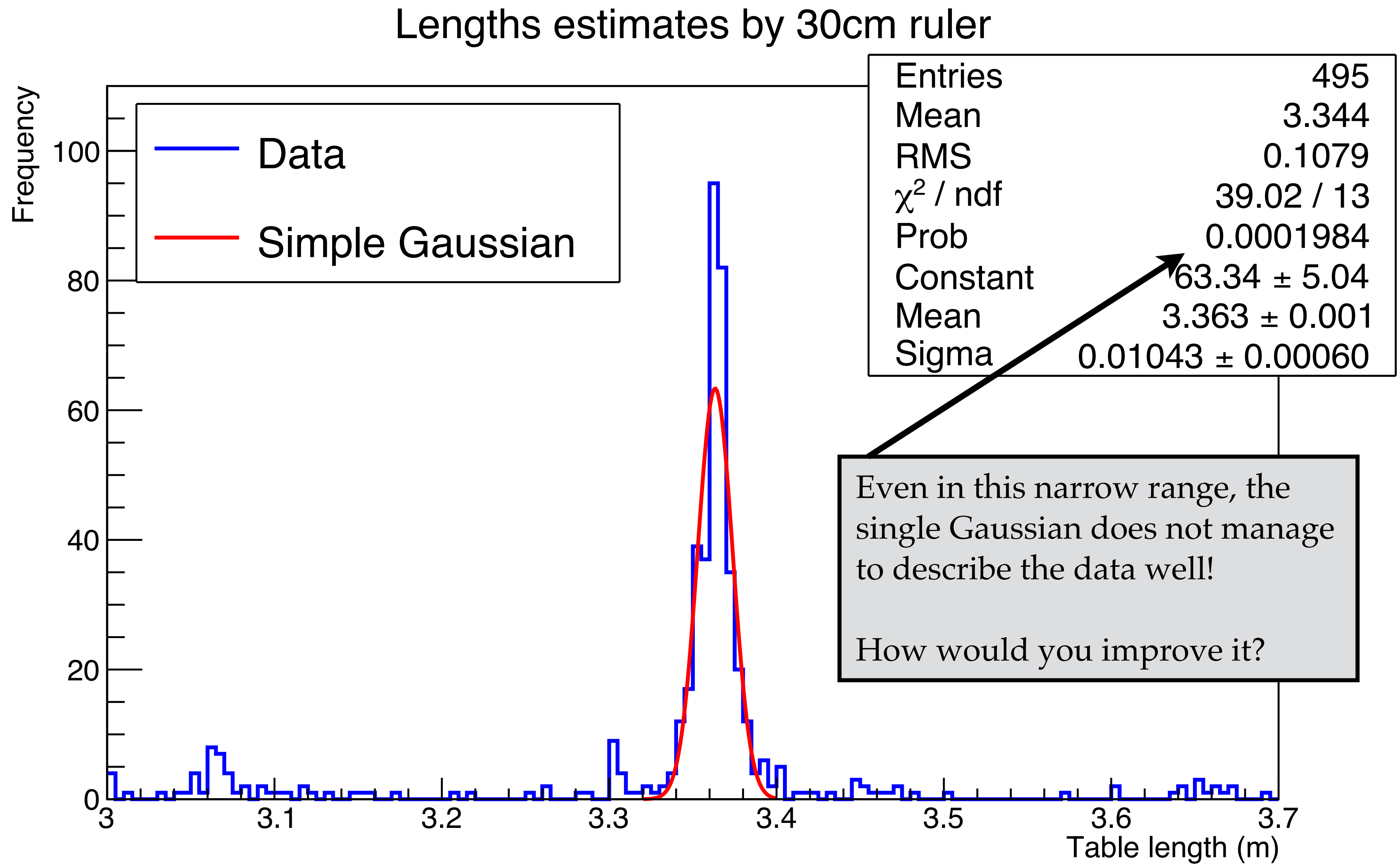
Challenges:
- Measurements has many different resolutions.
- There are several peaks in the data (30cm case).
- Some measurements are clearly rounded.

While all of these can be accommodated, it is still a challenge, at the following "fitting around" took me half and hour!

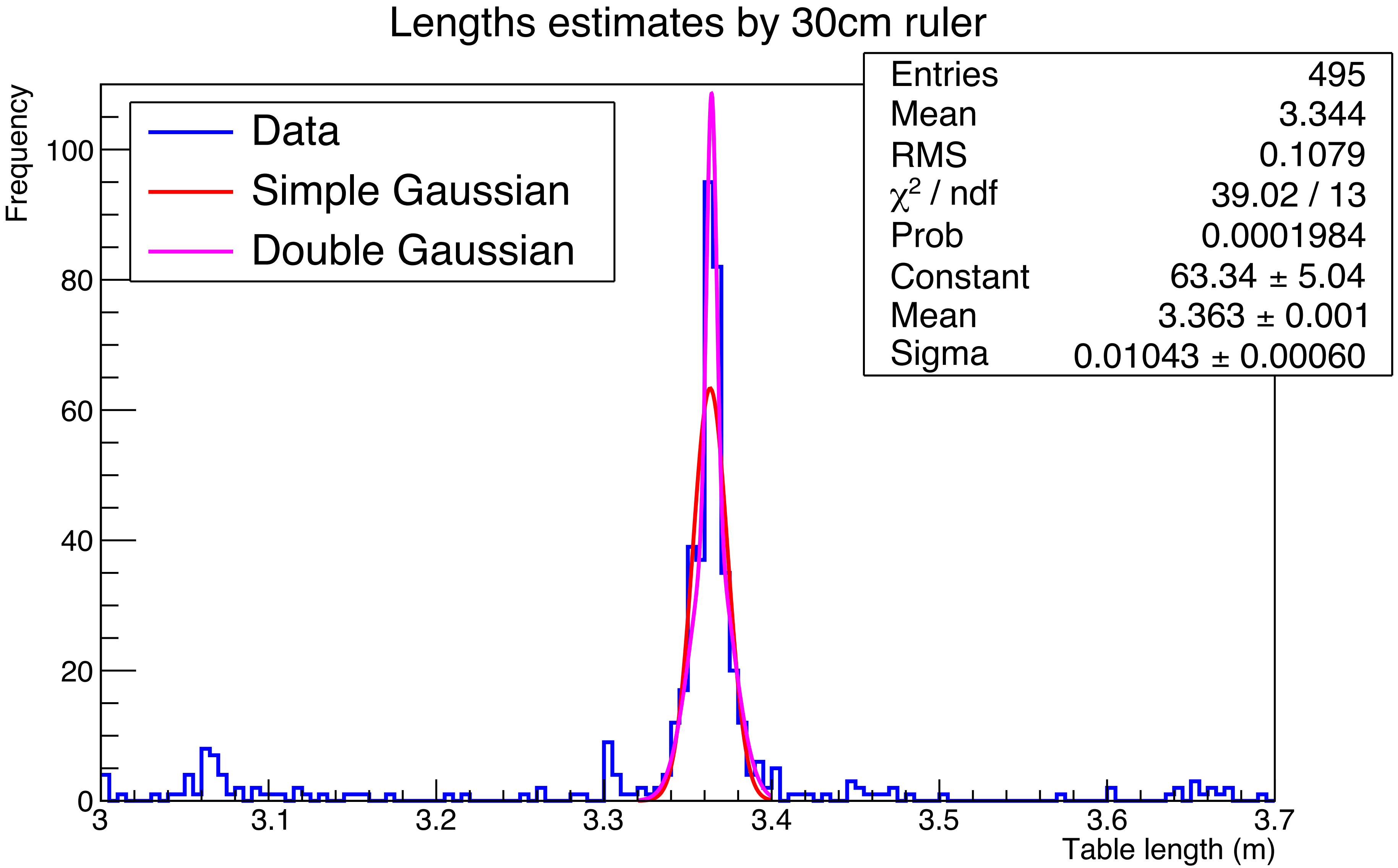
Fitting for a result



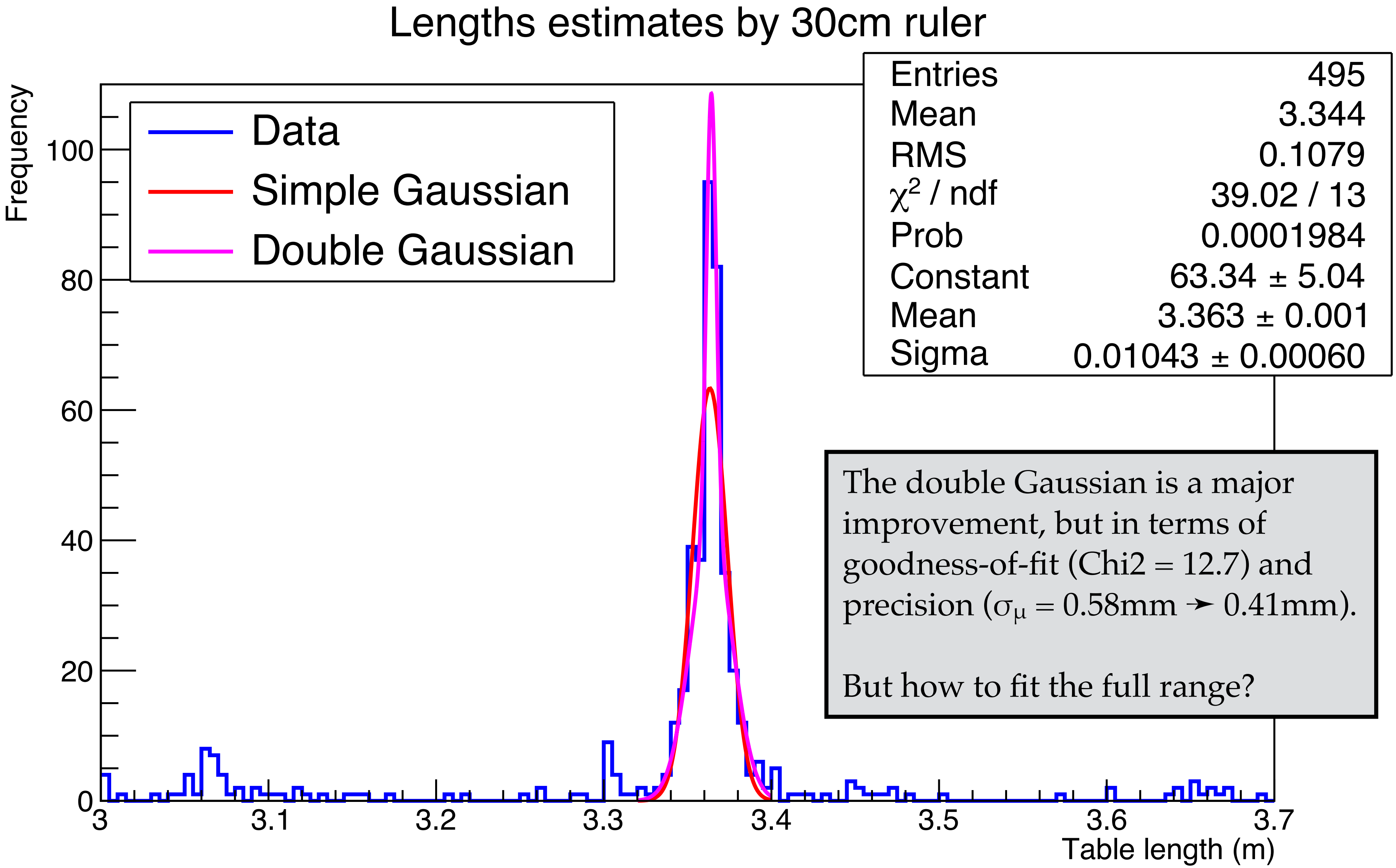
Fitting for a result



Fitting for a result



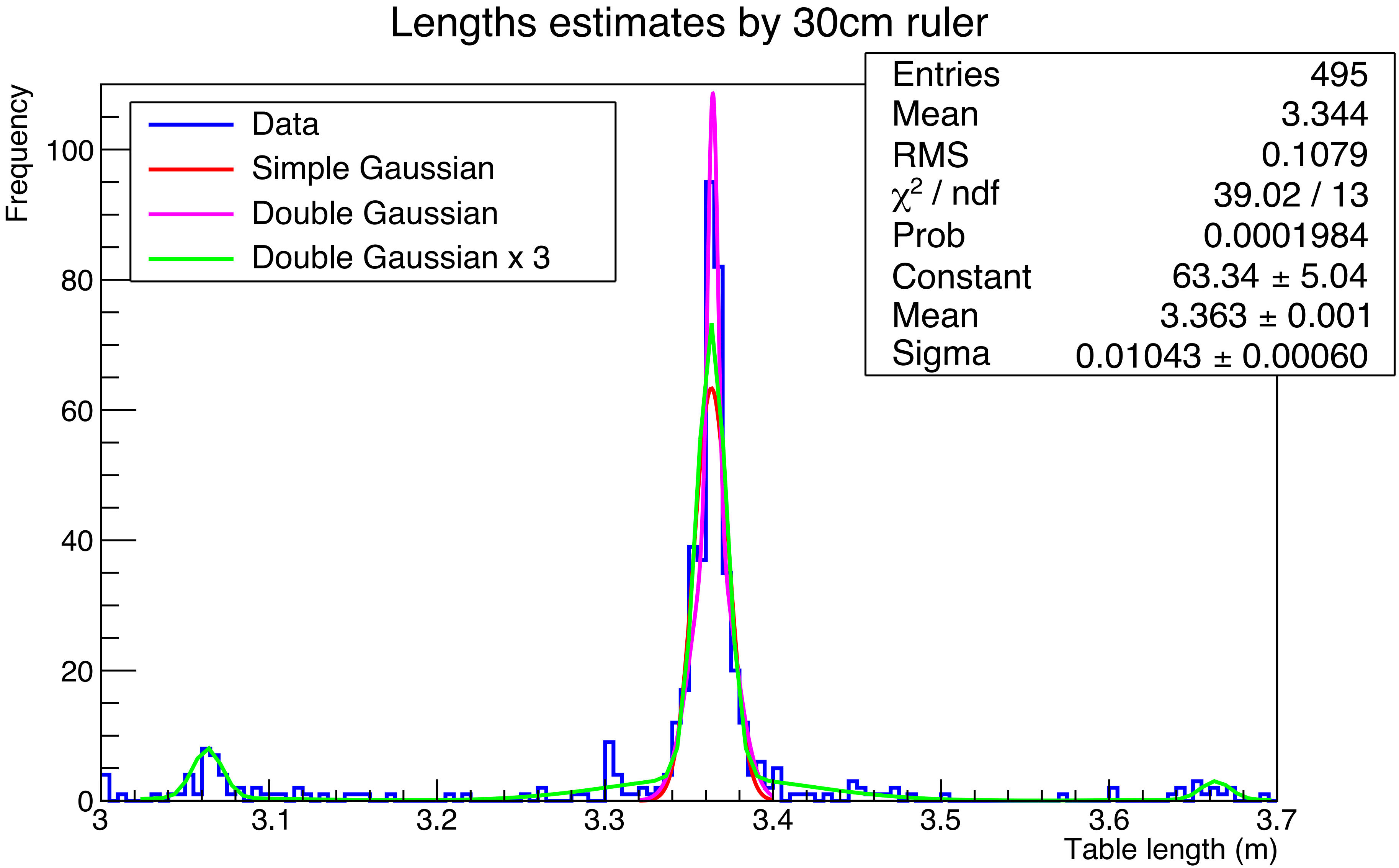
Fitting for a result



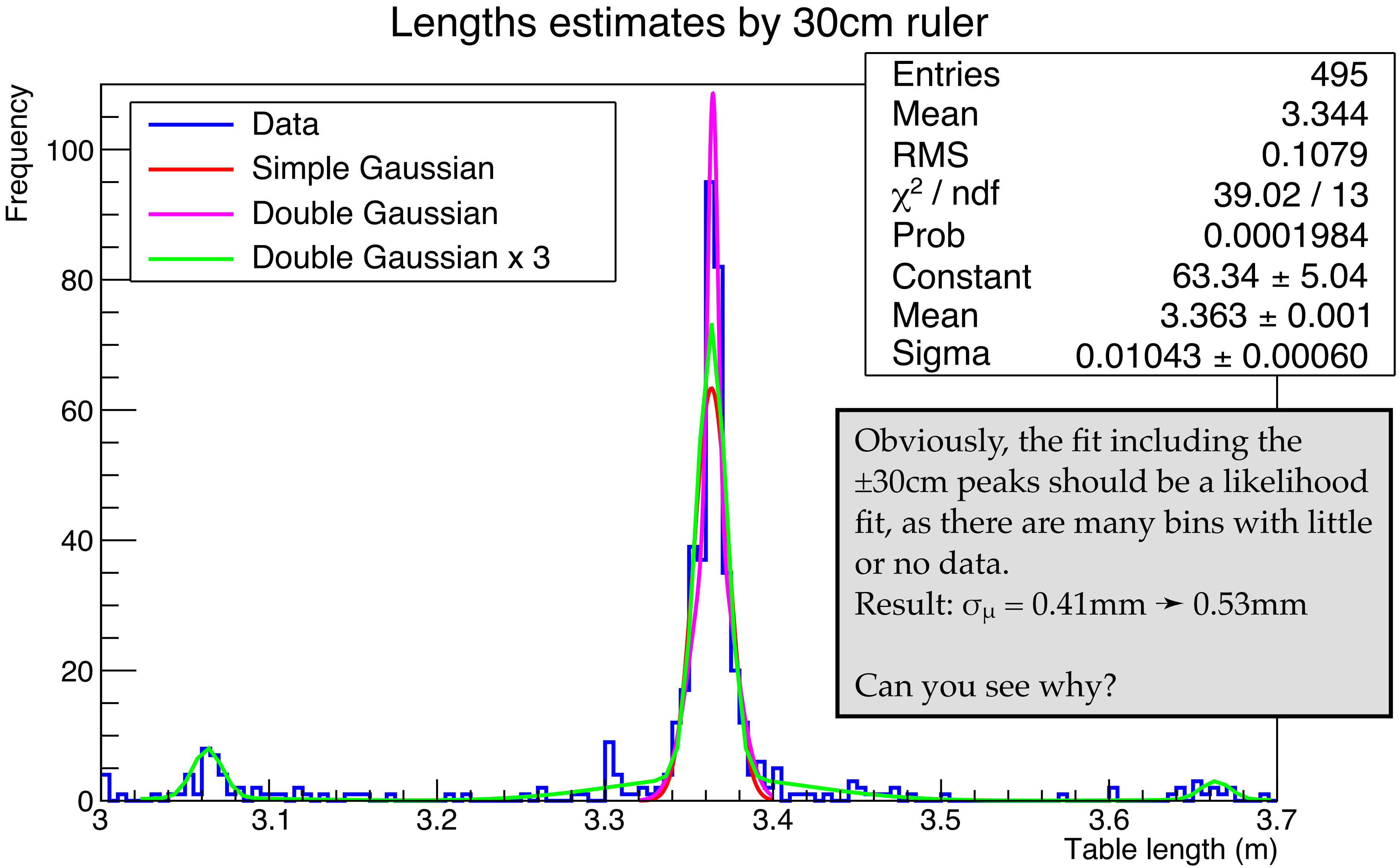
The double Gaussian is a major improvement, but in terms of goodness-of-fit (Chi2 = 12.7) and precision (σ_μ = 0.58mm ➛ 0.41mm).

But how to fit the full range?

Fitting for a result



Fitting for a result

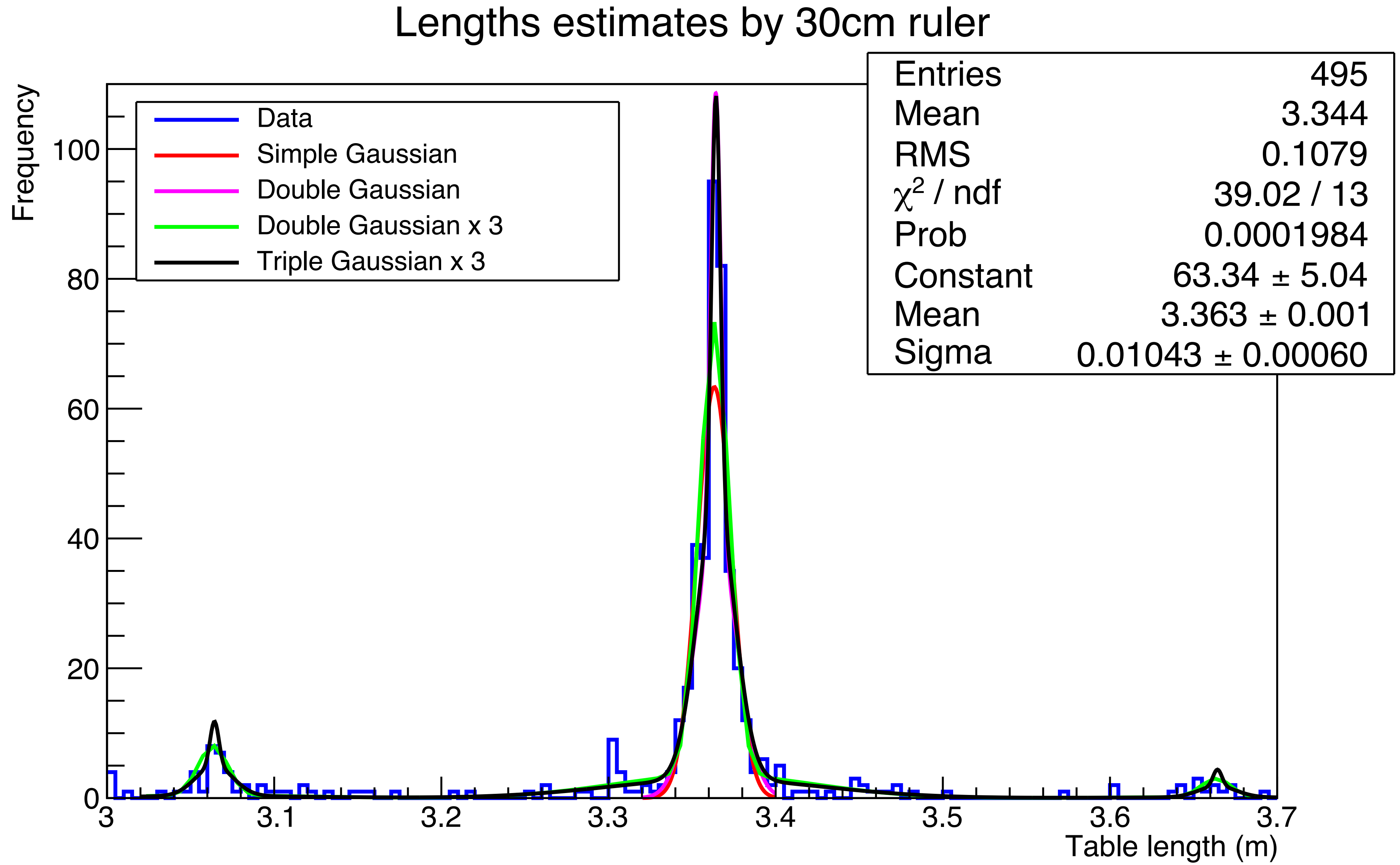


Obviously, the fit including the ±30cm peaks should be a likelihood fit, as there are many bins with little or no data.

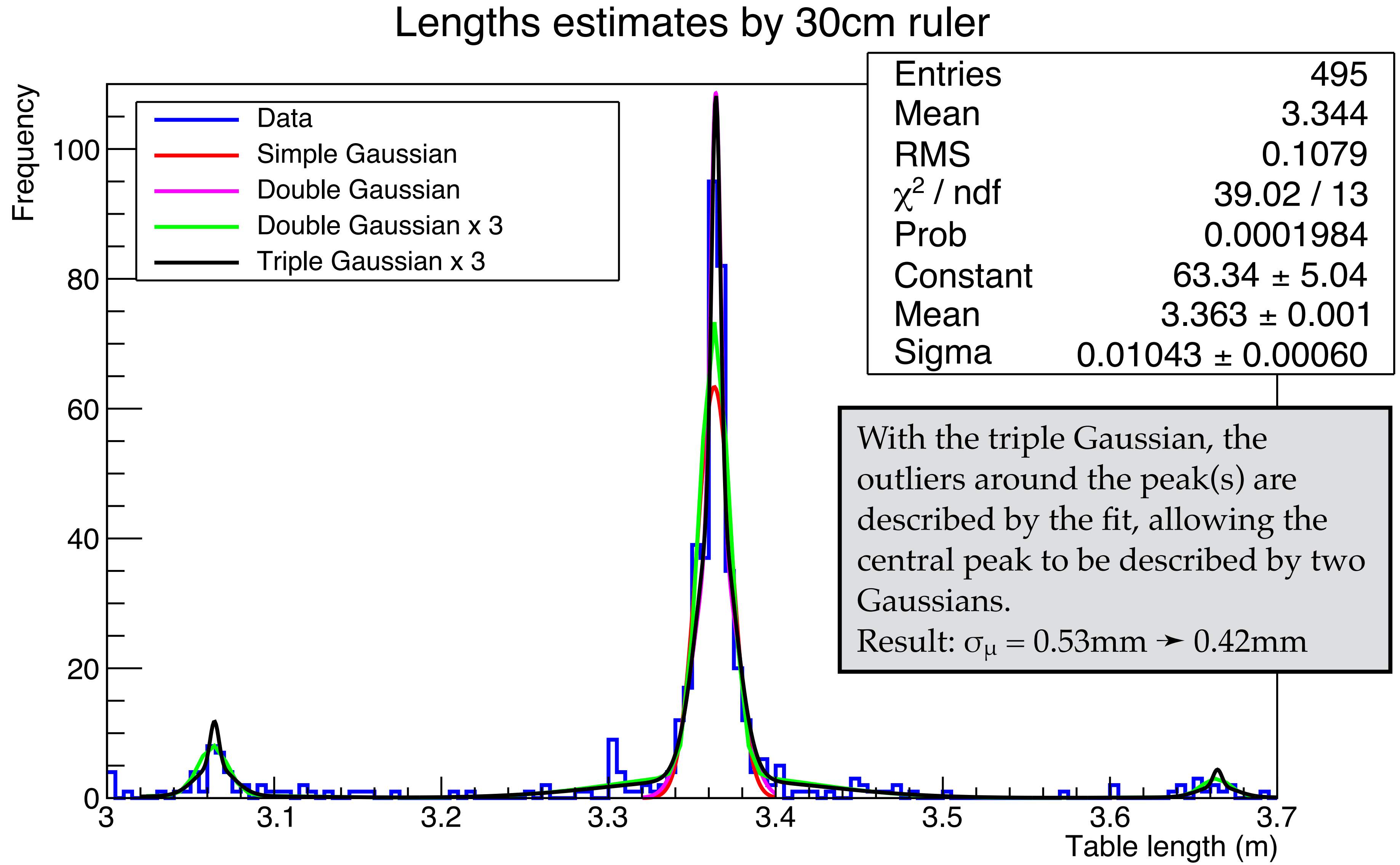
Result: σ_μ = 0.41mm ➛ 0.53mm

Can you see why?

Fitting for a result



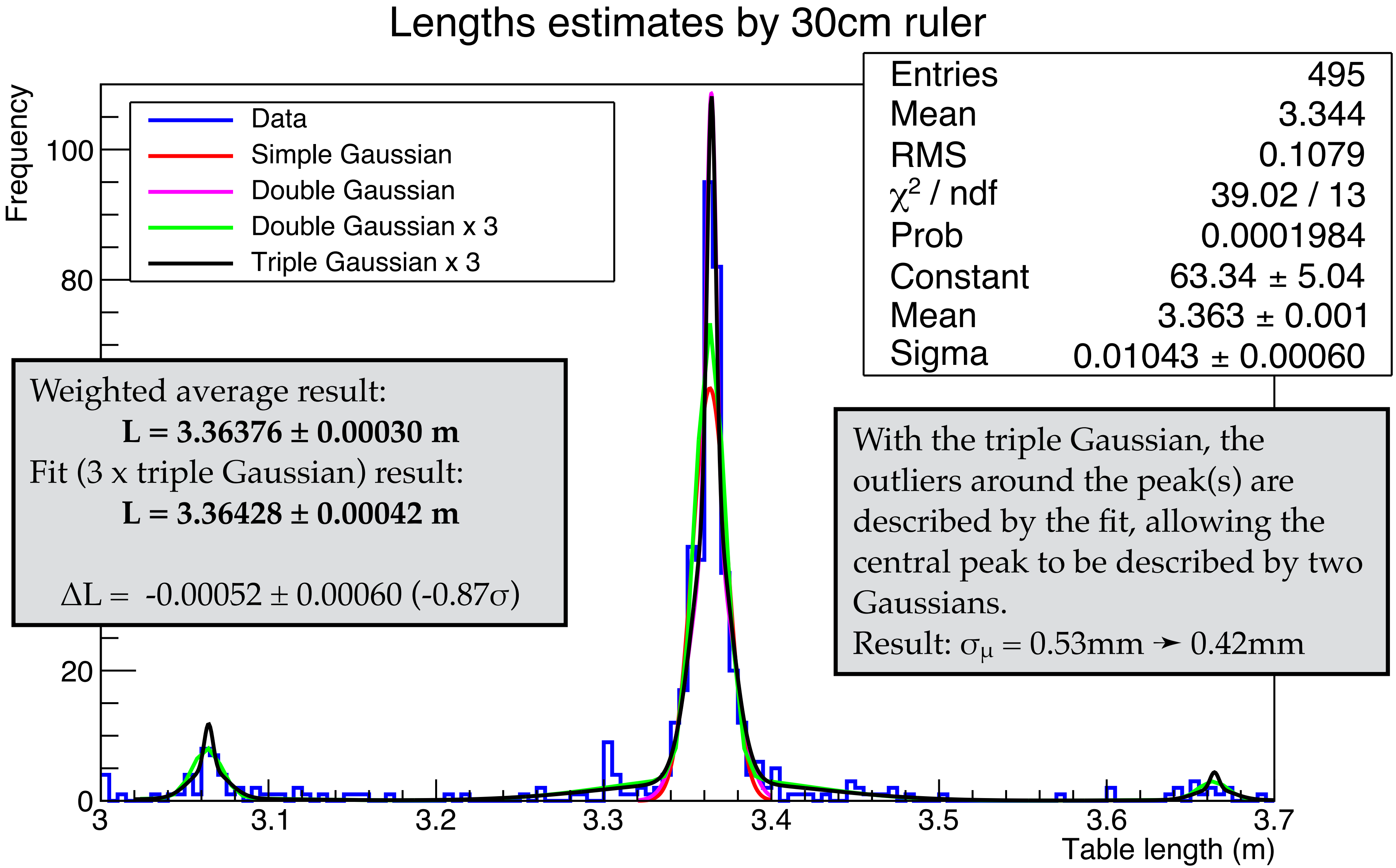
Fitting for a result



With the triple Gaussian, the outliers around the peak(s) are described by the fit, allowing the central peak to be described by two Gaussians.

Result: $\sigma_\mu = 0.53\text{mm} \rightarrow 0.42\text{mm}$

Fitting for a result



Weighted average result:

L = 3.36376 ± 0.00030 m

Fit (3 x triple Gaussian) result:

L = 3.36428 ± 0.00042 m

$\Delta L =$ -0.00052 ± 0.00060 (-0.87σ)

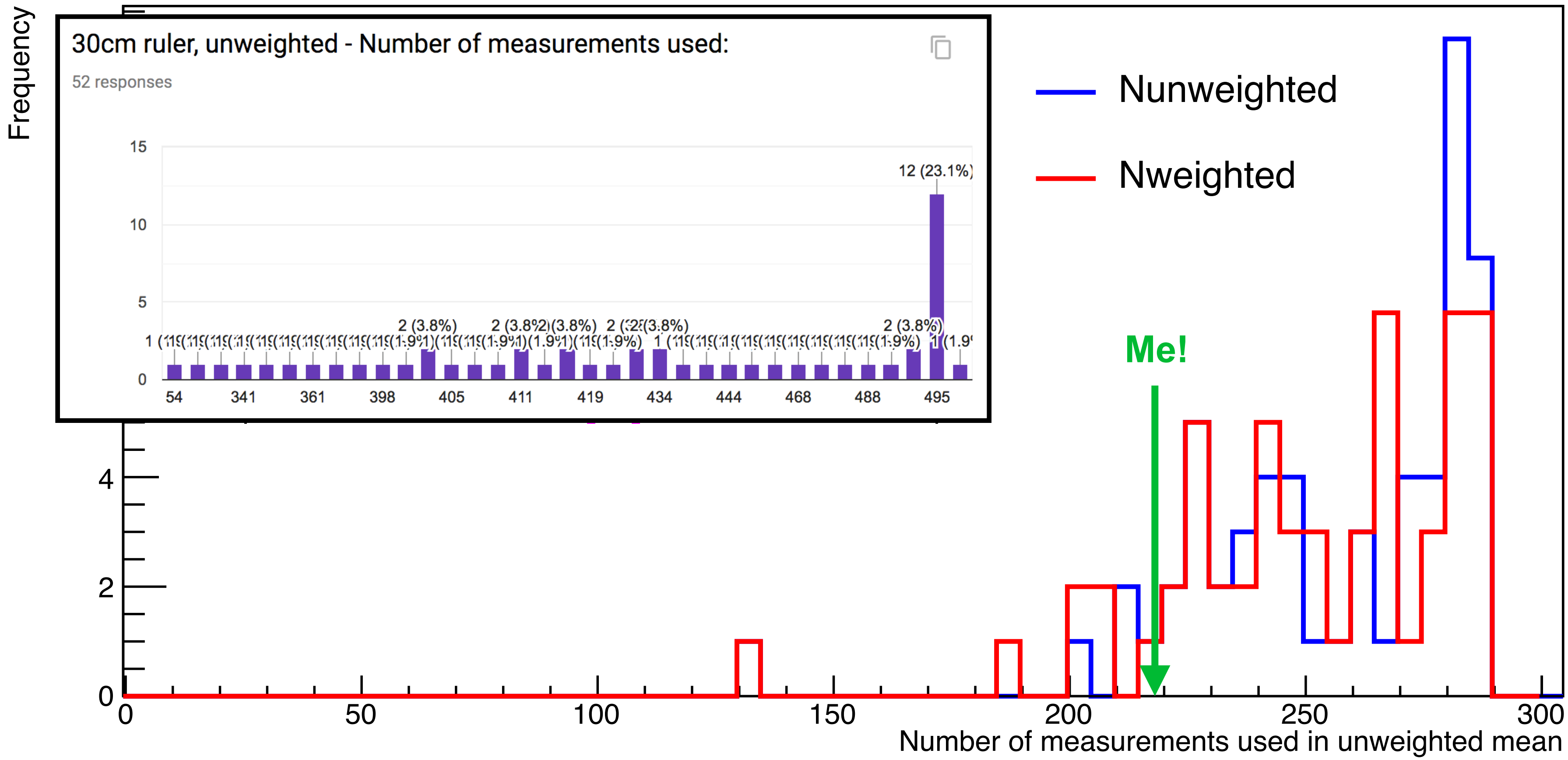
With the triple Gaussian, the outliers around the peak(s) are described by the fit, allowing the central peak to be described by two Gaussians.

Result: σ_μ = 0.53mm ➛ 0.42mm

Your analyses

Your measurement results

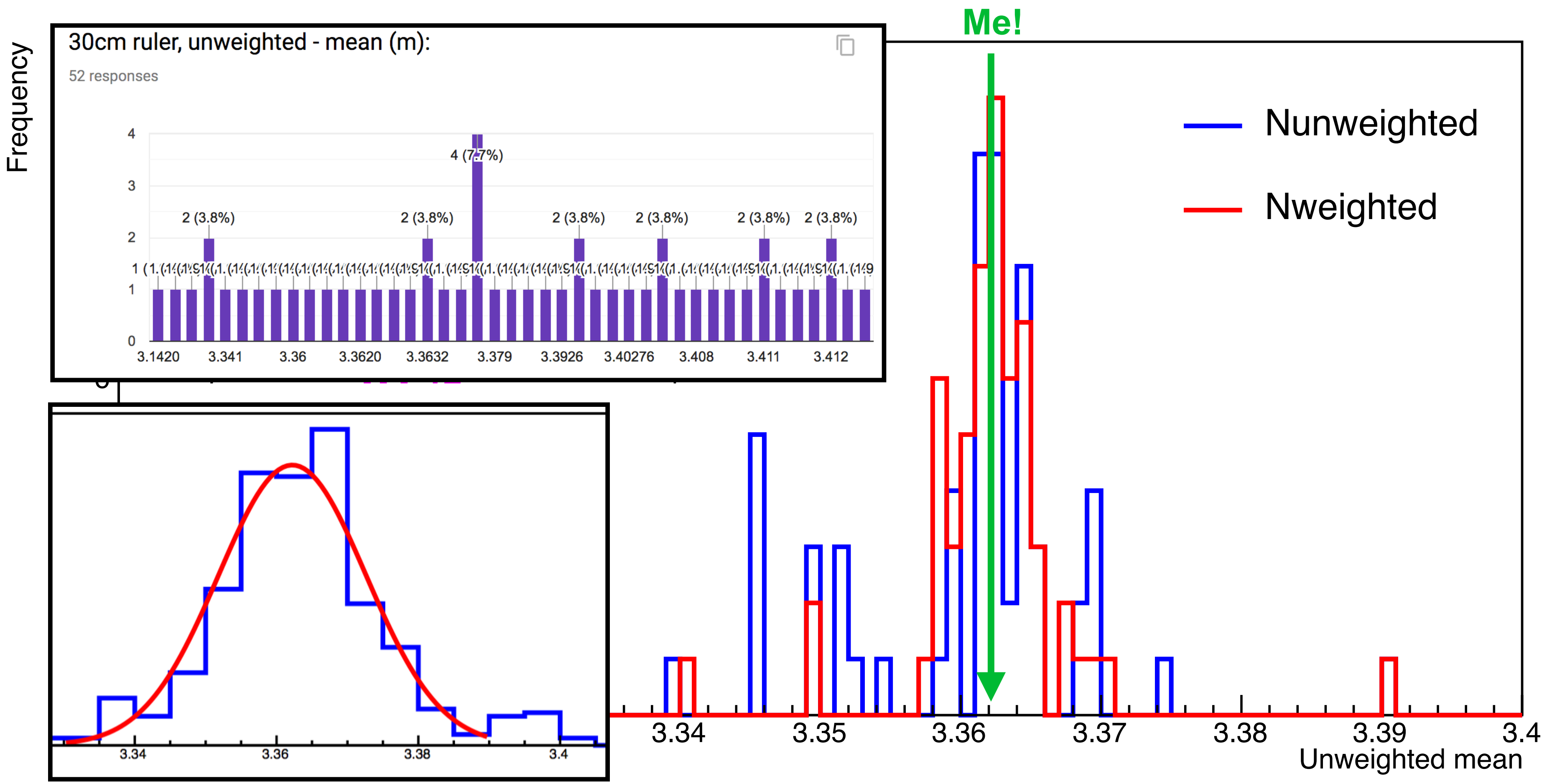
The number of measurements used varied quiet a bit.



But remember that the impact is only sqrt(N), and thus not that important!

Length results

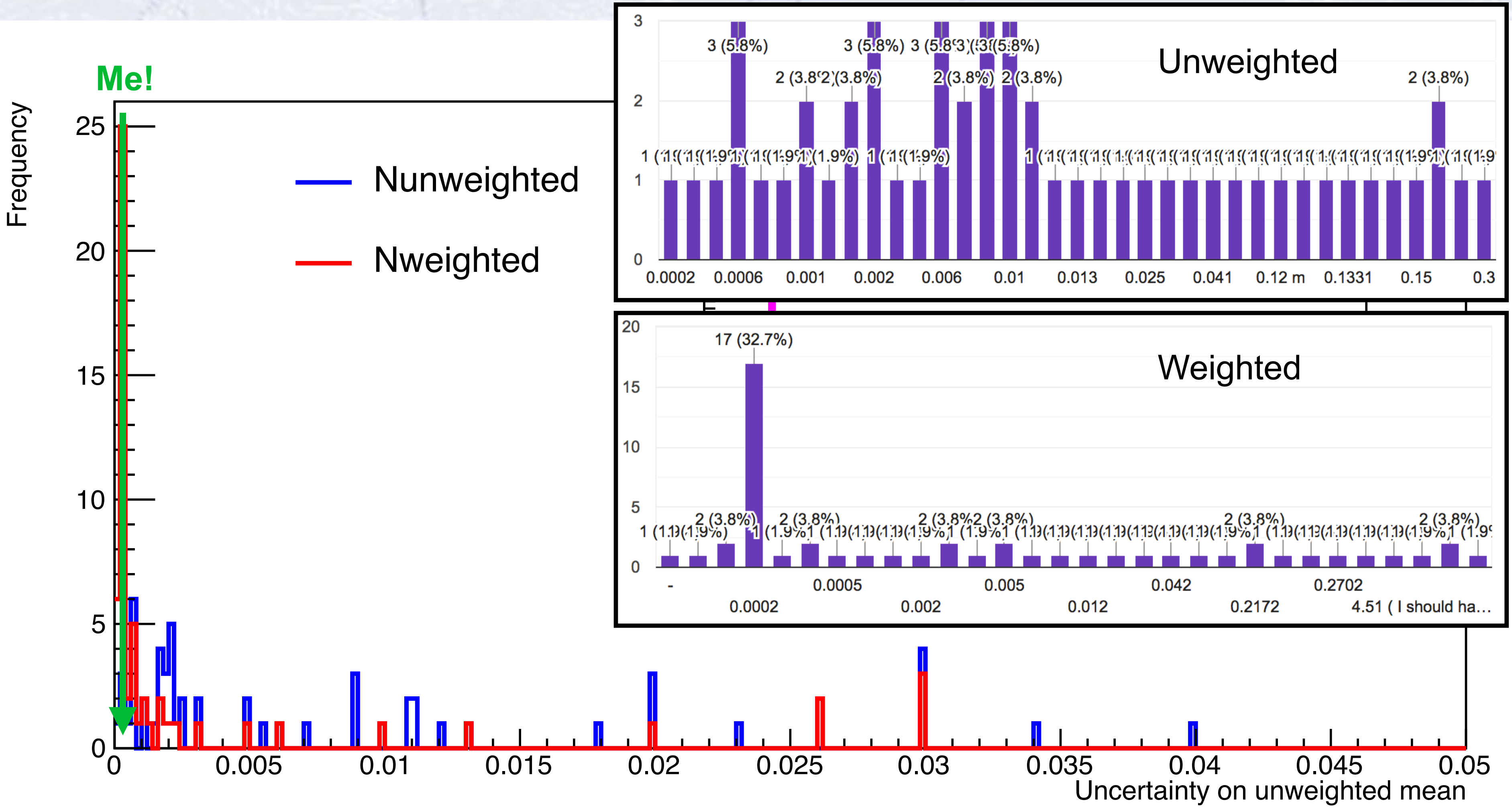
Results are relatively consistent… 80% of you get a value within 0.1% of “true”



I got: L(unweighted) = 3.36227 ± 0.00061 m, L(weighted) = 3.36371 ± 0.00035 m

Uncertainties

The uncertainties varied quite a bit - by more than a factor 100! Think about that.



I got: L(unweighted) = 3.36227 ± 0.00061 m, L(weighted) = 3.36371 ± 0.00035 m

Conclusions

<u>Specifically on the analysis:</u>

- Greatest improvement came from simply removing mis-measurements!
- Weighted result was a further improvement, but required good uncertainties.
- The uncertainties are accepted as "reasonable", as they have good pull distributions, and improve the result.
- The 30cm and 2m results match, giving credibility to the stated precision.

<u>More generally:</u>

- What appears to be a trivial task, turns out to require some thought anyhow.
  (Ask yourself how many fellow students would have been able to get a good result and error?)
- There were several choices to be made in the analysis:
  1. Which measurements to accept.
  2. Which uncertainties to accept.
  3. To correct or discard understood mis-measurements.
- All this can be solved with simple Python code.