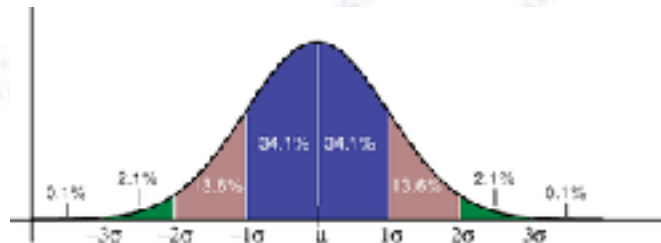


Applied Statistics

Measuring the length of a Table...

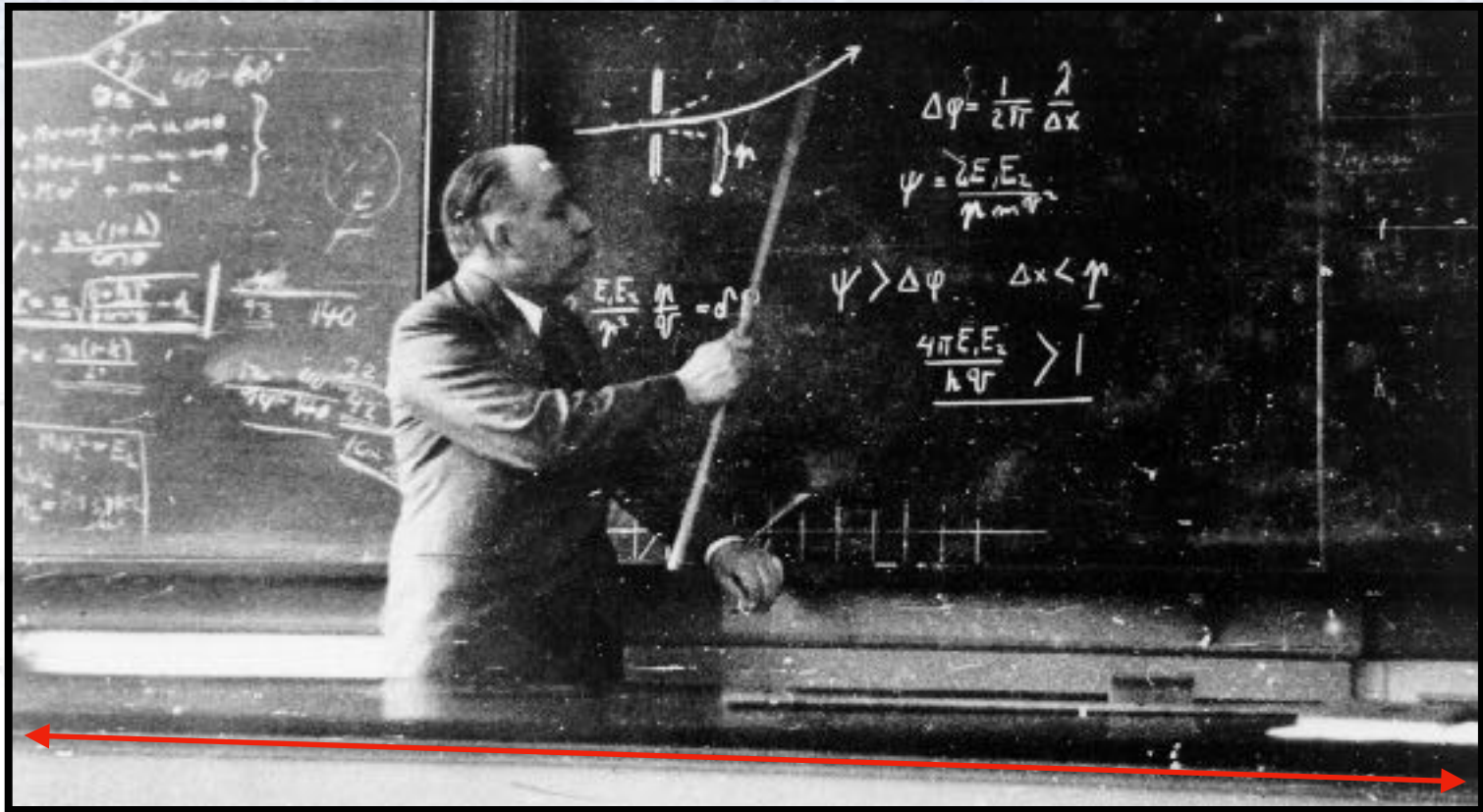


Troels C. Petersen (NBI)



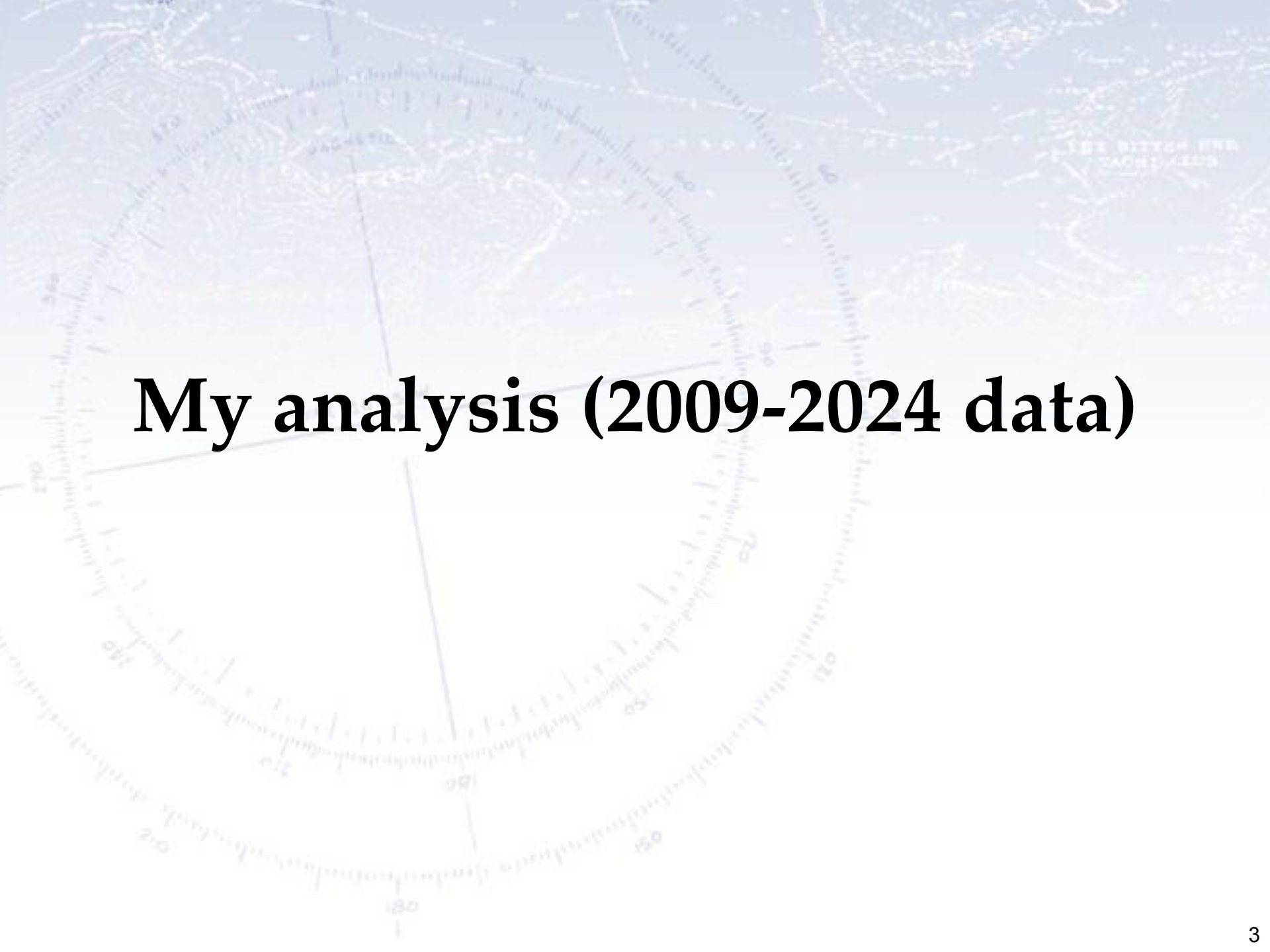
"Statistics is merely a quantisation of common sense"

The table in auditorium A



“Everything is vague to a degree you do not realise till you have tried to make it precise.”

[Bertrand Russell, 1872-1970]

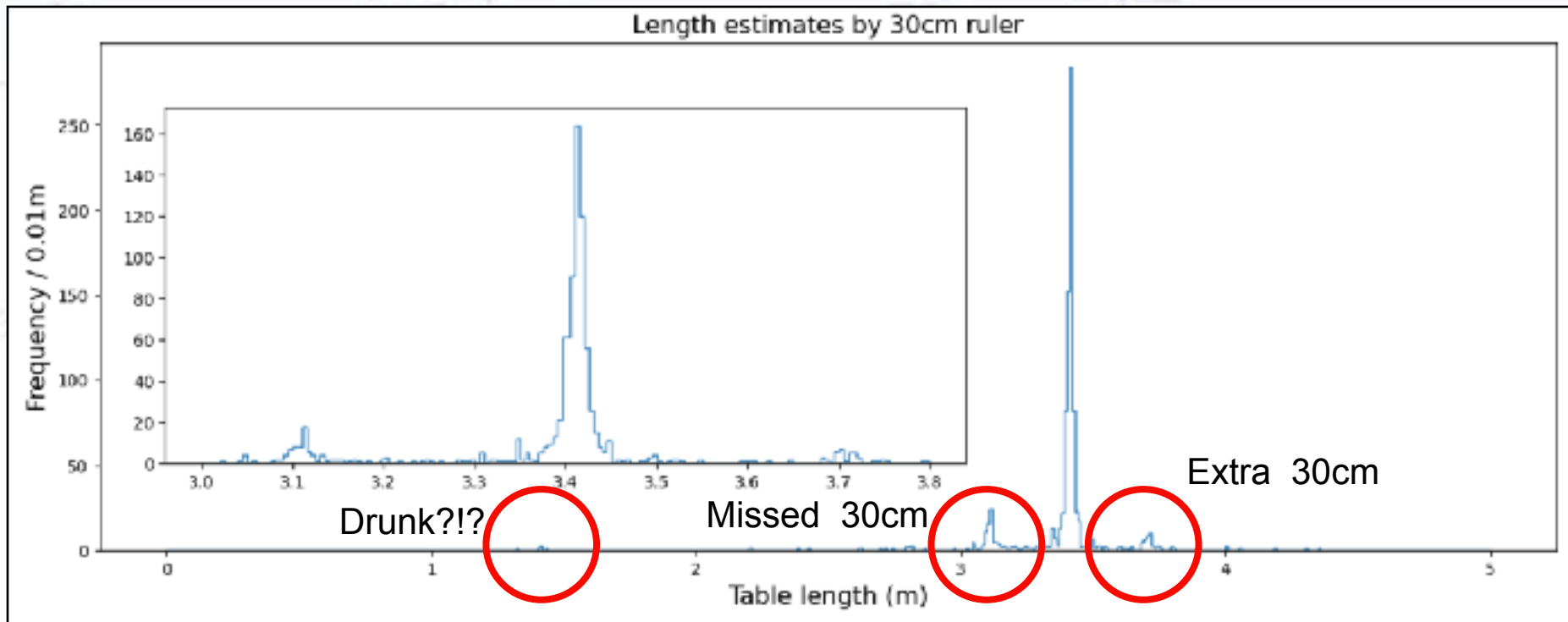
The background is a light blue map of the Pacific Ocean. A yellow line traces a path across the map, starting from the top left, moving towards the center, and then curving downwards towards the bottom left. The map includes various geographical features like coastlines and islands, as well as some text labels like 'PACIFIC OCEAN' and '181° 00' 00" W'.

My analysis (2009-2024 data)

The table measurement data

The initial dataset contains (valid measurements):

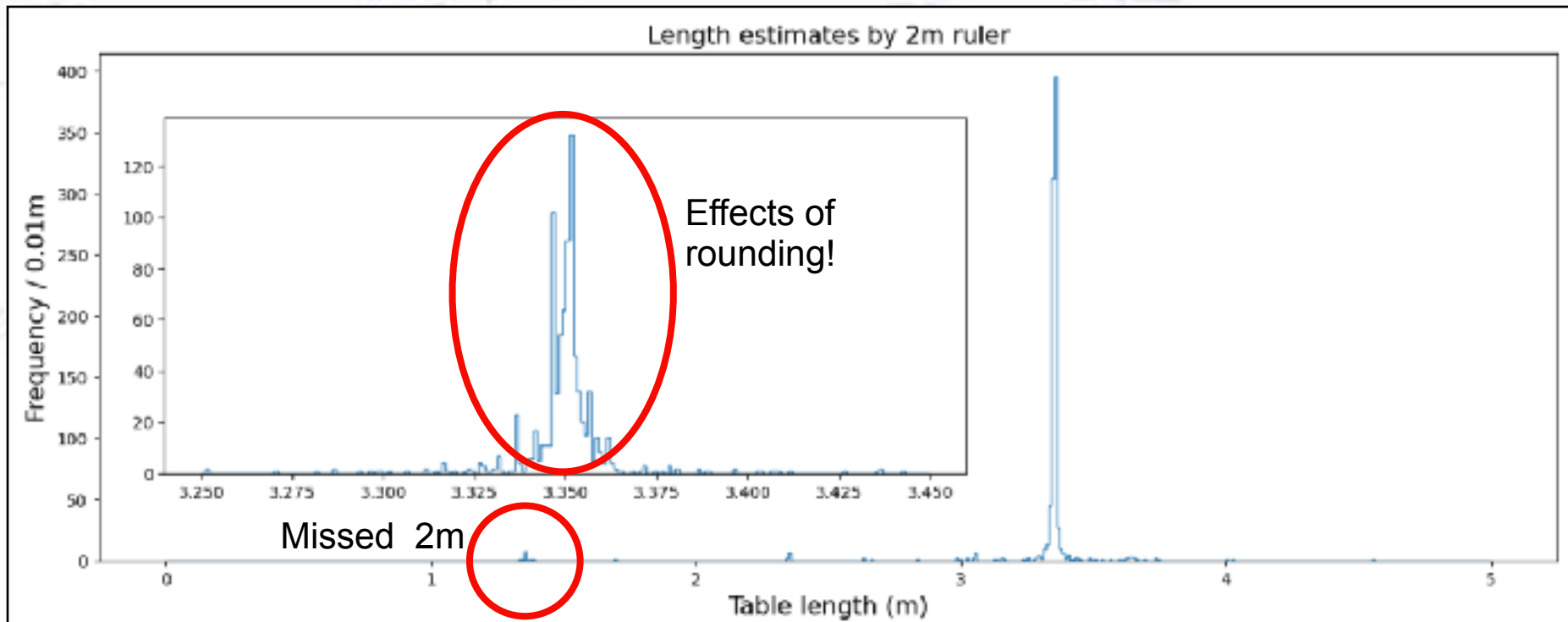
- 30cm measurements: **913** Range: [0.0, 5.0] m
- 2m measurements: **911** Range: [0.0, 5.0] m



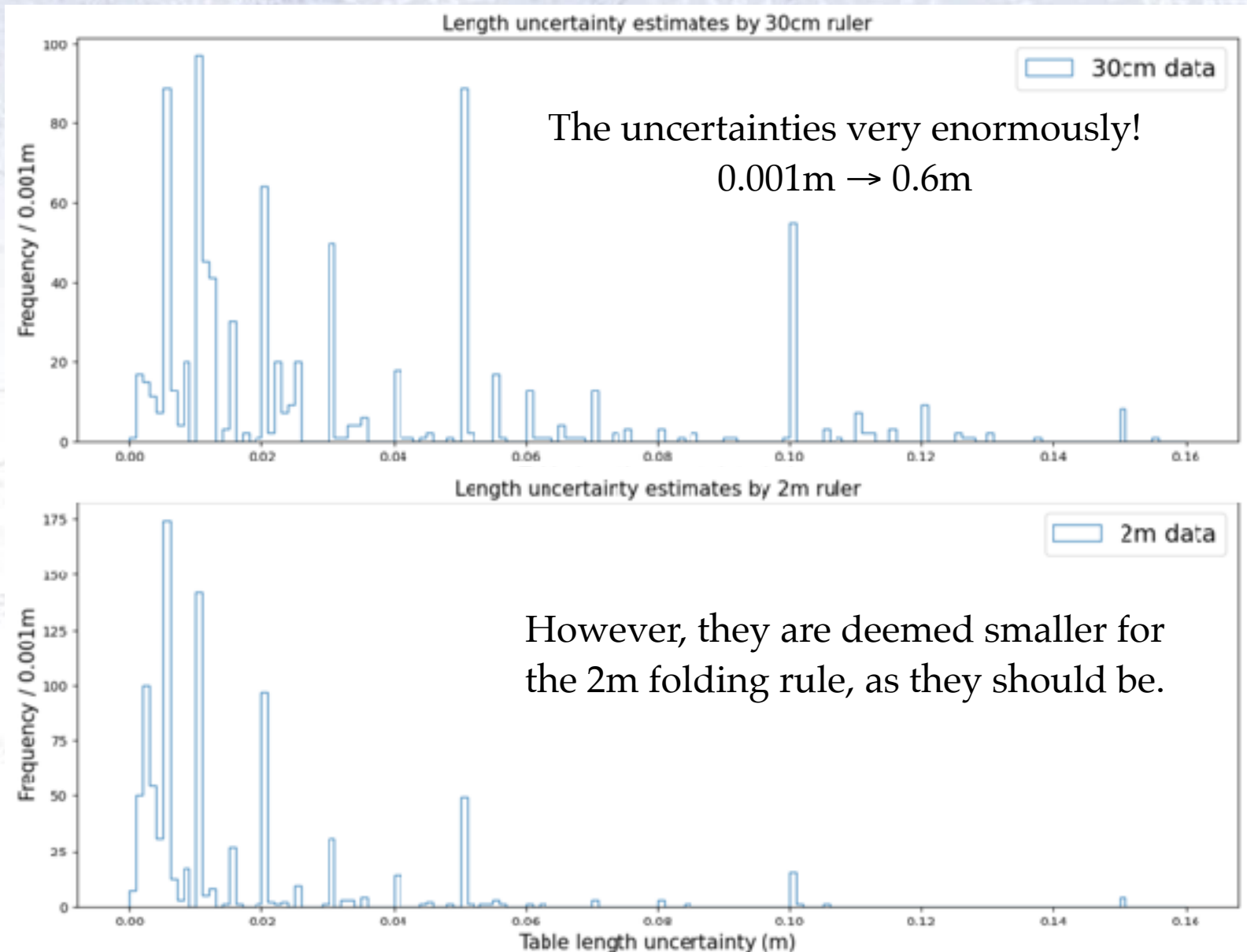
The table measurement data

The initial dataset contains (valid measurements):

- 30cm measurements: **913** Range: [0.0, 5.0] m
- 2m measurements: **911** Range: [0.0, 5.0] m



Uncertainties



Raw (“Naive”) results

30cm:

Mean = 3.3750 ± 0.0077 m

Std. = 0.23 m (N = 913)

2m:

Mean = 3.3155 ± 0.0086 m

Std. = 0.26 m (N = 911)

From the Std. values alone, it is clear that something is terribly wrong, which is also why the uncertainties on the mean are almost a centimeter!

The background is a faded map of the North Atlantic Ocean. It features magnetic isotherms, which are lines of equal magnetic declination, labeled with values such as 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1010, 1020, 1030, 1040, 1050, 1060, 1070, 1080, 1090, 1100, 1110, 1120, 1130, 1140, 1150, 1160, 1170, 1180, 1190, 1200, 1210, 1220, 1230, 1240, 1250, 1260, 1270, 1280, 1290, 1300, 1310, 1320, 1330, 1340, 1350, 1360, 1370, 1380, 1390, 1400, 1410, 1420, 1430, 1440, 1450, 1460, 1470, 1480, 1490, 1500, 1510, 1520, 1530, 1540, 1550, 1560, 1570, 1580, 1590, 1600, 1610, 1620, 1630, 1640, 1650, 1660, 1670, 1680, 1690, 1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010, 2020, 2030, 2040, 2050, 2060, 2070, 2080, 2090, 2100, 2110, 2120, 2130, 2140, 2150, 2160, 2170, 2180, 2190, 2200, 2210, 2220, 2230, 2240, 2250, 2260, 2270, 2280, 2290, 2300, 2310, 2320, 2330, 2340, 2350, 2360, 2370, 2380, 2390, 2400, 2410, 2420, 2430, 2440, 2450, 2460, 2470, 2480, 2490, 2500, 2510, 2520, 2530, 2540, 2550, 2560, 2570, 2580, 2590, 2600, 2610, 2620, 2630, 2640, 2650, 2660, 2670, 2680, 2690, 2700, 2710, 2720, 2730, 2740, 2750, 2760, 2770, 2780, 2790, 2800, 2810, 2820, 2830, 2840, 2850, 2860, 2870, 2880, 2890, 2900, 2910, 2920, 2930, 2940, 2950, 2960, 2970, 2980, 2990, 3000, 3010, 3020, 3030, 3040, 3050, 3060, 3070, 3080, 3090, 3100, 3110, 3120, 3130, 3140, 3150, 3160, 3170, 3180, 3190, 3200, 3210, 3220, 3230, 3240, 3250, 3260, 3270, 3280, 3290, 3300, 3310, 3320, 3330, 3340, 3350, 3360, 3370, 3380, 3390, 3400, 3410, 3420, 3430, 3440, 3450, 3460, 3470, 3480, 3490, 3500, 3510, 3520, 3530, 3540, 3550, 3560, 3570, 3580, 3590, 3600, 3610, 3620, 3630, 3640, 3650, 3660, 3670, 3680, 3690, 3700, 3710, 3720, 3730, 3740, 3750, 3760, 3770, 3780, 3790, 3800, 3810, 3820, 3830, 3840, 3850, 3860, 3870, 3880, 3890, 3900, 3910, 3920, 3930, 3940, 3950, 3960, 3970, 3980, 3990, 4000, 4010, 4020, 4030, 4040, 4050, 4060, 4070, 4080, 4090, 4100, 4110, 4120, 4130, 4140, 4150, 4160, 4170, 4180, 4190, 4200, 4210, 4220, 4230, 4240, 4250, 4260, 4270, 4280, 4290, 4300, 4310, 4320, 4330, 4340, 4350, 4360, 4370, 4380, 4390, 4400, 4410, 4420, 4430, 4440, 4450, 4460, 4470, 4480, 4490, 4500, 4510, 4520, 4530, 4540, 4550, 4560, 4570, 4580, 4590, 4600, 4610, 4620, 4630, 4640, 4650, 4660, 4670, 4680, 4690, 4700, 4710, 4720, 4730, 4740, 4750, 4760, 4770, 4780, 4790, 4800, 4810, 4820, 4830, 4840, 4850, 4860, 4870, 4880, 4890, 4900, 4910, 4920, 4930, 4940, 4950, 4960, 4970, 4980, 4990, 5000, 5010, 5020, 5030, 5040, 5050, 5060, 5070, 5080, 5090, 5100, 5110, 5120, 5130, 5140, 5150, 5160, 5170, 5180, 5190, 5200, 5210, 5220, 5230, 5240, 5250, 5260, 5270, 5280, 5290, 5300, 5310, 5320, 5330, 5340, 5350, 5360, 5370, 5380, 5390, 5400, 5410, 5420, 5430, 5440, 5450, 5460, 5470, 5480, 5490, 5500, 5510, 5520, 5530, 5540, 5550, 5560, 5570, 5580, 5590, 5600, 5610, 5620, 5630, 5640, 5650, 5660, 5670, 5680, 5690, 5700, 5710, 5720, 5730, 5740, 5750, 5760, 5770, 5780, 5790, 5800, 5810, 5820, 5830, 5840, 5850, 5860, 5870, 5880, 5890, 5900, 5910, 5920, 5930, 5940, 5950, 5960, 5970, 5980, 5990, 6000, 6010, 6020, 6030, 6040, 6050, 6060, 6070, 6080, 6090, 6100, 6110, 6120, 6130, 6140, 6150, 6160, 6170, 6180, 6190, 6200, 6210, 6220, 6230, 6240, 6250, 6260, 6270, 6280, 6290, 6300, 6310, 6320, 6330, 6340, 6350, 6360, 6370, 6380, 6390, 6400, 6410, 6420, 6430, 6440, 6450, 6460, 6470, 6480, 6490, 6500, 6510, 6520, 6530, 6540, 6550, 6560, 6570, 6580, 6590, 6600, 6610, 6620, 6630, 6640, 6650, 6660, 6670, 6680, 6690, 6700, 6710, 6720, 6730, 6740, 6750, 6760, 6770, 6780, 6790, 6800, 6810, 6820, 6830, 6840, 6850, 6860, 6870, 6880, 6890, 6900, 6910, 6920, 6930, 6940, 6950, 6960, 6970, 6980, 6990, 7000, 7010, 7020, 7030, 7040, 7050, 7060, 7070, 7080, 7090, 7100, 7110, 7120, 7130, 7140, 7150, 7160, 7170, 7180, 7190, 7200, 7210, 7220, 7230, 7240, 7250, 7260, 7270, 7280, 7290, 7300, 7310, 7320, 7330, 7340, 7350, 7360, 7370, 7380, 7390, 7400, 7410, 7420, 7430, 7440, 7450, 7460, 7470, 7480, 7490, 7500, 7510, 7520, 7530, 7540, 7550, 7560, 7570, 7580, 7590, 7600, 7610, 7620, 7630, 7640, 7650, 7660, 7670, 7680, 7690, 7700, 7710, 7720, 7730, 7740, 7750, 7760, 7770, 7780, 7790, 7800, 7810, 7820, 7830, 7840, 7850, 7860, 7870, 7880, 7890, 7900, 7910, 7920, 7930, 7940, 7950, 7960, 7970, 7980, 7990, 8000, 8010, 8020, 8030, 8040, 8050, 8060, 8070, 8080, 8090, 8100, 8110, 8120, 8130, 8140, 8150, 8160, 8170, 8180, 8190, 8200, 8210, 8220, 8230, 8240, 8250, 8260, 8270, 8280, 8290, 8300, 8310, 8320, 8330, 8340, 8350, 8360, 8370, 8380, 8390, 8400, 8410, 8420, 8430, 8440, 8450, 8460, 8470, 8480, 8490, 8500, 8510, 8520, 8530, 8540, 8550, 8560, 8570, 8580, 8590, 8600, 8610, 8620, 8630, 8640, 8650, 8660, 8670, 8680, 8690, 8700, 8710, 8720, 8730, 8740, 8750, 8760, 8770, 8780, 8790, 8800, 8810, 8820, 8830, 8840, 8850, 8860, 8870, 8880, 8890, 8900, 8910, 8920, 8930, 8940, 8950, 8960, 8970, 8980, 8990, 9000, 9010, 9020, 9030, 9040, 9050, 9060, 9070, 9080, 9090, 9100, 9110, 9120, 9130, 9140, 9150, 9160, 9170, 9180, 9190, 9200, 9210, 9220, 9230, 9240, 9250, 9260, 9270, 9280, 9290, 9300, 9310, 9320, 9330, 9340, 9350, 9360, 9370, 9380, 9390, 9400, 9410, 9420, 9430, 9440, 9450, 9460, 9470, 9480, 9490, 9500, 9510, 9520, 9530, 9540, 9550, 9560, 9570, 9580, 9590, 9600, 9610, 9620, 9630, 9640, 9650, 9660, 9670, 9680, 9690, 9700, 9710, 9720, 9730, 9740, 9750, 9760, 9770, 9780, 9790, 9800, 9810, 9820, 9830, 9840, 9850, 9860, 9870, 9880, 9890, 9900, 9910, 9920, 9930, 9940, 9950, 9960, 9970, 9980, 9990, 10000. The text "MAGNETIC" is visible in the upper left. In the upper right, there is a small rectangular box containing the text "181 BITE-4 EWA" and "YACHT-4123".

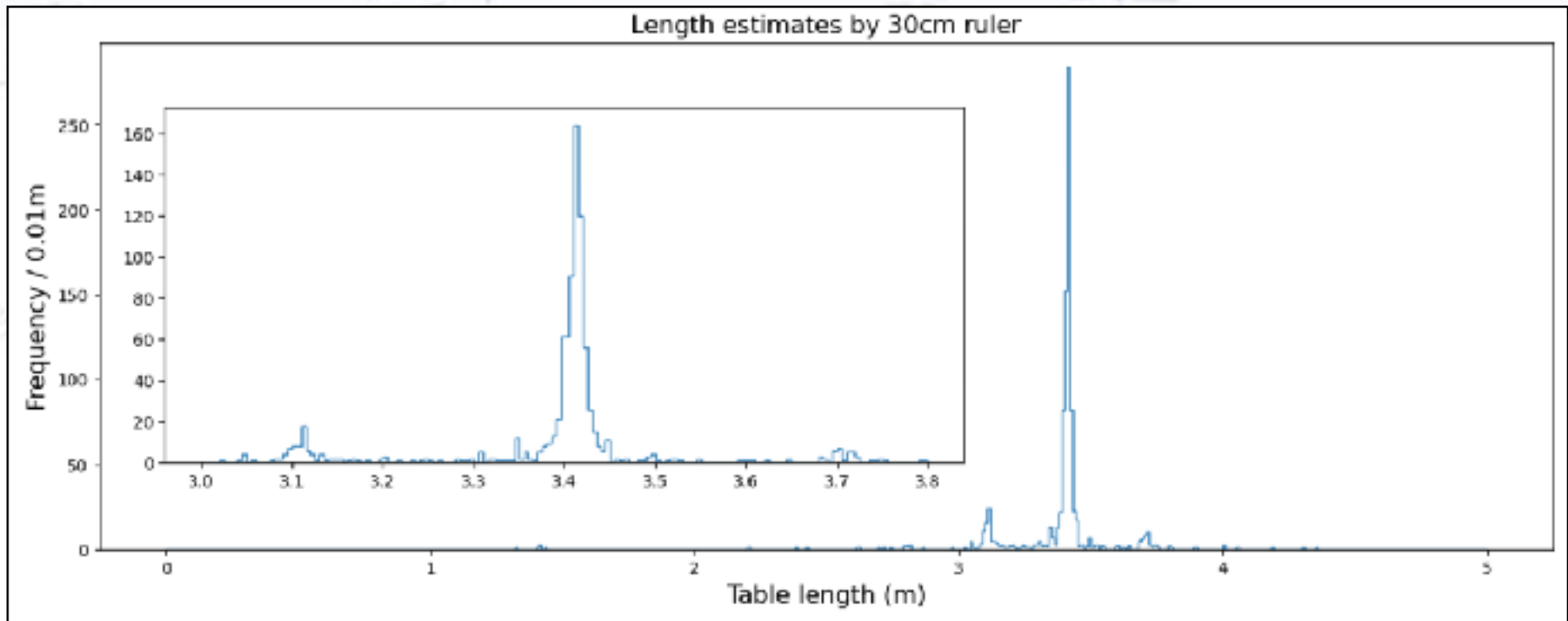
Unweighted analysis

Include offsets?

There are some clear and understandable mis-measurements.

Should one correct and include these? Or reject the values?

Depends on situation, but decide without seeing the final result.

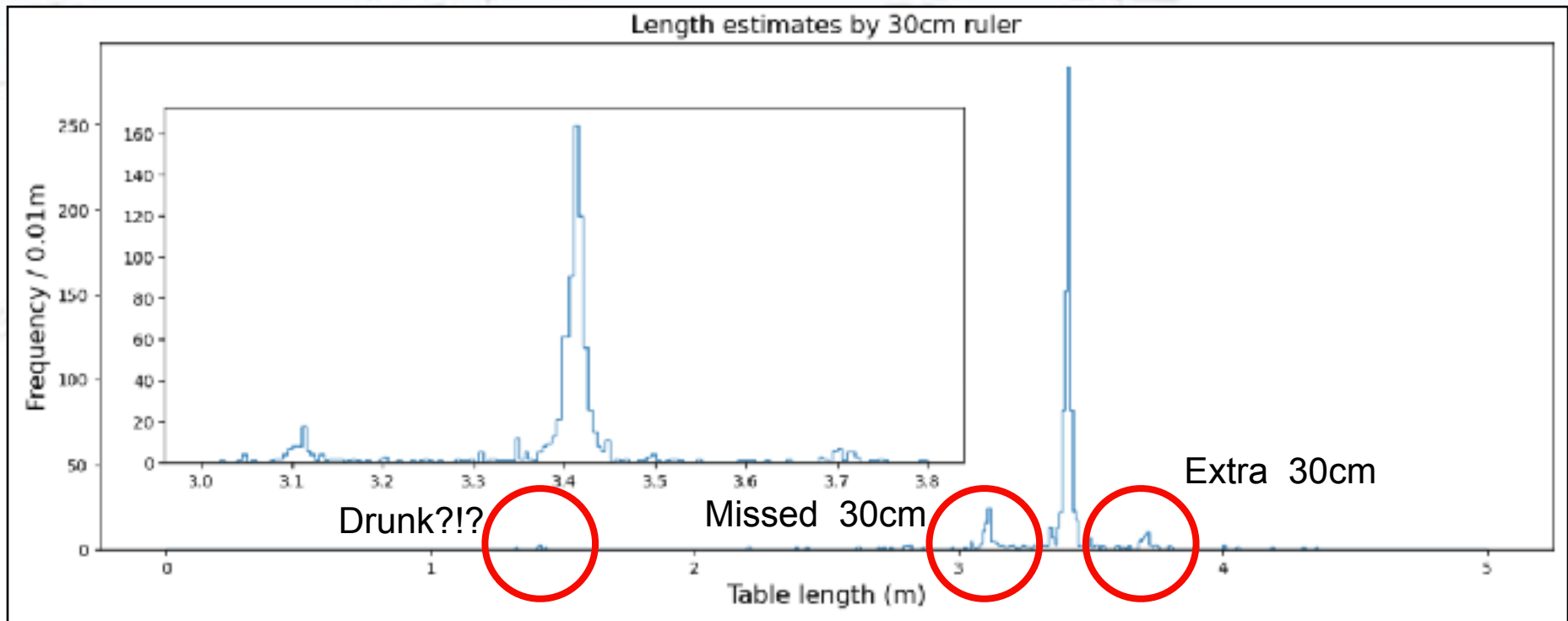


Include offsets?

There are some clear and understandable mis-measurements.

Should one correct and include these? Or reject the values?

Depends on situation, but decide without seeing the final result.

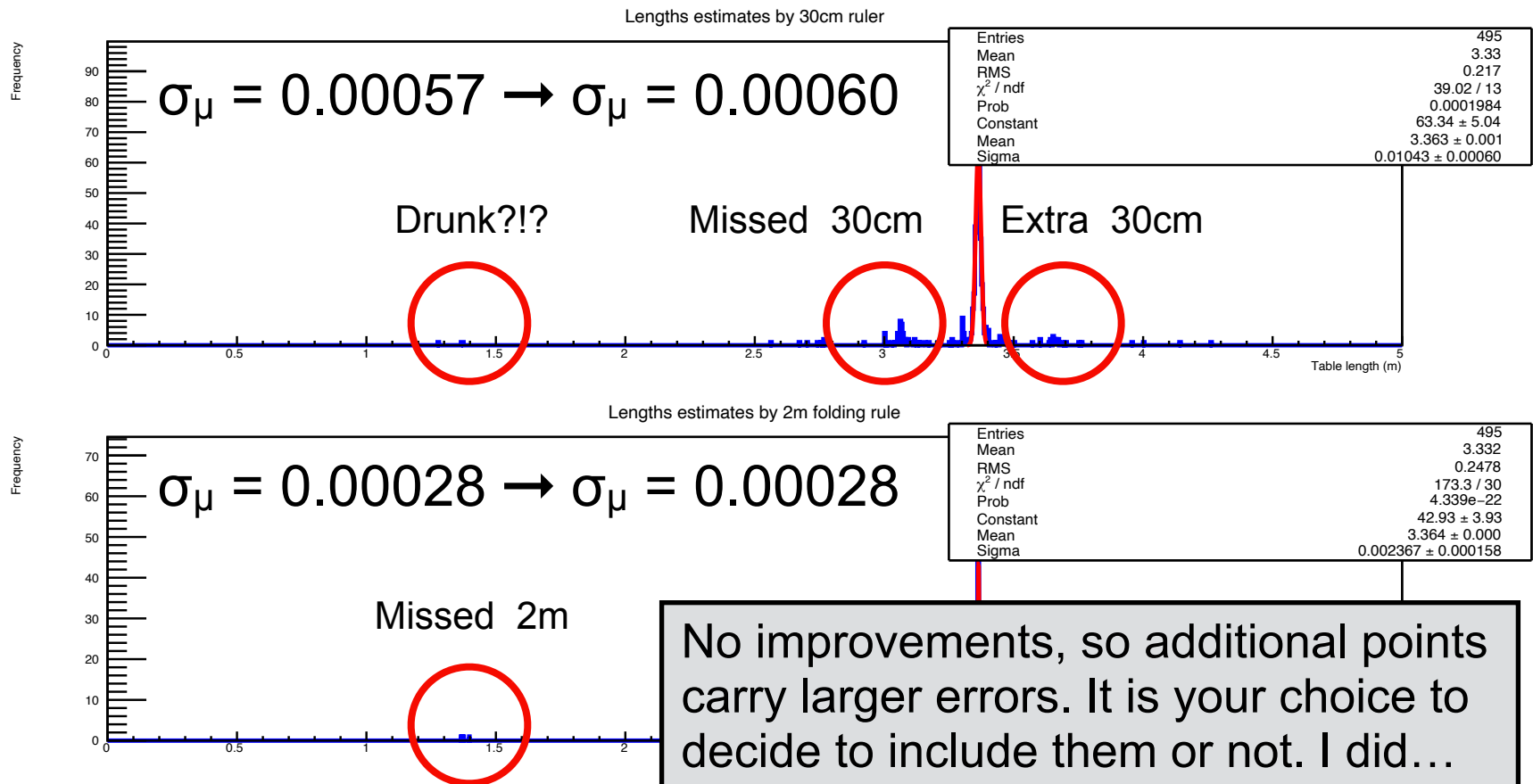


Include offsets?

There are some clear and understandable mis-measurements.

Should one correct and include these?

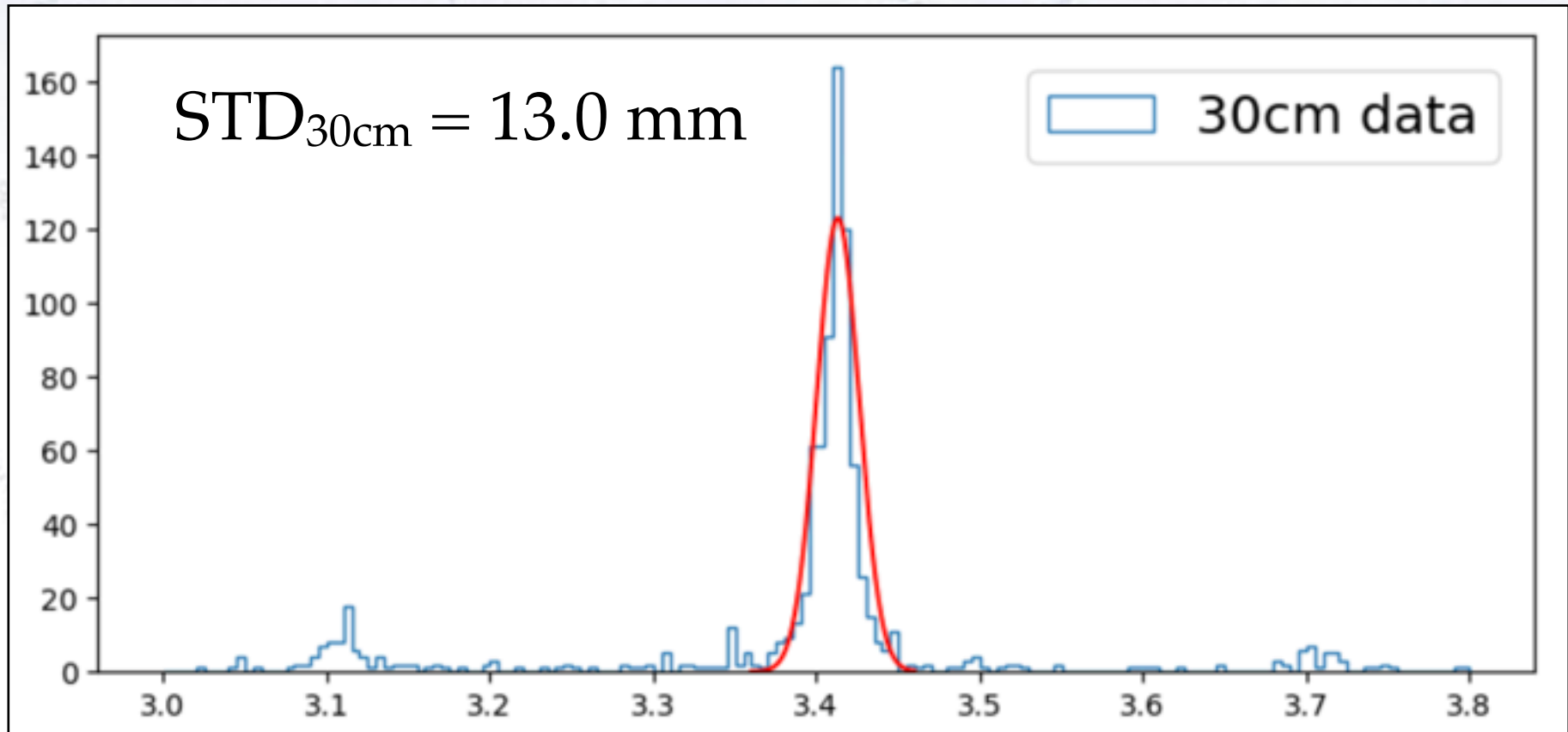
Depends on resulting improvement, but decide without seeing the final result.



Inspecting the data

The 30cm peak seems somewhat Gaussian ($p=2.4\%$) with binning 0.005m (smaller binning shows discontinuities, i.e. gives peaks).

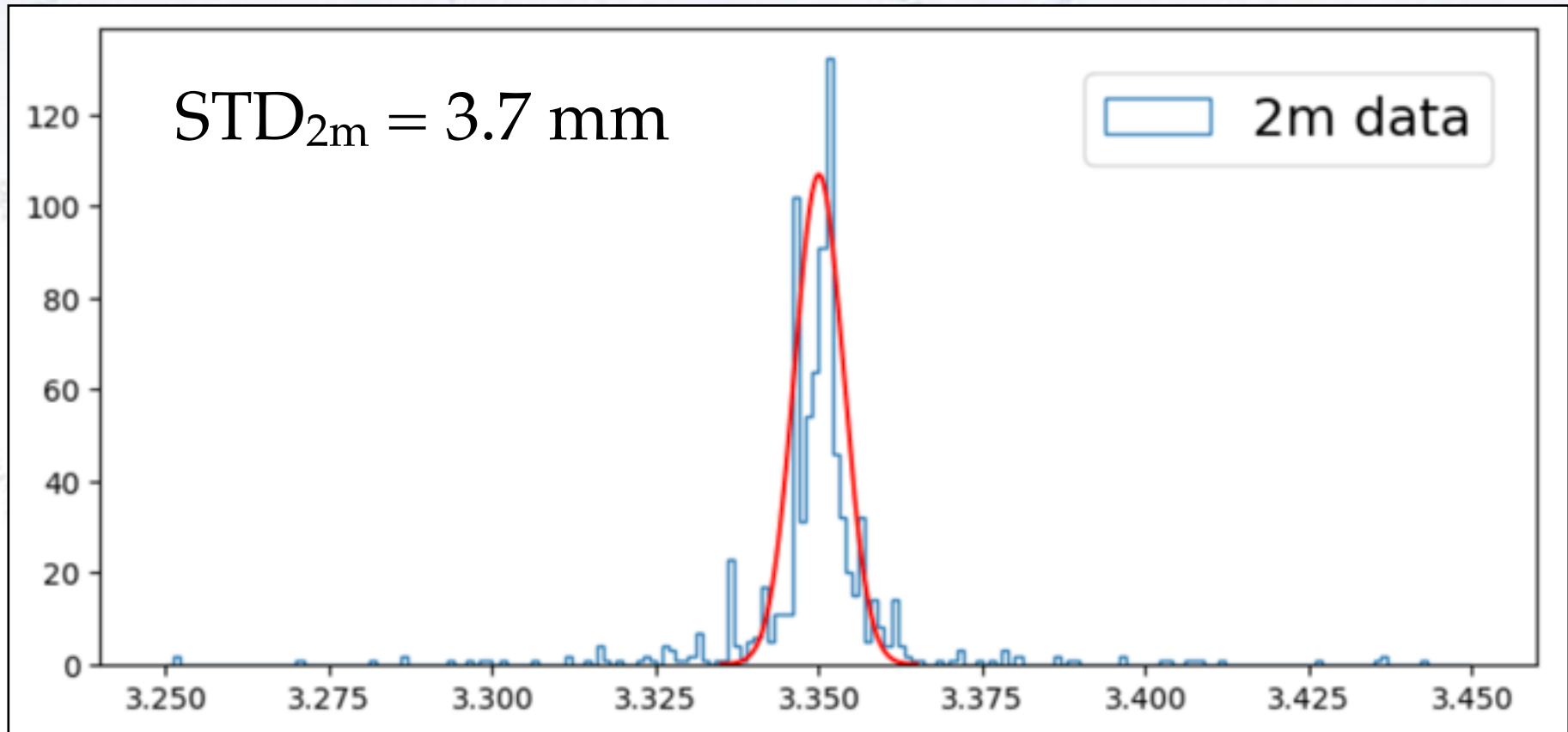
The 2m peak does not seem Gaussian with any binning (here 0.005), yet “collected”.



Inspecting the data

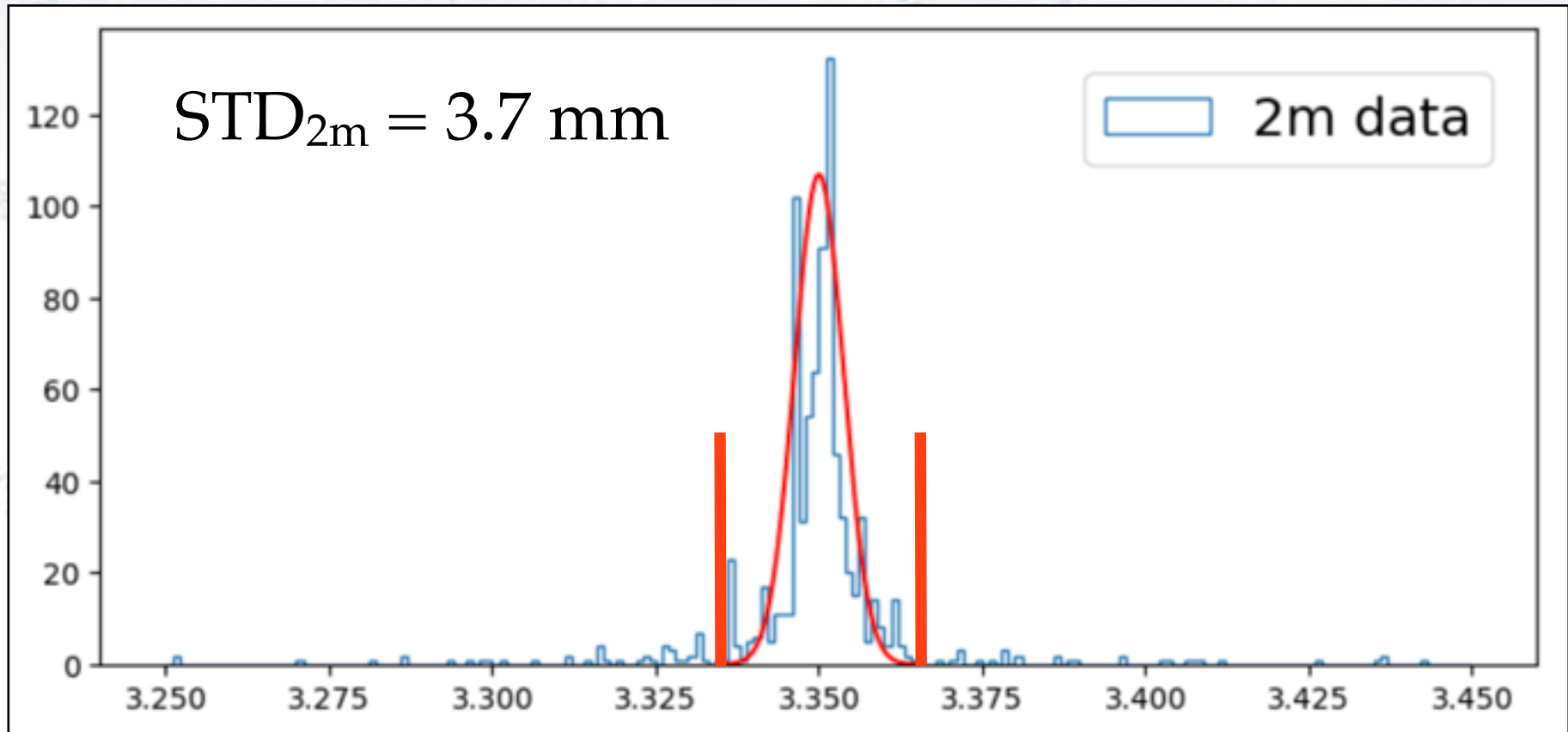
The 30cm peak seems somewhat Gaussian ($p=2.4\%$) with binning 0.005m (smaller binning shows discontinuities, i.e. gives peaks).

The 2m peak does not seem Gaussian with any binning (here 0.005), 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 $4 \times \text{STD}$:



Removing data - General

Some (very “purist”) scientists would never allow for the reject of data points! They would argue, that data reflects reality, and that one should simply model this, including imperfections.

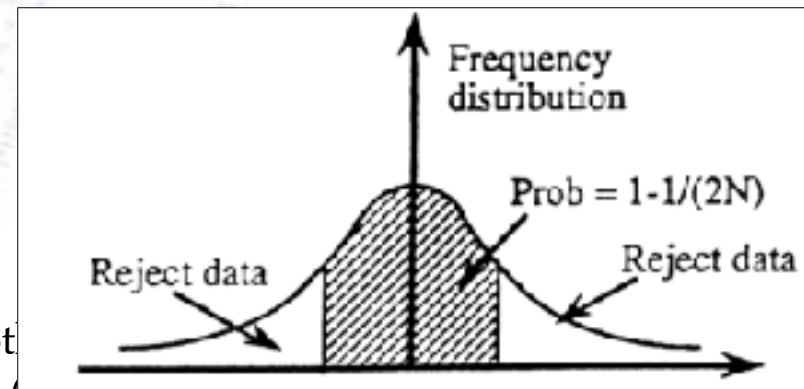
Less “purist” scientists accept exclusion of some data points. However, **one should always be very careful about removing data points**, and only be willing to do so, if very good arguments can be found:

- It is clearly an erroneous measurement.
- Measurement is highly improbable.

Preferably, one would like to understand why data points seem faulty.

The procedures for removing points are:

- Without errors: **Chauvenet's Criterion**, though ot
- With errors: Simply reject based on the z-value = $(x - \mu) / \sigma$ or the point



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

Removing data - without errors

Removing improbable data points when no error is given is formalised in **Chauvenet's Criterion**, though many other methods exists (Pierce, Grubbs, etc.)

The overall idea is to assume that the distribution is Gaussian.

One calculates the mean (μ) and standard deviation (σ), and then:

1. Ask what the probability of the farthest point is (given the number of points)
2. Remove point, if it is below some value (e.g. 0.05, preferably decided in advance)
3. If the furthest point was removed, then recalculate μ and σ , and go to 1.

How to calculate the probability of the furthest point with value x (given μ and σ)?

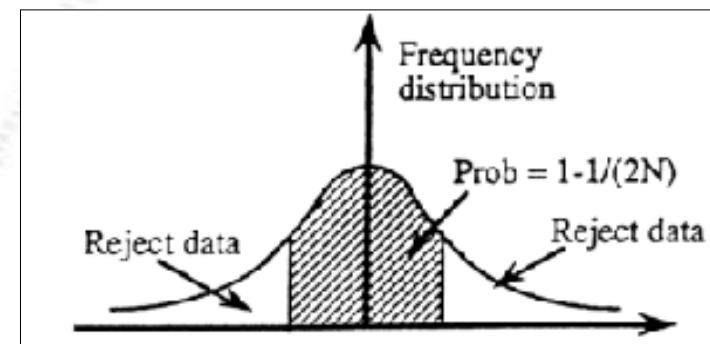
1. Calculate z :
$$z = (x - \mu) / \sigma$$

2. Find the probability of this z , p_{local} :

$$p_{\text{local}} = \int_{-\infty}^{-z} G(z) dz + \int_z^{\infty} G(z) dz$$

3. Take number of point into account, to get p_{global} :

$$p_{\text{global}} = 1 - (1 - p_{\text{local}})^{N_{\text{points}}}$$



Key question:
Is $p_{\text{global}} < 0.05$?

Removing data - with errors

Removing improbable data points when each point has an associated uncertainty is much simpler.

The overall idea is that all points should be consistent with a mean value.

One calculates the weighted mean (μ), and then removes all points that are more than z_{cut} sigma away. Done!

No iterative procedure is needed. One can calculate the value of z_{cut} ahead of applying it as:

$$p_{\text{local}} = 1 - (1 - p_{\text{global}})^{1/N_{\text{points}}}$$

Example:

You have 1000 measurements, all with uncertainties, and decide to discard all points which are less likely than $p_{\text{global}} = 0.05$. This yields a cut at $p_{\text{local}} = 0.000051$ or 4.05σ . Thus, one would reject all data, which are more than 4.05σ away from the mean.

```
p_local = 1.0 - (1.0 - p_global)**(1.0/Ndata)
Nsigma = np.abs(stats.norm.ppf(p_local/2.0))
```


...a fair hearing?

Chauvenet's Criterion (30cm)									
600: L=1.325	dL=2.050	Nsig= 8.85	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=913	mean=3.3750	std=0.2317	-> Rejected
61: L=1.405	dL=1.973	Nsig= 8.90	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=912	mean=3.3773	std=0.2216	-> Rejected
97: L=1.413	dL=1.967	Nsig= 9.28	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=911	mean=3.3795	std=0.2119	-> Rejected
552: L=1.415	dL=1.967	Nsig= 9.75	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=910	mean=3.3816	std=0.2017	-> Rejected
829: L=1.420	dL=1.964	Nsig=10.28	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=909	mean=3.3838	std=0.1910	-> Rejected
189: L=1.434	dL=1.952	Nsig=10.87	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=908	mean=3.3859	std=0.1796	-> Rejected
162: L=2.206	dL=1.182	Nsig= 7.05	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=907	mean=3.3881	std=0.1676	-> Rejected
261: L=2.383	dL=1.007	Nsig= 6.18	p_loc=0.00000000	p_glob=0.00000030	>? pmin=0.100	N=906	mean=3.3894	std=0.1630	-> Rejected
30: L=2.422	dL=0.969	Nsig= 6.07	p_loc=0.00000000	p_glob=0.00000058	>? pmin=0.100	N=905	mean=3.3905	std=0.1596	-> Rejected
110: L=4.355	dL=0.963	Nsig= 6.16	p_loc=0.00000000	p_glob=0.00000034	>? pmin=0.100	N=904	mean=3.3916	std=0.1564	-> Rejected
826: L=4.310	dL=0.919	Nsig= 6.00	p_loc=0.00000000	p_glob=0.00000090	>? pmin=0.100	N=903	mean=3.3925	std=0.1532	-> Rejected
786: L=4.190	dL=0.800	Nsig= 5.33	p_loc=0.00000005	p_glob=0.00004501	>? pmin=0.100	N=902	mean=3.3895	std=0.1502	-> Rejected
773: L=2.613	dL=0.776	Nsig= 5.25	p_loc=0.00000008	p_glob=0.00006997	>? pmin=0.100	N=901	mean=3.3886	std=0.1479	-> Rejected
119: L=2.625	dL=0.765	Nsig= 5.25	p_loc=0.00000008	p_glob=0.00006887	>? pmin=0.100	N=900	mean=3.3895	std=0.1457	-> Rejected
212: L=2.700	dL=0.691	Nsig= 4.81	p_loc=0.00000075	p_glob=0.00067274	>? pmin=0.100	N=899	mean=3.3903	std=0.1435	-> Rejected
768: L=2.720	dL=0.671	Nsig= 4.74	p_loc=0.00000109	p_glob=0.00097587	>? pmin=0.100	N=898	mean=3.3911	std=0.1418	-> Rejected
599: L=4.054	dL=0.662	Nsig= 4.73	p_loc=0.00000115	p_glob=0.00102772	>? pmin=0.100	N=897	mean=3.3918	std=0.1400	-> Rejected
496: L=2.750	dL=0.641	Nsig= 4.64	p_loc=0.00000178	p_glob=0.00159297	>? pmin=0.100	N=896	mean=3.3911	std=0.1384	-> Rejected
294: L=4.010	dL=0.618	Nsig= 4.52	p_loc=0.00000313	p_glob=0.00279838	>? pmin=0.100	N=895	mean=3.3918	std=0.1368	-> Rejected
179: L=4.009	dL=0.618	Nsig= 4.56	p_loc=0.00000250	p_glob=0.00223068	>? pmin=0.100	N=894	mean=3.3911	std=0.1353	-> Rejected
601: L=4.005	dL=0.614	Nsig= 4.59	p_loc=0.00000220	p_glob=0.00195965	>? pmin=0.100	N=893	mean=3.3904	std=0.1338	-> Rejected
445: L=2.794	dL=0.596	Nsig= 4.51	p_loc=0.00000329	p_glob=0.00292612	>? pmin=0.100	N=892	mean=3.3897	std=0.1323	-> Rejected
350: L=2.800	dL=0.591	Nsig= 4.52	p_loc=0.00000315	p_glob=0.00280203	>? pmin=0.100	N=891	mean=3.3904	std=0.1308	-> Rejected
241: L=2.808	dL=0.583	Nsig= 4.51	p_loc=0.00000325	p_glob=0.00288751	>? pmin=0.100	N=890	mean=3.3911	std=0.1294	-> Rejected
423: L=2.809	dL=0.583	Nsig= 4.56	p_loc=0.00000260	p_glob=0.00230684	>? pmin=0.100	N=889	mean=3.3917	std=0.1280	-> Rejected
16: L=2.815	dL=0.578	Nsig= 4.57	p_loc=0.00000248	p_glob=0.00220346	>? pmin=0.100	N=888	mean=3.3924	std=0.1265	-> Rejected
654: L=2.818	dL=0.575	Nsig= 4.60	p_loc=0.00000212	p_glob=0.00187759	>? pmin=0.100	N=887	mean=3.3930	std=0.1251	-> Rejected
557: L=2.819	dL=0.575	Nsig= 4.65	p_loc=0.00000166	p_glob=0.00147061	>? pmin=0.100	N=886	mean=3.3937	std=0.1237	-> Rejected
169: L=2.864	dL=0.531	Nsig= 4.34	p_loc=0.00000705	p_glob=0.00522428	>? pmin=0.100	N=885	mean=3.3943	std=0.1222	-> Rejected
205: L=3.891	dL=0.496	Nsig= 4.10	p_loc=0.00002005	p_glob=0.01825865	>? pmin=0.100	N=884	mean=3.3949	std=0.1210	-> Rejected
700: L=2.971	dL=0.424	Nsig= 3.53	p_loc=0.00020441	p_glob=0.16515717	>? pmin=0.100	N=883	mean=3.3944	std=0.1199	-> Accepted
Chauvenet's Criterion (2m)									
191: L=1.337	dL=1.978	Nsig= 7.64	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=911	mean=3.3154	std=0.2590	-> Rejected
400: L=1.341	dL=1.976	Nsig= 7.80	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=910	mean=3.3176	std=0.2507	-> Rejected
98: L=1.350	dL=1.970	Nsig= 8.14	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=909	mean=3.3197	std=0.2421	-> Rejected
61: L=1.351	dL=1.971	Nsig= 8.45	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=908	mean=3.3219	std=0.2332	-> Rejected
199: L=1.351	dL=1.973	Nsig= 8.81	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=907	mean=3.3241	std=0.2240	-> Rejected
894: L=1.351	dL=1.975	Nsig= 9.22	p_loc=0.00000000	p_glob=0.00000000	>? pmin=0.100	N=906	mean=3.3263	std=0.2143	-> Rejected

...a fair hearing?

Chauvenet's Criterion (30cm)											
600:	L=1.325	dL=2.050	Nsig= 8.85	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=913	mean=3.3750	std=0.2317	-> Rejected
61:	L=1.405	dL=1.973	Nsig= 8.90	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=912	mean=3.3773	std=0.2216	-> Rejected
97:	L=1.413	dL=1.967	Nsig= 9.28	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=911	mean=3.3795	std=0.2119	-> Rejected
552:	L=1.415	dL=1.967	Nsig= 9.75	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=910	mean=3.3816	std=0.2017	-> Rejected
829:	L=1.420	dL=1.964	Nsig=10.28	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=909	mean=3.3838	std=0.1910	-> Rejected
189:	L=1.434	dL=1.952	Nsig=10.87	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=908	mean=3.3859	std=0.1796	-> Rejected
162:	L=2.206	dL=1.182	Nsig= 7.05	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=907	mean=3.3881	std=0.1676	-> Rejected
261:	L=2.383	dL=1.007	Nsig= 6.18	p_loc=0.00000000	p_glob=0.00000030	>?	pmin=0.100	N=906	mean=3.3894	std=0.1630	-> Rejected
30:	L=2.422	dL=0.969	Nsig= 6.07	p_loc=0.00000000	p_glob=0.00000058	>?	pmin=0.100	N=905	mean=3.3905	std=0.1596	-> Rejected
110:	L=4.355	dL=0.963	Nsig= 6.16	p_loc=0.00000000	p_glob=0.00000034	>?	pmin=0.100	N=904	mean=3.3916	std=0.1564	-> Rejected
826:	L=4.310	dL=0.919	Nsig= 6.00	p_loc=0.00000000	p_glob=0.00000090	>?	pmin=0.100	N=903	mean=3.3925	std=0.1532	-> Rejected
786:	L=4.190	dL=0.800	Nsig= 5.33	p_loc=0.00000005	p_glob=0.00004501	>?	pmin=0.100	N=902	mean=3.3895	std=0.1502	-> Rejected
773:	L=2.613	dL=0.776	Nsig= 5.25	p_loc=0.00000008	p_glob=0.00006907	>?	pmin=0.100	N=901	mean=3.3886	std=0.1479	-> Rejected
119:	L=2.625	dL=0.765								std=0.1457	-> Rejected
212:	L=2.700	dL=0.691								std=0.1435	-> Rejected
768:	L=2.720	dL=0.671								std=0.1418	-> Rejected
599:	L=4.054	dL=0.662								std=0.1400	-> Rejected
496:	L=2.750	dL=0.641								std=0.1384	-> Rejected
294:	L=4.010	dL=0.618								std=0.1368	-> Rejected
179:	L=4.009	dL=0.618								std=0.1353	-> Rejected
601:	L=4.005	dL=0.614								std=0.1338	-> Rejected
445:	L=2.794	dL=0.596								std=0.1323	-> Rejected
350:	L=2.800	dL=0.591								std=0.1308	-> Rejected
241:	L=2.808	dL=0.583								std=0.1294	-> Rejected
423:	L=2.809	dL=0.583								std=0.1280	-> Rejected
16:	L=2.815	dL=0.578	Nsig= 4.57	p_loc=0.00000248	p_glob=0.00220346	>?	pmin=0.100	N=888	mean=3.3924	std=0.1265	-> Rejected
654:	L=2.818	dL=0.575	Nsig= 4.60	p_loc=0.00000212	p_glob=0.00187759	>?	pmin=0.100	N=887	mean=3.3930	std=0.1251	-> Rejected
557:	L=2.819	dL=0.575	Nsig= 4.65	p_loc=0.00000166	p_glob=0.00147061	>?	pmin=0.100	N=886	mean=3.3937	std=0.1237	-> Rejected
169:	L=2.864	dL=0.531	Nsig= 4.34	p_loc=0.00000705	p_glob=0.00522428	>?	pmin=0.100	N=885	mean=3.3943	std=0.1222	-> Rejected
205:	L=3.891	dL=0.496	Nsig= 4.10	p_loc=0.00002085	p_glob=0.01825865	>?	pmin=0.100	N=884	mean=3.3949	std=0.1210	-> Rejected
700:	L=2.971	dL=0.424	Nsig= 3.53	p_loc=0.00020441	p_glob=0.16515717	>?	pmin=0.100	N=883	mean=3.3944	std=0.1199	-> Accepted
Chauvenet's Criterion (2m)											
191:	L=1.337	dL=1.978	Nsig= 7.64	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=911	mean=3.3154	std=0.2590	-> Rejected
400:	L=1.341	dL=1.976	Nsig= 7.80	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=910	mean=3.3176	std=0.2507	-> Rejected
98:	L=1.350	dL=1.970	Nsig= 8.14	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=909	mean=3.3197	std=0.2421	-> Rejected
61:	L=1.351	dL=1.971	Nsig= 8.45	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=908	mean=3.3219	std=0.2332	-> Rejected
199:	L=1.351	dL=1.973	Nsig= 8.81	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=907	mean=3.3241	std=0.2240	-> Rejected
894:	L=1.351	dL=1.975	Nsig= 9.22	p_loc=0.00000000	p_glob=0.00000000	>?	pmin=0.100	N=906	mean=3.3263	std=0.2143	-> Rejected

I rejected:
30 data points from the 30cm sample (*),
132 data points from the 2m sample.
And I inspected each and every one!

I rejected:
 30 data points from the 30cm sample (*),
 132 data points from the 2m sample.
 And I inspected each and every one!

Unweighted results

30cm:

Mean = 3.39438 ± 0.00404 m

Std. = 0.120 m (N = 882)

2m:

Mean = 3.34942 ± 0.00020 m

Std. = 0.0055 m (N = 779)

Without corrections and Chavenet's ($p=0.10$)

While the 2m result starts looking realistic (Std. of 5.5mm) and precise (1 / 5th of a millimeter), the 30cm result is still terrible.

This is because the two ± 30 cm peaks are not removed by Chavenet's Criterion.

30cm: Reject or correct?

The poor 30cm result is due to the (many) mis-measurements of $\pm 30\text{cm}$, which are not rejected by the Chauvenet's Criterion with $p_{\text{global}} = 0.10$.

At least two solutions exist:

1. Decide to **reject** all measurements more than 15cm away from the mean and run Chauvenet's Criterion again.
2. Decide to **correct** measurements 15-45cm away from the mean, and run Chauvenet's Criterion again.

Rejecting the events removes a total of $172 (\pm 15\text{cm}) + 64 (\text{CC})$ measurements. Correcting the events first removes a total of 91 measurements.

The uncertain on the mean resulting from each strategy is:

1. Rejection: 0.49 mm
2. Correction: 0.62 mm

So, it is worthwhile to focus on the good measurements, if this can be argued.

Unweighted results

30cm:

Mean = 3.41190 ± 0.00049 m

Std. = 0.0126 m (N = 677)

2m:

Mean = 3.34942 ± 0.00020 m

Std. = 0.0055 m (N = 779)

Without corrections and Chavenet's ($p=0.10$)

Now the results are precise, and the 2m result is about a factor 2.5 more so, as would also be expected from the initial Std. observed for the peaks.

The improvement over the naive 30cm / 2m results are factors of 19 / 43

Cross Check

Now we have gotten two precision results. How to cross check if there is any realism in the values and uncertainties?

We compare the 30cm and 2m results.

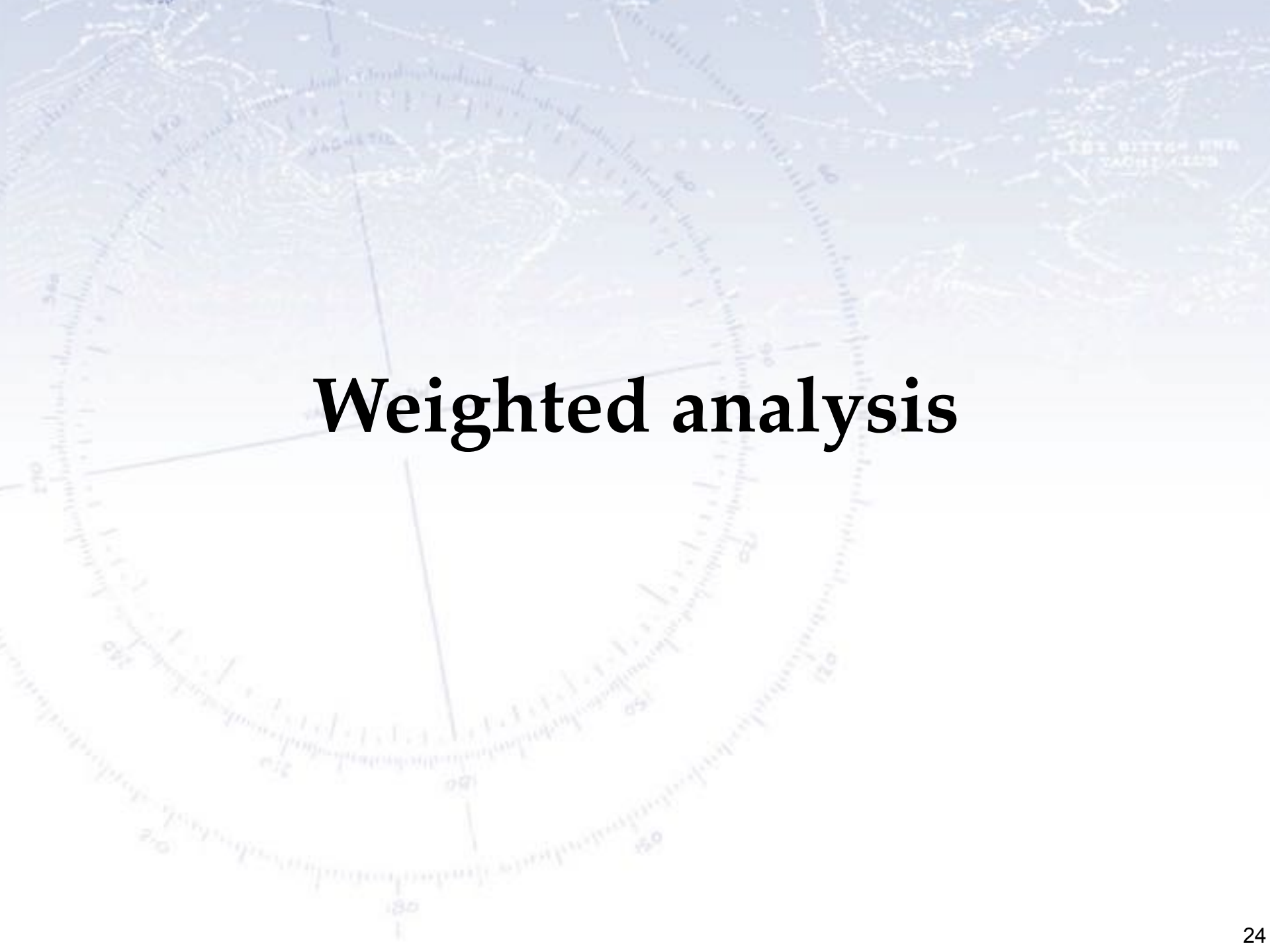
So far, the results have been blinded, and so the difference is very large:

$$\text{L30cm} - \text{L2m (fully blinded)} = 0.06248 \pm 0.00052 \text{ m} \quad (119.4\sigma \rightarrow \text{prob}=0.0000)$$

Subtracting the *difference in blinding value* yields the real difference:

$$\text{L30cm} - \text{L2m (partially blinded)} = \mathbf{-0.00102 \pm 0.00052 \text{ m}} \quad (2.0\sigma \rightarrow \text{prob}=0.0508)$$

This means that the two results are reasonably within each other, and the results and their uncertainty are (more) trustworthy.



Weighted analysis

24

Checking for valid errors

In order to do a weighted analyst, the measurements of course have to have valid uncertainties.

You may wonder why there are negative uncertainties!

The reason is, that this is a (good?) way of putting measurements without uncertainties, without putting NaNs into the table.

```
The 30cm entry L = 3.413 +- -1.000 was not considered valid!  
The 30cm entry L = 3.412 +- -1.000 was not considered valid!  
The 30cm entry L = 3.388 +- -1.000 was not considered valid!  
The 30cm entry L = 3.416 +- -1.000 was not considered valid!  
The 30cm entry L = 3.421 +- -1.000 was not considered valid!  
The 30cm entry L = 3.415 +- -1.000 was not considered valid!  
The 30cm entry L = 3.443 +- -1.000 was not considered valid!  
The 30cm entry L = 3.422 +- -1.000 was not considered valid!  
The 30cm entry L = 3.410 +- -1.000 was not considered valid!  
The 30cm entry L = 3.416 +- -1.000 was not considered valid!  
The 30cm entry L = 3.416 +- -1.000 was not considered valid!  
The 30cm entry L = 3.405 +- -1.000 was not considered valid!  
The 30cm entry L = 3.417 +- -1.000 was not considered valid!  
The 30cm entry L = 3.414 +- -1.000 was not considered valid!  
The 30cm entry L = 3.430 +- -1.000 was not considered valid!  
The 30cm entry L = 3.720 +- -1.000 was not considered valid!  
The number of accepted / rejected 30cm points is 897 / 16
```

```
The 2m entry L = 3.350 +- -1.000 was not considered valid!  
The 2m entry L = 3.361 +- -1.000 was not considered valid!  
The 2m entry L = 3.351 +- 0.000 was not considered valid!  
The 2m entry L = 3.350 +- -1.000 was not considered valid!  
The 2m entry L = 1.361 +- 0.000 was not considered valid!  
The 2m entry L = 3.342 +- -1.000 was not considered valid!  
The 2m entry L = 3.350 +- -1.000 was not considered valid!  
The 2m entry L = -1.014 +- -1.000 was not considered valid!  
The 2m entry L = 3.355 +- -1.000 was not considered valid!  
The 2m entry L = 3.353 +- -1.000 was not considered valid!  
The 2m entry L = 3.349 +- -1.000 was not considered valid!  
The 2m entry L = 3.356 +- -1.000 was not considered valid!  
The 2m entry L = 3.348 +- -1.000 was not considered valid!  
The 2m entry L = 3.350 +- -1.000 was not considered valid!  
The 2m entry L = -1.014 +- -0.010 was not considered valid!  
The 2m entry L = 3.336 +- 0.000 was not considered valid!  
The 2m entry L = 3.341 +- -0.010 was not considered valid!  
The 2m entry L = 3.346 +- -0.010 was not considered valid!  
The 2m entry L = 3.351 +- -0.010 was not considered valid!  
The 2m entry L = 3.426 +- -0.010 was not considered valid!  
The 2m entry L = 3.626 +- -0.010 was not considered valid!  
The number of accepted / rejected 2m points is 892 / 21
```

“Naive” weighted results

30cm:

Mean = 3.38764 ± 0.00017 m

RMS = undefined! (N = 896)

2m:

Mean = 3.31486 ± 0.00009 m

RMS = undefined! (N = 891)

Now the results are really precise, and the 2m result is about a factor 2.5 more so, as would also be expected from the initial Std. observed for the peaks.

The improvement over the naive 30cm / 2m results are factors of 19 / 43

“Naive” weighted results

30cm:

Mean = 3.38764 ± 0.00017 m

Chi2 = 1743866.2, Ndof = 896, Prob = 0.0!

2m:

Mean = 3.31486 ± 0.00009 m

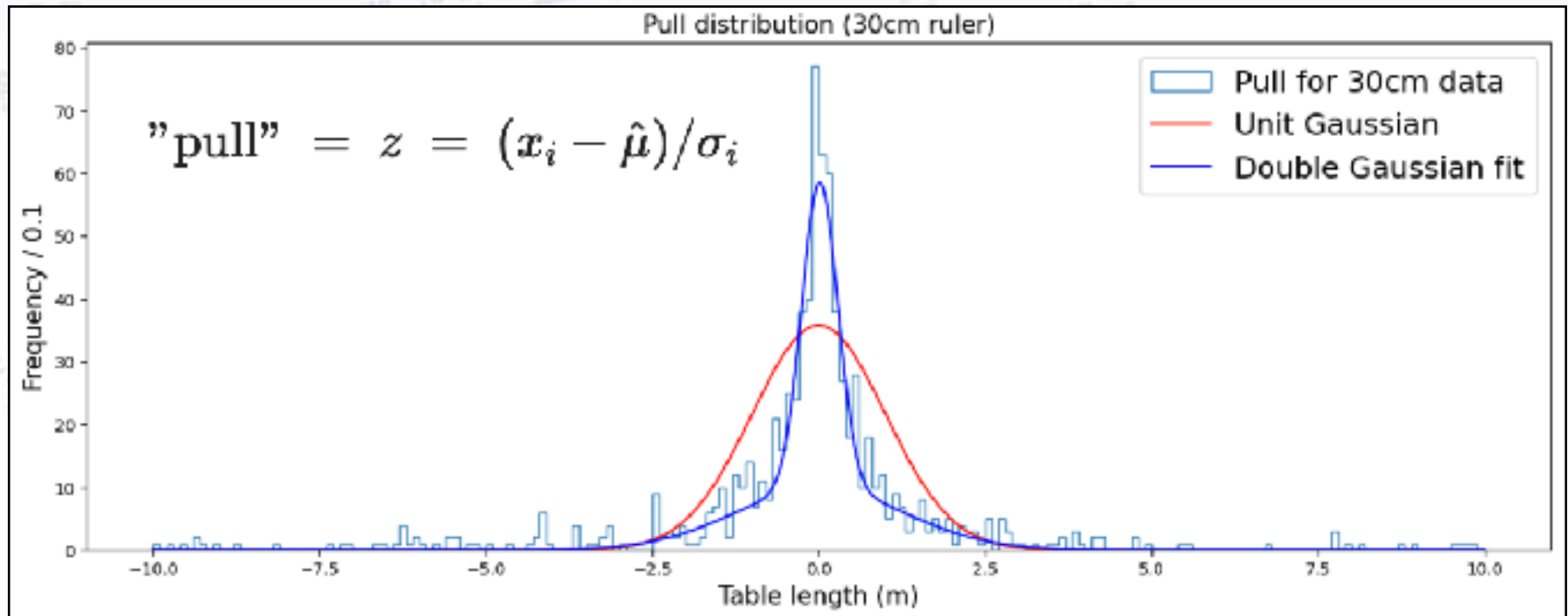
Chi2 = 7595341.9, Ndof = 891, Prob = 0.0!

While the values of the results may look "alluring" and the uncertainties amazingly small, the ChiSquare reveals that this is not the case. **The measurements disagree enormously when uncertainties are taken into account.** Clearly, the naive approach is way off.

The pull distribution

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-values.

The pulls should be unit Gaussian. However, it is far from. In fact, **most pull values are small, which is caused by an overestimation of the uncertainty.** We are too conservative and don't trust, that we can do things fairly accurately.

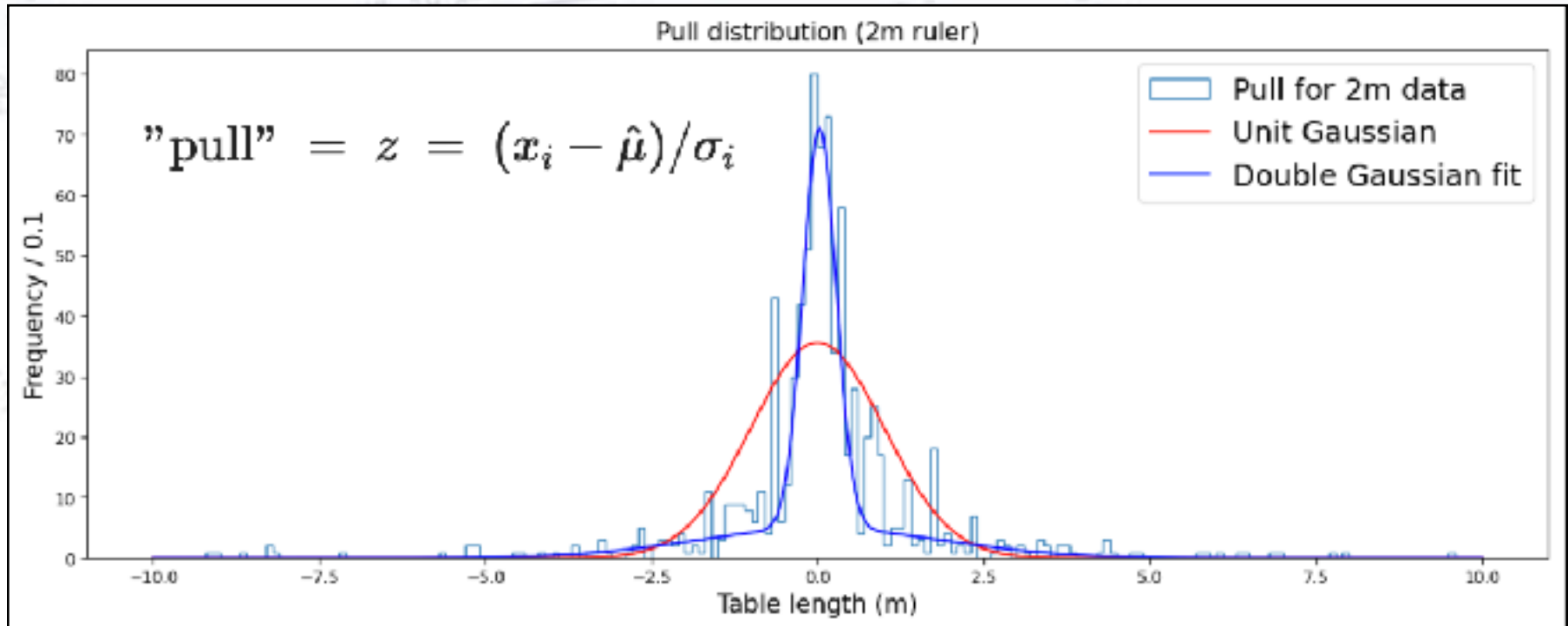


In the case at hand, we take the mean to be the unweighted best result.

The pull distribution

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-values.

The pulls should be unit Gaussian. However, it is far from. In fact, **most pull values are small, which is caused by an overestimation of the uncertainty.** We are too conservative and don't trust, that we can do things fairly accurately.



In the case at hand, we take the mean to be the unweighted best result.

Where to select?

Assuming unit Gaussian distributions (not the case, but still) we calculate at what level it is reasonable to discard individual measurement based on their z-value, i.e. how many sigmas they are away from the mean.

This only depends on the number of measurements and $p_{\text{global}} = 0.05$, and the result is 4.0σ for both 30cm and 2m data.

```
p_local = 1.0 - (1.0 - p_global)**(1.0/Ndata)
Nsigma = np.abs(stats.norm.ppf(p_local/2.0))
```

Given that the assumption is not really fulfilled, the real level should be set below this value, as many low z-values will make the ChiSquare unnaturally low. I chose 80% and 90% of the 4.0σ , i.e. 3.2σ and 3.6σ

Once the selection level is fitting, we then discard unlikely events (i.e. beyond a certain number of sigmas) and then proceed to calculate the weighted mean (with error, Chi2, and ProbChi2 of course!).

Excluded data due to bad pull

Warning! Large pull: L = 3.750 +- 0.013 m	z = 25.98	Warning! Large pull: L = 3.696 +- 0.050 m	z = 6.93
Warning! Large pull: L = 3.697 +- 0.020 m	z = 14.24	Warning! Large pull: L = 3.371 +- 0.085 m	z = 4.35
Warning! Large pull: L = 3.323 +- 0.009 m	z = -9.91	Warning! Large pull: L = 2.350 +- 0.015 m	z = -66.62
Warning! Large pull: L = 2.815 +- 0.091 m	z = -6.56	Warning! Large pull: L = 3.746 +- 0.020 m	z = 19.84
Warning! Large pull: L = 3.250 +- 0.002 m	z = -81.11	Warning! Large pull: L = 3.051 +- 0.002 m	z = -149.13
Warning! Large pull: L = 3.395 +- 0.005 m	z = -3.45	Warning! Large pull: L = 2.350 +- 0.001 m	z = -999.25
Warning! Large pull: L = 2.422 +- 0.005 m	z = -198.05	Warning! Large pull: L = 1.351 +- 0.020 m	z = -99.91
Warning! Large pull: L = 3.753 +- 0.010 m	z = 34.08	Warning! Large pull: L = 3.122 +- 0.025 m	z = -9.09
Warning! Large pull: L = 3.425 +- 0.003 m	z = 4.26	Warning! Large pull: L = 3.631 +- 0.006 m	z = 46.96
Warning! Large pull: L = 3.085 +- 0.070 m	z = -4.67	Warning! Large pull: L = 3.535 +- 0.005 m	z = 37.15
Warning! Large pull: L = 3.115 +- 0.023 m	z = -12.92	Warning! Large pull: L = 1.350 +- 0.010 m	z = -199.93
Warning! Large pull: L = 3.595 +- 0.005 m	z = 36.55	Warning! Large pull: L = 4.550 +- 0.023 m	z = 52.21
Warning! Large pull: L = 3.701 +- 0.030 m	z = 9.63	Warning! Large pull: L = 3.322 +- 0.001 m	z = -27.25
Warning! Large pull: L = 1.405 +- 0.000 m	z = 0.00	Warning! Large pull: L = 3.166 +- 0.020 m	z = -9.16
Warning! Large pull: L = 3.356 +- 0.000 m	z = 0.00	Warning! Large pull: L = 3.344 +- 0.001 m	z = -5.25
Warning! Large pull: L = 3.089 +- 0.000 m	z = 0.00	Warning! Large pull: L = 3.653 +- 0.025 m	z = 12.15
Warning! Large pull: L = 3.112 +- 0.000 m	z = 0.00	Warning! Large pull: L = 3.386 +- 0.010 m	z = 3.67
Warning! Large pull: L = 3.417 +- 0.000 m	z = 0.00	Warning! Large pull: L = 3.660 +- 0.004 m	z = 3.70
Warning! Large pull: L = 3.405 +- 0.000 m	z = 0.00	Warning! Large pull: L = 3.353 +- 0.001 m	z = 3.75
Warning! Large pull: L = 3.409 +- 0.000 m	z = 0.00	Warning! Large pull: L = 3.324 +- 0.001 m	z = -25.25
Warning! Large pull: L = 3.743 +- 0.000 m	z = 0.00	Warning! Large pull: L = 2.351 +- 0.015 m	z = -66.55
Warning! Large pull: L = 3.721 +- 0.005 m	z = 61.75	Warning! Large pull: L = 1.337 +- 0.003 m	z = -670.75
Warning! Large pull: L = 3.134 +- 0.050 m	z = -5.56	Warning! Large pull: L = 3.641 +- 0.014 m	z = 20.04
Warning! Large pull: L = 1.413 +- 0.013 m	z = -153.79	Warning! Large pull: L = 1.351 +- 0.005 m	z = -399.65
Warning! Large pull: L = 3.142 +- 0.005 m	z = -54.05	Warning! Large pull: L = 3.357 +- 0.002 m	z = 3.87
Warning! Large pull: L = 3.107 +- 0.005 m	z = -61.05	Warning! Large pull: L = 3.299 +- 0.002 m	z = -25.13
Warning! Large pull: L = 4.355 +- 0.260 m	z = 3.63	Warning! Large pull: L = 3.286 +- 0.005 m	z = -12.65
Warning! Large pull: L = 3.795 +- 0.100 m	z = 3.83	Warning! Large pull: L = 2.636 +- 0.050 m	z = -14.27
Warning! Large pull: L = 3.091 +- 0.015 m	z = -21.42	Warning! Large pull: L = 3.361 +- 0.002 m	z = 5.87
Warning! Large pull: L = 2.625 +- 0.090 m	z = -8.75	Warning! Large pull: L = 2.631 +- 0.010 m	z = -71.83
Warning! Large pull: L = 3.106 +- 0.073 m	z = -4.19	Warning! Large pull: L = 2.665 +- 0.150 m	z = -4.56
Warning! Large pull: L = 3.132 +- 0.022 m	z = -12.74	Warning! Large pull: L = 3.358 +- 0.002 m	z = 4.37
Warning! Large pull: L = 3.168 +- 0.045 m	z = -5.43	Warning! Large pull: L = 3.341 +- 0.001 m	z = -8.25
Warning! Large pull: L = 2.206 +- 0.002 m	z = -603.11	Warning! Large pull: L = 3.344 +- 0.001 m	z = -5.25
Warning! Large pull: L = 3.094 +- 0.002 m	z = -159.11	Warning! Large pull: L = 2.341 +- 0.025 m	z = -40.33
Warning! Large pull: L = 3.085 +- 0.005 m	z = -65.45	Warning! Large pull: L = 3.618 +- 0.013 m	z = 20.67
Warning! Large pull: L = 3.320 +- 0.005 m	z = -18.45	Warning! Large pull: L = 3.341 +- 0.002 m	z = -4.13
Warning! Large pull: L = 3.106 +- 0.050 m	z = -6.12	Warning! Large pull: L = 3.378 +- 0.002 m	z = 14.37
Warning! Large pull: L = 3.110 +- 0.010 m	z = -30.22	Warning! Large pull: L = 3.335 +- 0.002 m	z = -7.13
Warning! Large pull: L = 3.717 +- 0.061 m	z = 5.00	Warning! Large pull: L = 3.354 +- 0.001 m	z = 4.75

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

Excluded data due to bad pull

Warning! Large pull: L = 3.750 +- 0.013 m z = 25.98	Warning! Large pull: L = 3.696 +- 0.050 m z = 6.93
Warning! Large pull: L = 3.750 +- 0.013 m z = 25.98	Warning! Large pull: L = 3.371 +- 0.085 m z = 4.35
Warning! Large pull: L = 3.750 +- 0.013 m z = 25.98	Warning! Large pull: L = 2.350 +- 0.015 m z = -66.62
Warning! Large pull: L = 2.350 +- 0.015 m z = -66.62	Warning! Large pull: L = 3.746 +- 0.020 m z = 19.84
Warning! Large pull: L = 3.746 +- 0.020 m z = 19.84	Warning! Large pull: L = 3.851 +- 0.082 m z = -148.13
Warning! Large pull: L = 3.851 +- 0.082 m z = -148.13	Warning! Large pull: L = 2.350 +- 0.081 m z = -999.25
Warning! Large pull: L = 2.350 +- 0.081 m z = -999.25	Warning! Large pull: L = 2.351 +- 0.020 m z = -99.91
Warning! Large pull: L = 2.351 +- 0.020 m z = -99.91	Warning! Large pull: L = 3.122 +- 0.025 m z = -9.09
Warning! Large pull: L = 3.122 +- 0.025 m z = -9.09	Warning! Large pull: L = 3.631 +- 0.086 m z = 46.96
Warning! Large pull: L = 3.631 +- 0.086 m z = 46.96	Warning! Large pull: L = 3.535 +- 0.085 m z = 37.15
Warning! Large pull: L = 3.535 +- 0.085 m z = 37.15	Warning! Large pull: L = 1.350 +- 0.010 m z = -199.93
Warning! Large pull: L = 1.350 +- 0.010 m z = -199.93	Warning! Large pull: L = 4.550 +- 0.023 m z = 52.21
Warning! Large pull: L = 4.550 +- 0.023 m z = 52.21	Warning! Large pull: L = 3.322 +- 0.081 m z = -27.25
Warning! Large pull: L = 3.322 +- 0.081 m z = -27.25	Warning! Large pull: L = 3.166 +- 0.020 m z = -9.16
Warning! Large pull: L = 3.166 +- 0.020 m z = -9.16	Warning! Large pull: L = 3.344 +- 0.081 m z = -5.25
Warning! Large pull: L = 3.344 +- 0.081 m z = -5.25	Warning! Large pull: L = 3.653 +- 0.025 m z = 12.15
Warning! Large pull: L = 3.653 +- 0.025 m z = 12.15	Warning! Large pull: L = 3.386 +- 0.010 m z = 3.67
Warning! Large pull: L = 3.386 +- 0.010 m z = 3.67	Warning! Large pull: L = 3.660 +- 0.084 m z = 3.70
Warning! Large pull: L = 3.660 +- 0.084 m z = 3.70	Warning! Large pull: L = 3.353 +- 0.081 m z = 3.75
Warning! Large pull: L = 3.353 +- 0.081 m z = 3.75	Warning! Large pull: L = 3.324 +- 0.081 m z = -25.25
Warning! Large pull: L = 3.324 +- 0.081 m z = -25.25	Warning! Large pull: L = 2.351 +- 0.015 m z = -66.55
Warning! Large pull: L = 2.351 +- 0.015 m z = -66.55	Warning! Large pull: L = 1.337 +- 0.083 m z = -670.75
Warning! Large pull: L = 1.337 +- 0.083 m z = -670.75	Warning! Large pull: L = 3.641 +- 0.014 m z = 20.84
Warning! Large pull: L = 3.641 +- 0.014 m z = 20.84	Warning! Large pull: L = 1.351 +- 0.085 m z = -399.65
Warning! Large pull: L = 1.351 +- 0.085 m z = -399.65	Warning! Large pull: L = 3.357 +- 0.082 m z = 3.87
Warning! Large pull: L = 3.357 +- 0.082 m z = 3.87	Warning! Large pull: L = 3.299 +- 0.082 m z = -25.13
Warning! Large pull: L = 3.299 +- 0.082 m z = -25.13	Warning! Large pull: L = 3.286 +- 0.085 m z = -12.65
Warning! Large pull: L = 3.286 +- 0.085 m z = -12.65	Warning! Large pull: L = 2.636 +- 0.050 m z = -14.27
Warning! Large pull: L = 2.636 +- 0.050 m z = -14.27	Warning! Large pull: L = 3.361 +- 0.082 m z = 5.87
Warning! Large pull: L = 3.361 +- 0.082 m z = 5.87	Warning! Large pull: L = 2.631 +- 0.010 m z = -71.83
Warning! Large pull: L = 2.631 +- 0.010 m z = -71.83	Warning! Large pull: L = 2.665 +- 0.150 m z = -4.56
Warning! Large pull: L = 2.665 +- 0.150 m z = -4.56	Warning! Large pull: L = 3.358 +- 0.082 m z = 4.37
Warning! Large pull: L = 3.358 +- 0.082 m z = 4.37	Warning! Large pull: L = 3.341 +- 0.081 m z = -8.25
Warning! Large pull: L = 3.341 +- 0.081 m z = -8.25	Warning! Large pull: L = 3.344 +- 0.081 m z = -5.25
Warning! Large pull: L = 3.344 +- 0.081 m z = -5.25	Warning! Large pull: L = 2.341 +- 0.025 m z = -40.33
Warning! Large pull: L = 2.341 +- 0.025 m z = -40.33	Warning! Large pull: L = 3.618 +- 0.013 m z = 20.67
Warning! Large pull: L = 3.618 +- 0.013 m z = 20.67	Warning! Large pull: L = 3.341 +- 0.082 m z = -4.13
Warning! Large pull: L = 3.341 +- 0.082 m z = -4.13	Warning! Large pull: L = 3.378 +- 0.082 m z = 14.37
Warning! Large pull: L = 3.378 +- 0.082 m z = 14.37	Warning! Large pull: L = 3.335 +- 0.082 m z = -7.13
Warning! Large pull: L = 3.335 +- 0.082 m z = -7.13	Warning! Large pull: L = 3.354 +- 0.081 m z = 4.75
Warning! Large pull: L = 3.354 +- 0.081 m z = 4.75	

Largest pull is about 1000, the result of a fine measurement with tiny uncertainty...
...but 1 meter wrong!

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

Weighted results

30cm:

Mean = 3.41300 ± 0.00027 m

RMS = undefined! (N = 677)

2m:

Mean = 3.34985 ± 0.00011 m

RMS = undefined! (N = 779)

Now the results are really precise, and the 2m result is about a factor 2 more so.

Improvement over the unweighted 30cm / 2m results are factors of 1.8 / 1.8
So the uncertainties carry information about the measurement quality.

Weighted results

30cm:

$$\text{Mean} = 3.41300 \pm 0.00027 \text{ m}$$

Chi2 = 642.8, Ndof = 720, Prob = 0.98

2m:

$$\text{Mean} = 3.34985 \pm 0.00011 \text{ m}$$

Chi2 = 715.9, Ndof = 789, Prob = 0.97

Now the results are really precise, and the 2m result is about a factor 2 more so.

Improvement over the unweighted 30cm / 2m results are factors of 1.8 / 1.8
So the uncertainties carry information about the measurement quality.

Cross Checks

Once again, we compare the 30cm and 2m weighted results.

Subtracting the *difference in blinding value* yields the real difference:

$$L_{30\text{cm}} - L_{2\text{m}} (\text{partially blinded}) = -0.00034 \pm 0.00029 \text{ m} \quad (1.2\sigma \rightarrow \text{prob}=0.24)$$

This means that the two results are reasonably within each other, and the results and their uncertainty are (more) trustworthy.

We can also check the unweighted against the weighted results. Here, there is not even a partial unblinding, as they have the same offsets.

$$30\text{cm: Unweighted-Weighted} = -0.00110 \pm 0.00056 \text{ m} \quad (2.0\sigma \rightarrow \text{prob} = 0.048)$$

$$2\text{m: Unweighted-Weighted} = -0.00042 \pm 0.00023 \text{ m} \quad (1.9\sigma \rightarrow \text{prob} = 0.062)$$

Thus, now the four results (30cm, 2m) x (unweighted, weighted) seem to be in agreement.

A problem?

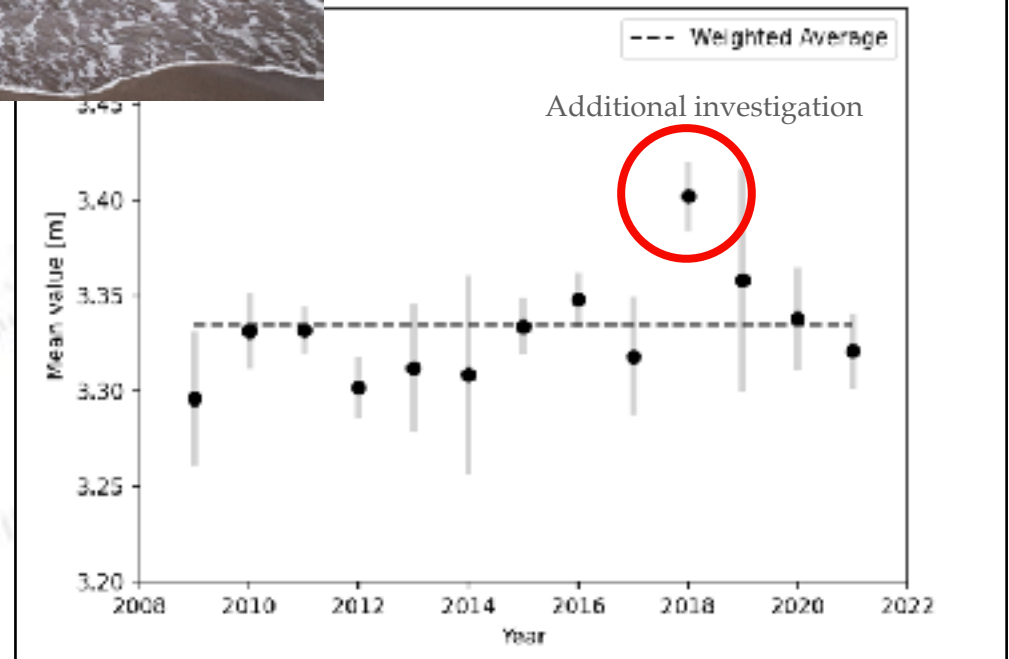


Things may look very good, yet it remains to investigate the data further.

A question is the homogeneity of the data. And here we find problems!

Somehow, the 2018 data seems different (read: biased) compared to the other years.

One could also ask, if the order in the data file mattered.





Fitting analysis

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.

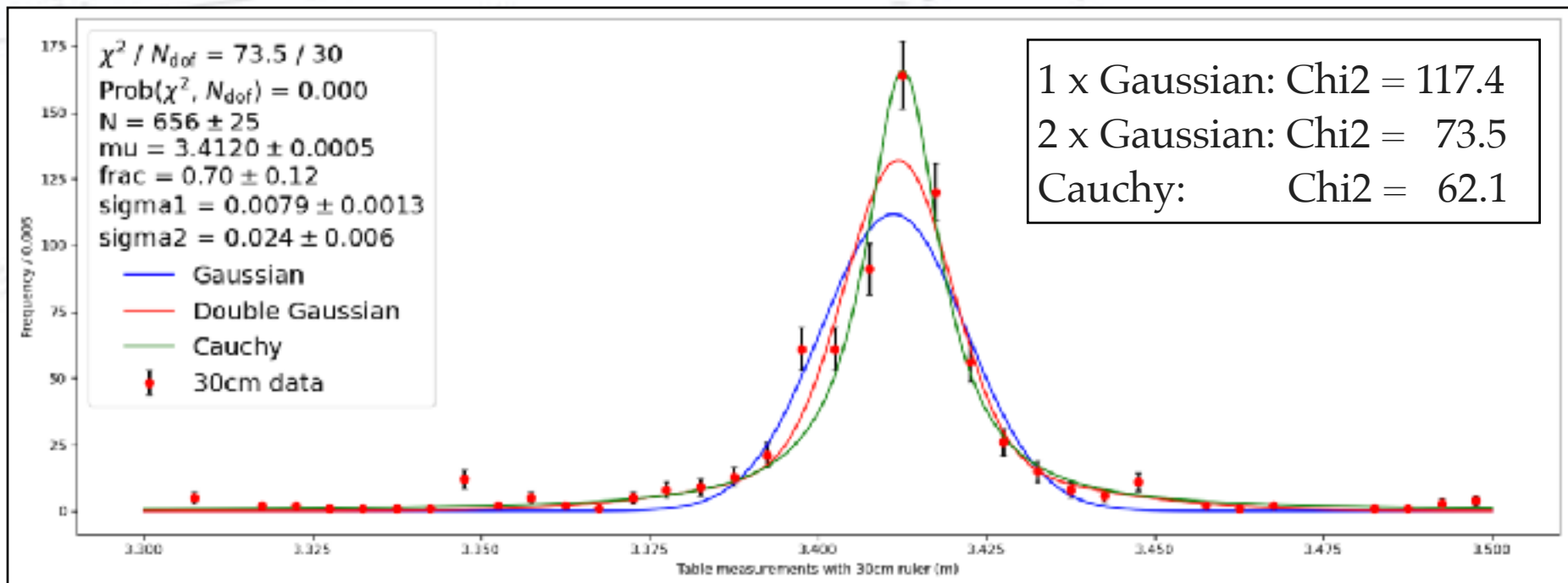
While all of these can be accommodated, it is still a challenge, at the following “fitting around” took me several hours!

Fitting for a result

First step is to establish what PDF the measurements follow.

I have tried the following three:

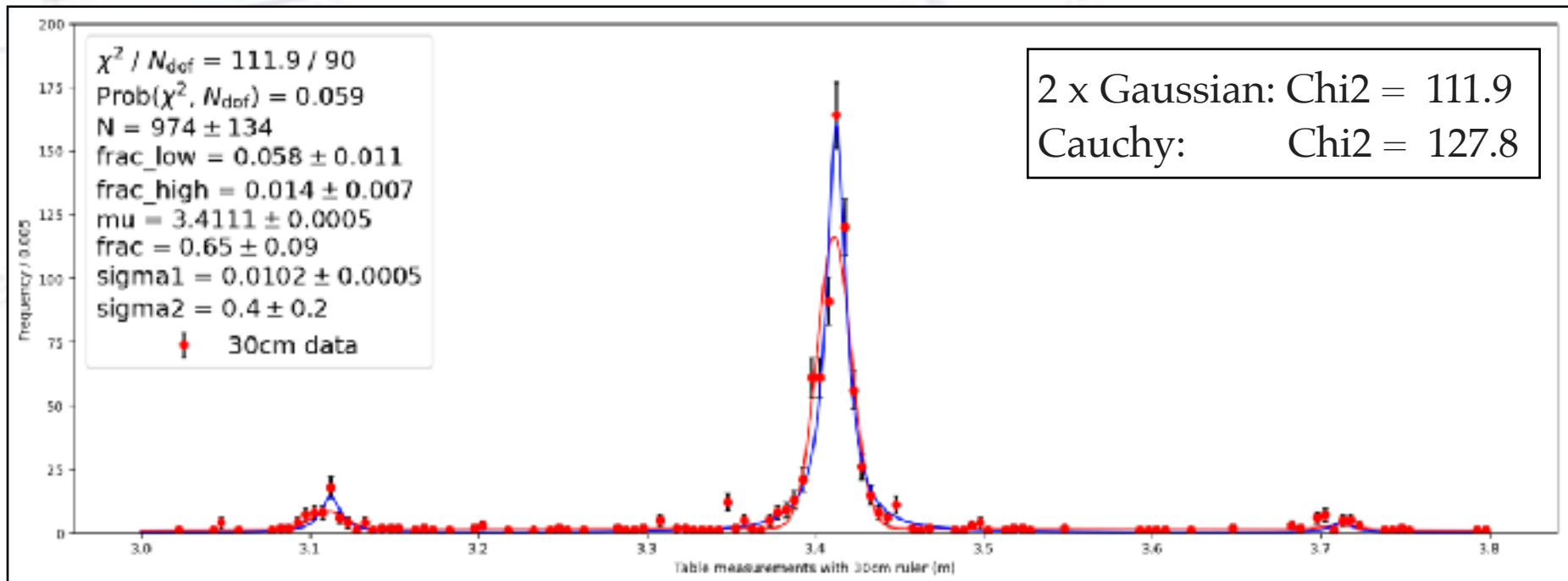
- Single Gaussian: Simplest and mandatory first step.
- Double Gaussian: To accommodate different resolutions.
- Cauchy: Alternative to Gaussian with long tails as expected.



Fitting for a result

The fits converge and gives OK values. However, both models have a problem modelling the far outliers. The second Gaussian starts being used for this, thus not matching the peak.

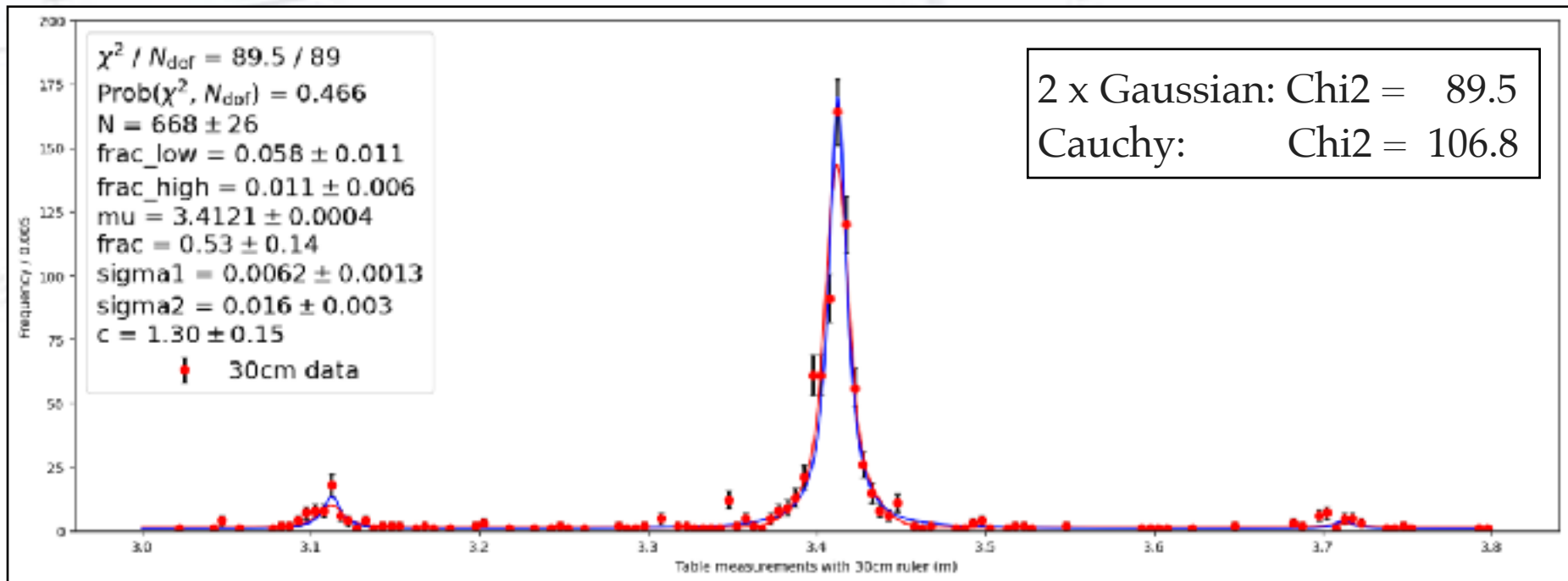
A better model, which avoids this problem should have a separate PDF for the far outliers.



Fitting for a result

The fits converge and gives OK values. However, both models have a problem modelling the far outliers. The second Gaussian starts being used for this, thus not matching the peak.

A better model, which avoids this problem should have a separate PDF for the far outliers. Adding a constant improves the fits, especially the double Gaussian.



Fitting results

Summarising all the fitting results (below), it is clear that the quality of the fit slowly improves.

Chi2 - Single Gaussian:	Prob(chi2 = 117.4, Ndof = 94) = 0.052	Mu = 3.411227 +- 0.000455
Chi2 - Double Gaussian:	Prob(chi2 = 73.5, Ndof = 92) = 0.922	Mu = 3.411950 +- 0.000547
Chi2 - 3 x Double Gaussian:	Prob(chi2 = 111.9, Ndof = 90) = 0.059	Mu = 3.411138 +- 0.000452
Chi2 - 3 x Double Gaussian + c:	Prob(chi2 = 89.5, Ndof = 89) = 0.466	Mu = 3.412124 +- 0.000448
ULLH - 3 x Double Gaussian + c:	Likelihood value = -6327.9	Mu = 3.413071 +- 0.000321
ULLH - 3 x Cauchy + c:	Likelihood value = -6241.9	Mu = 3.413009 +- 0.000318

The double Gaussian tripple peak fit has a good ChiSquare, but the statistics is often low, and hence a likelihood fit is used.

Even with a good PDF, this was not easy to get running, and amendments were needed. However, the result is significantly more precise, and in the end we reach an uncertainty of 0.32mm.

Both value and uncertainty are remarkably comparable to the weighted mean result: Weighted mean = 3.41300 +- 0.00027m Fit: 3.413071 +- 0.00032m

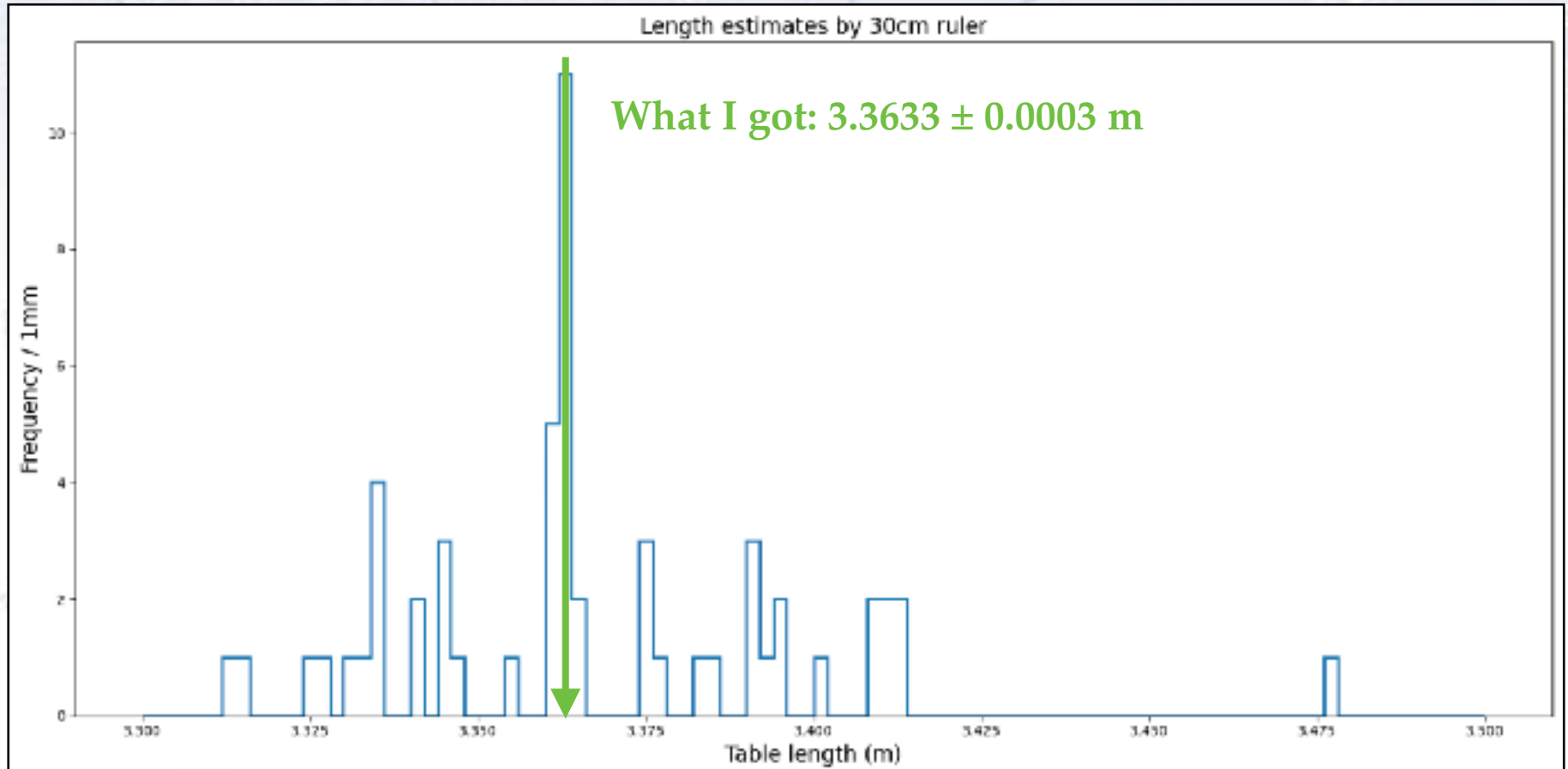
The fitting method starts being a significant systematic uncertainty!

The background is a faded, light blue map of the North Atlantic Ocean. It features numerous depth contour lines labeled with values such as 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5100, 5200, 5300, 5400, 5500, 5600, 5700, 5800, 5900, 6000, 6100, 6200, 6300, 6400, 6500, 6600, 6700, 6800, 6900, 7000, 7100, 7200, 7300, 7400, 7500, 7600, 7700, 7800, 7900, 8000, 8100, 8200, 8300, 8400, 8500, 8600, 8700, 8800, 8900, 9000, 9100, 9200, 9300, 9400, 9500, 9600, 9700, 9800, 9900, 10000. There are also several shipping routes marked with lines and dots, and some text labels like "101 BATTLESHIP" and "YACHT 1423".

Student analyses comparison

Your measurement value

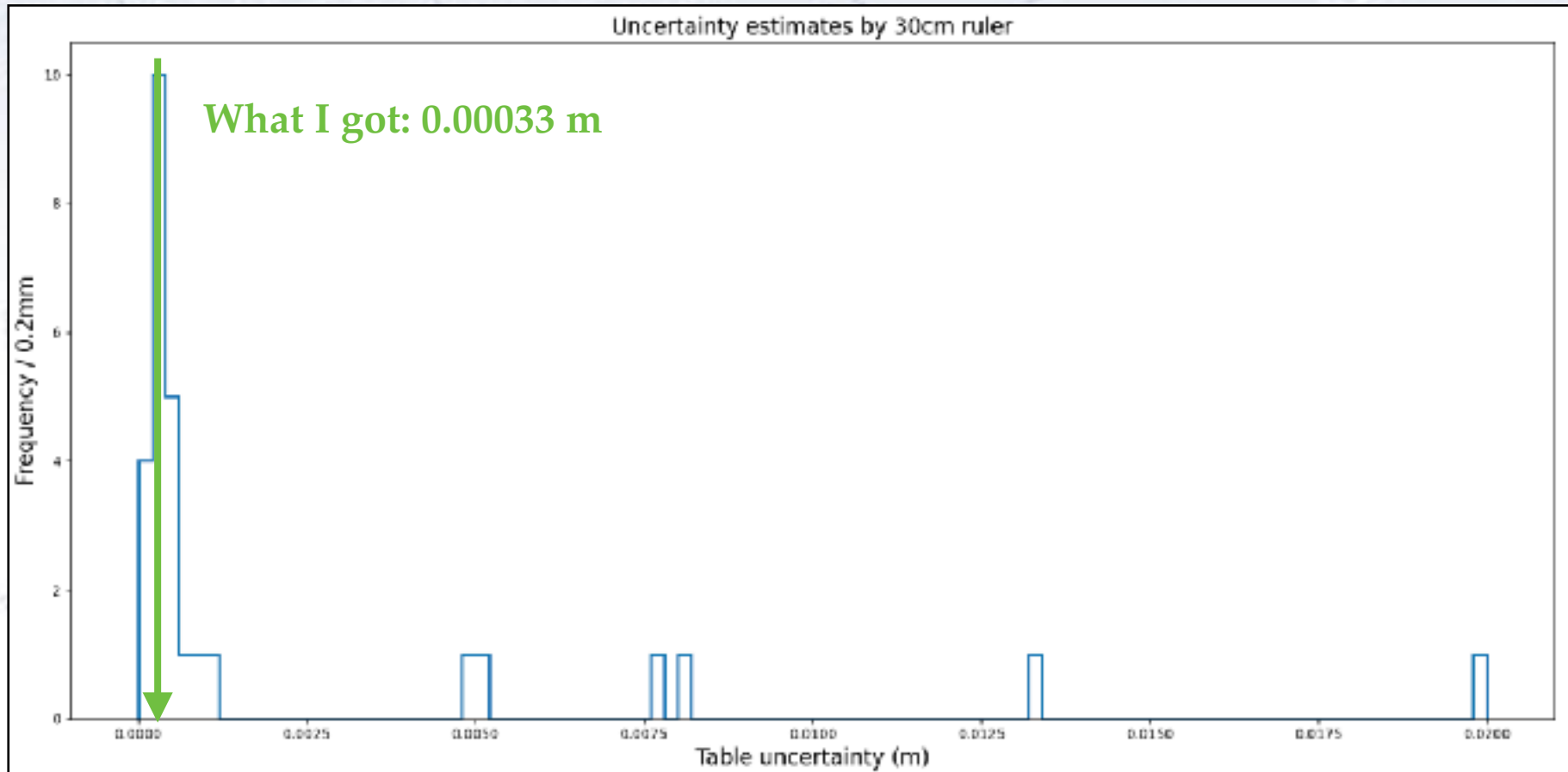
The uncertainties varied quite a bit.... from 3.3 to beyond 3.4.



Estimating uncertainties is (still) hard.

Your measurement uncertainty

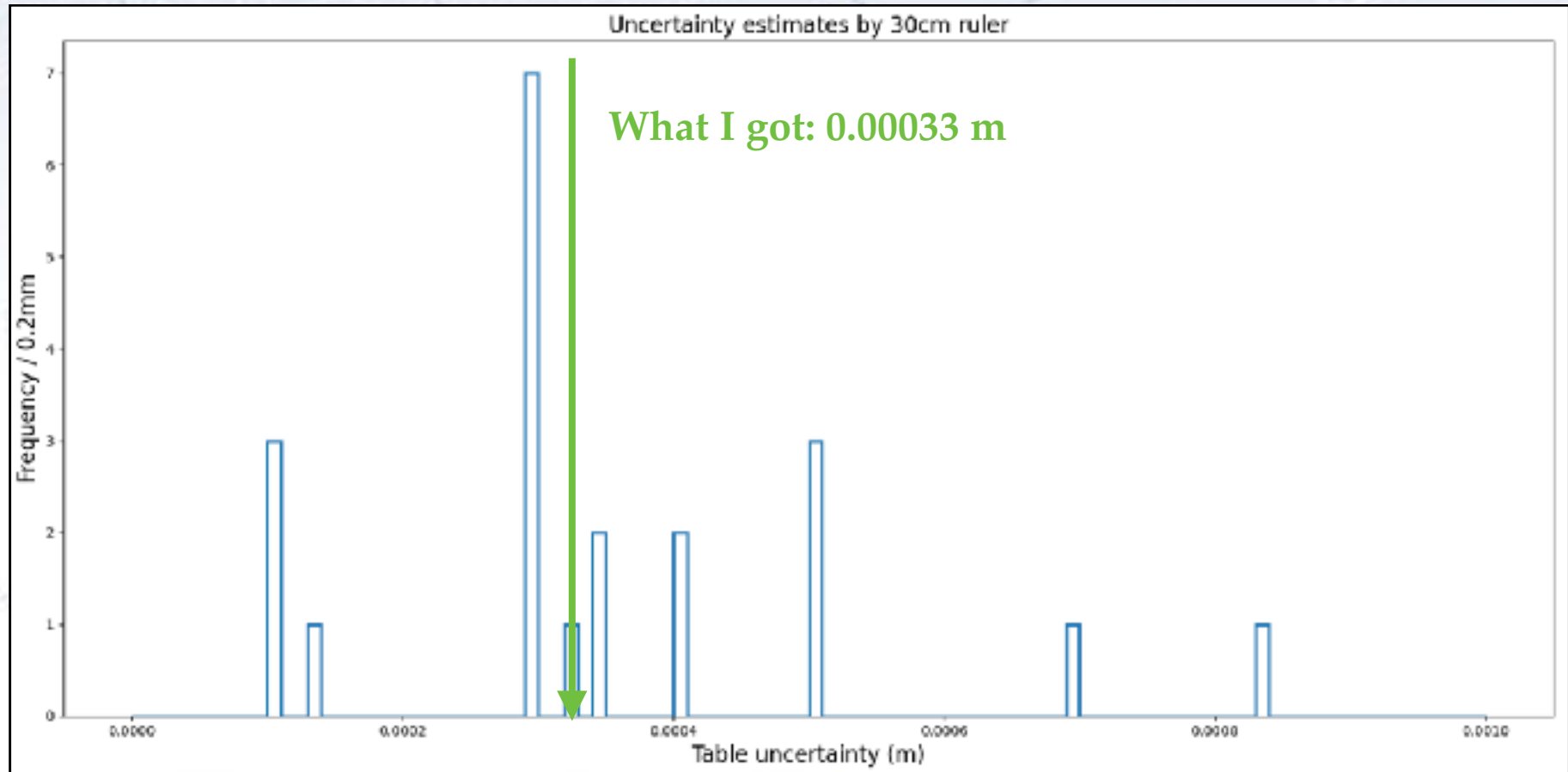
The measurement uncertainties varied even more wildly!!!



The lowest was 0.0001, while the highest was 0.02 (two orders of magnitude).

Your measurement uncertainty

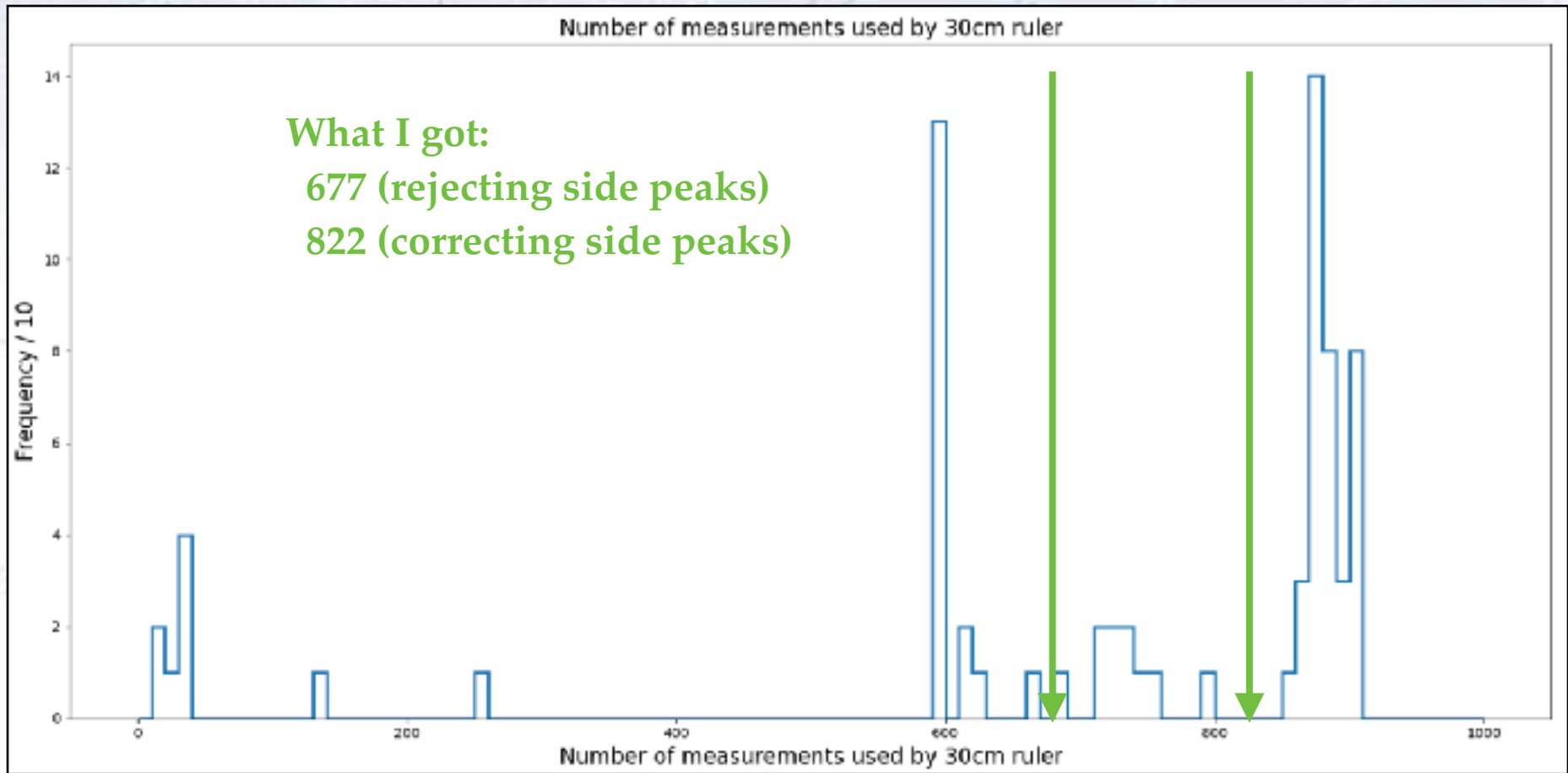
The measurement uncertainties varied even more wildly!!!



The lowest was 0.0001, while the highest was 0.02 (two orders of magnitude).

Your number of measurements

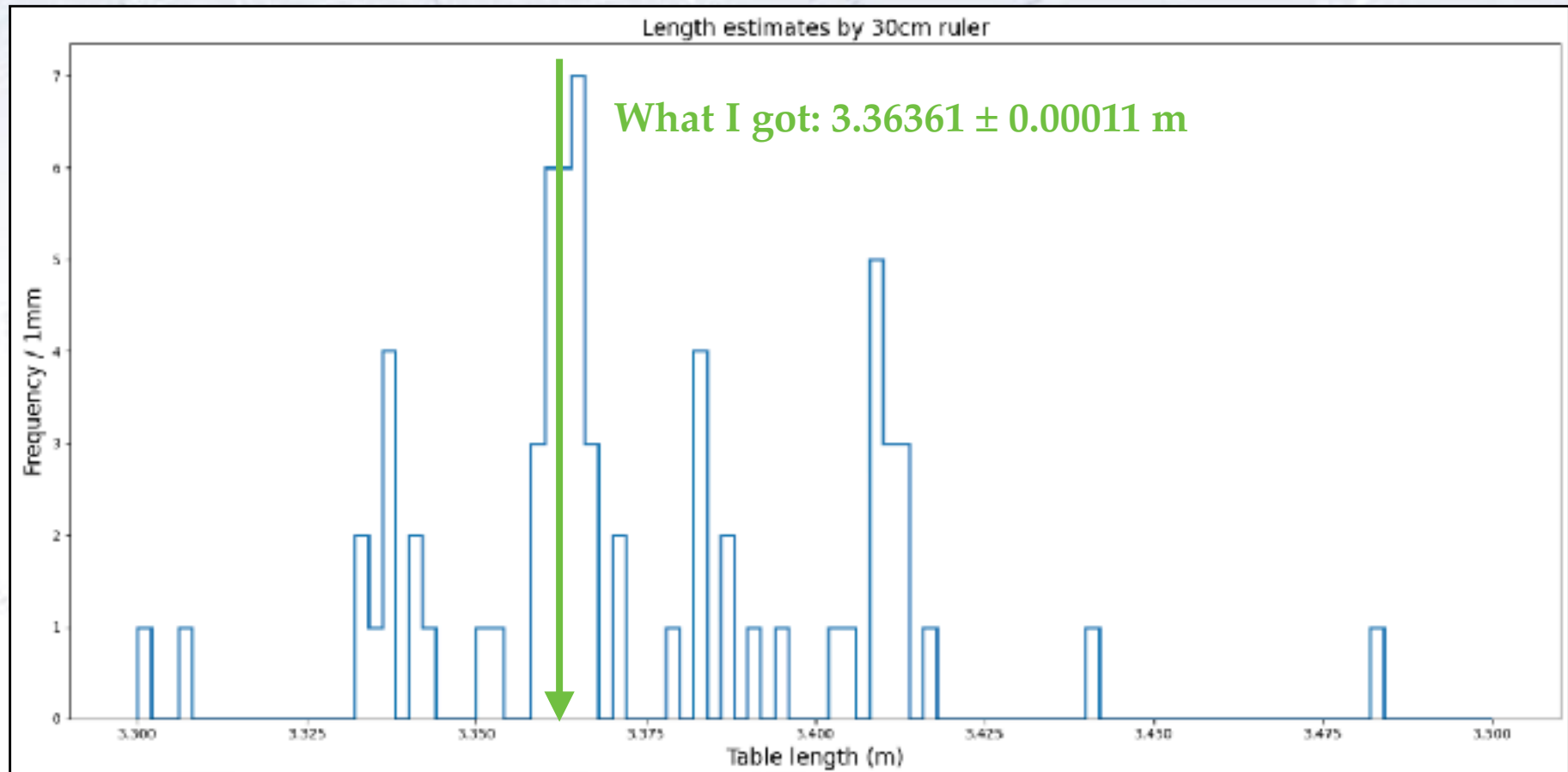
The number of measurements also varied, but some were in the right ballpark.



Remember that the impact is only \sqrt{N} , and thus not overly important!

Your measurement uncertainty

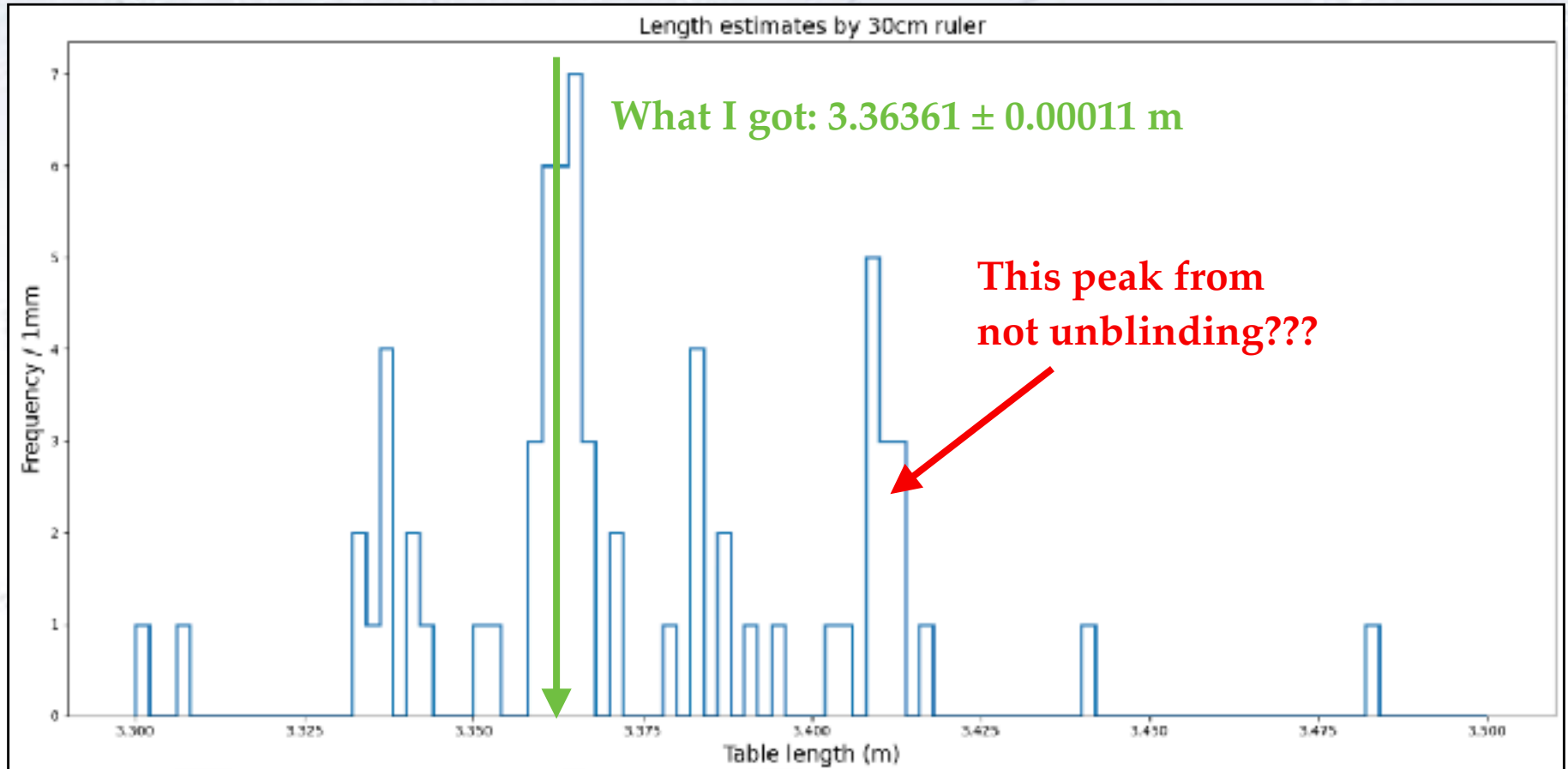
The uncertainties varied quite a bit.... from 3.3 to beyond 3.4.



Estimating uncertainties is (still) hard.

Your measurement uncertainty

The uncertainties varied quite a bit.... from 3.3 to beyond 3.4.



Estimating uncertainties is (still) hard.

The background is a detailed nautical chart of the North Atlantic Ocean. It features a grid of latitude and longitude lines, with latitude marked from 30 to 60 degrees North and longitude from 10 to 30 degrees West. Depth contours are shown in fathoms, with labels such as 100, 200, 300, and 400. A prominent shipping lane is indicated by a series of small, dark, rectangular symbols, likely representing shipwrecks or navigational hazards. The text "THE NORTH OCEAN" is visible in the upper right corner, and "YACHTS" is written in the lower right corner. The overall tone of the map is a light blue-grey.

The Quick & Dirty

The quick and dirty solution(s):

The above analysis is some work, but once you get the hang of it, and have previously produced (or copied/understood) code for the task, it is less cumbersome. Once you see a plot of the data and understand what is happening, **the essence of this data analysis is to only consider the reasonable measurements** and extract a value from these.

The below code does that in a quick and dirty manner, which fails to do all the checks that are needed, if the data is important and the situation calls for it.

```
# Looking at the initial (30cm in particular) plots, it is clear that there is a central peak of valid measurements.
# By eye, the 30cm / 2m peak is at 3.415m / 3.350m (blinded!) and the width about 2.5cm / 1cm, so discarding all measurements
# outside +- 7.5cm / 3cm is a crude but fast way forward.
m30cm = np.abs(L30cm - 3.415) < 0.075
mu30cm = np.mean(np.array(L30cm)[m30cm])
sig30cm = np.std(np.array(L30cm)[m30cm])
print(f" The crude (unweighted) mean = {mu30cm:7.5f} +- {sig30cm/np.sqrt(len(L30cm)):7.5f} m" +
      f" (brutally raw (gu)estimate) from {len(np.array(L30cm)[m30cm]):d} measurements")

m2m = np.abs(L2m - 3.350) < 0.030
mu2m = np.mean(np.array(L2m)[m2m])
sig2m = np.std(np.array(L2m)[m2m])
print(f" The crude (unweighted) mean = {mu2m:7.5f} +- {sig2m/np.sqrt(len(L2m)):7.5f} m" +
      f" (brutally raw (gu)estimate) from {len(np.array(L2m)[m2m]):d} measurements")
```

```
The crude (unweighted) mean = 3.41055 +- 0.00057 m (brutally raw (gu)estimate) from 705 measurements
The crude (unweighted) mean = 3.34932 +- 0.00022 m (brutally raw (gu)estimate) from 801 measurements
```

However, this approach is not advisable. Give it a little more consideration, and **promise yourself that you'll at the very minimum always do the following three key things:**

1. Blind the data
2. Plot the data
3. Make reassuring cross check and info print statements throughout your code.

By doing these things, you might only have to debug your way out of the unforeseen (e.g. negative uncertainties) to get to a decent result, that you can convince yourself and others is in the right ballpark. Good luck.

Notice that even though there has been a lot of analysis, comparison, and discussion of the result, the actual value of the table length has not yet been **unblinded!**

Conclusions

Specifically on the analysis:

- 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.

More generally:

- 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.

