Absolute structure determination without heavy atoms

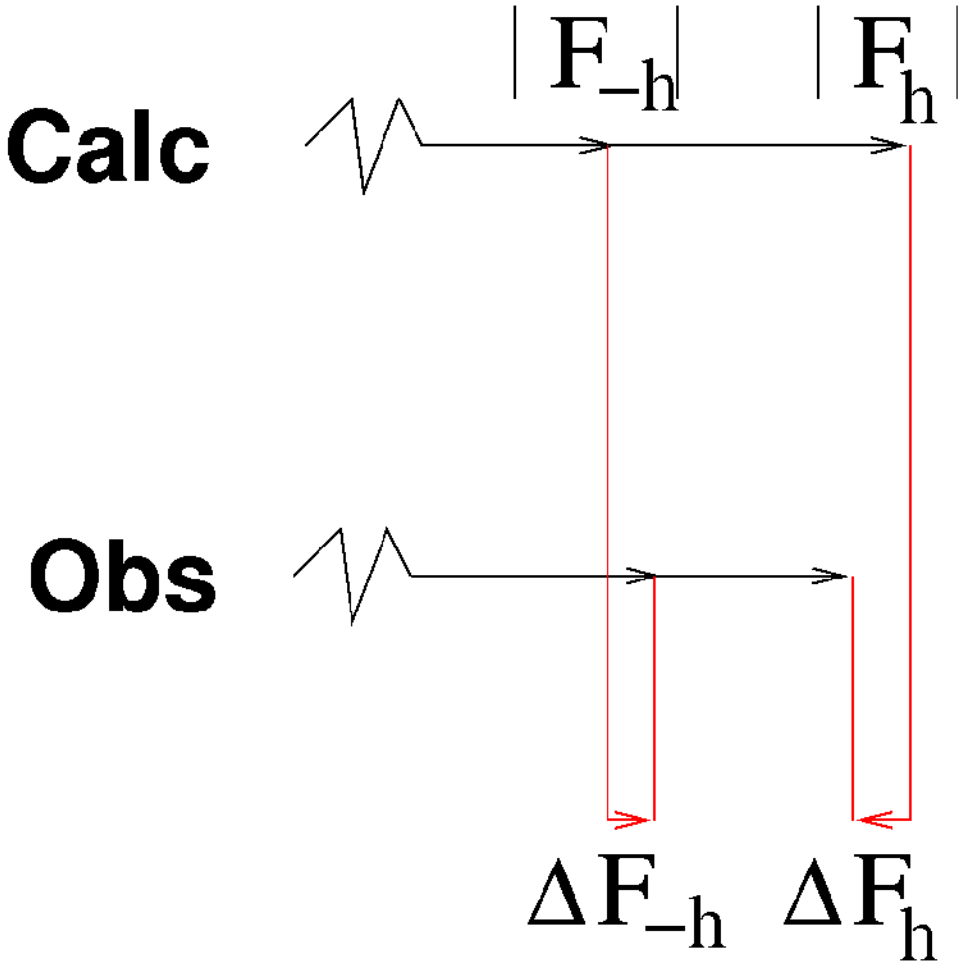
Rob W.W. Hooft, Bruker AXS, Delft
Leo H. Straver, Bruker AXS, Delft
Anthony L. Spek, Utrecht University

How our statistical analysis using Student's t-distribution can improve your absolute structure determinations.

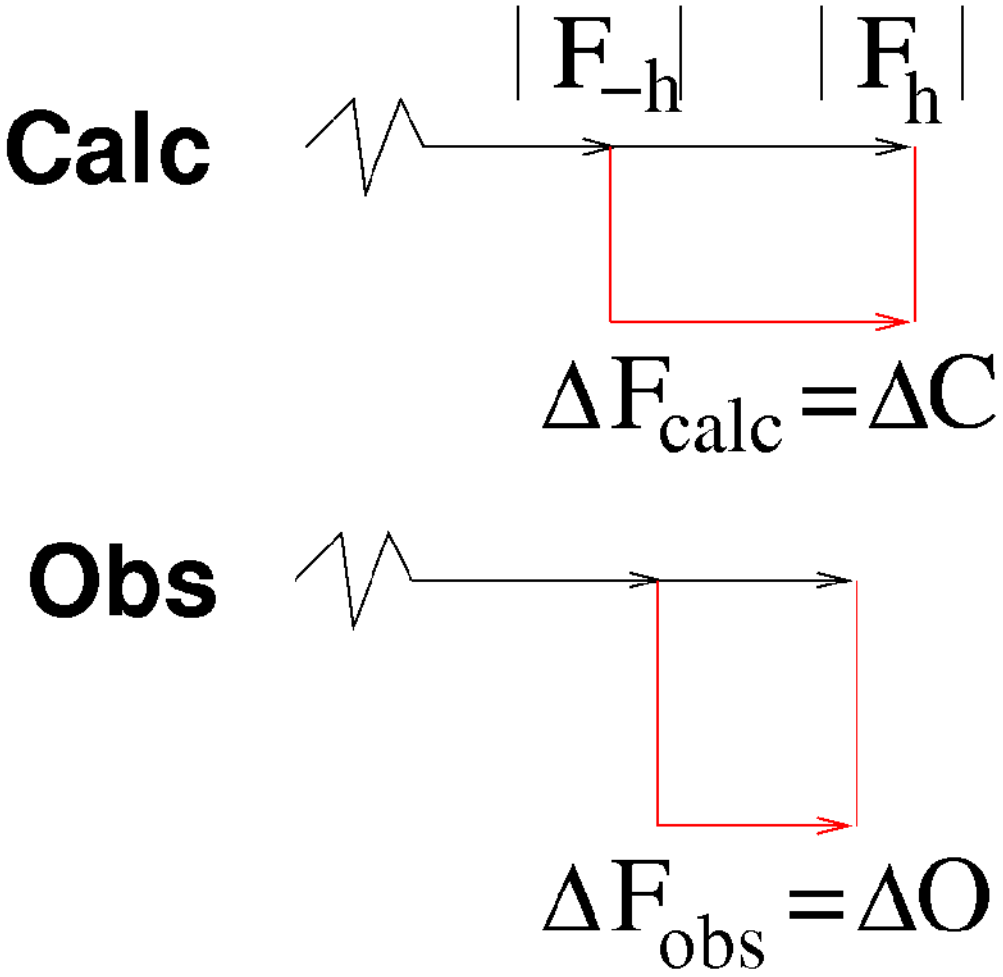
The resonant scattering signal: f”

	f”(CuKa)	f”(MoKa)
Se	1.14	2.23
Cl	0.70	0.16
S	0.56	0.12
O	0.032	0.006
N	0.018	0.003
C	0.009	0.002

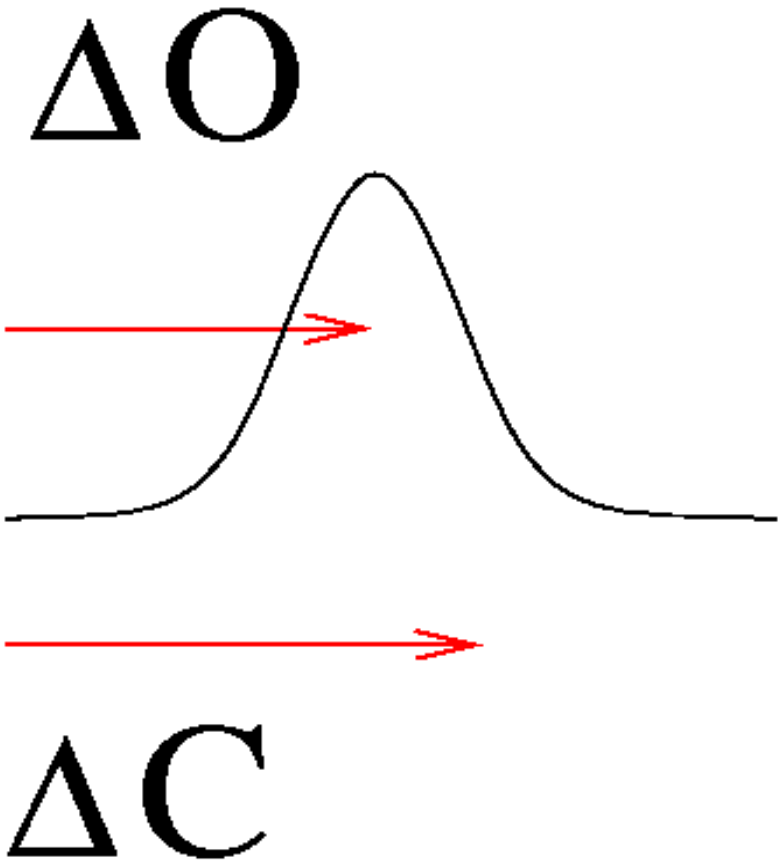
Structure refinement



Observed and calculated differences



Likelihood of correspondence



Normal distribution for one reflection

$$z = \frac{\Delta C - \Delta O}{\sigma_{\Delta O}} \qquad p(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- Multiply all probabilities for all reflection pairs.

- This can be extended to different structure models:

  Inverted model

  - replace ΔC by -ΔC

  Mixed model

  - multiply ΔC by γ; calculate best γ and y

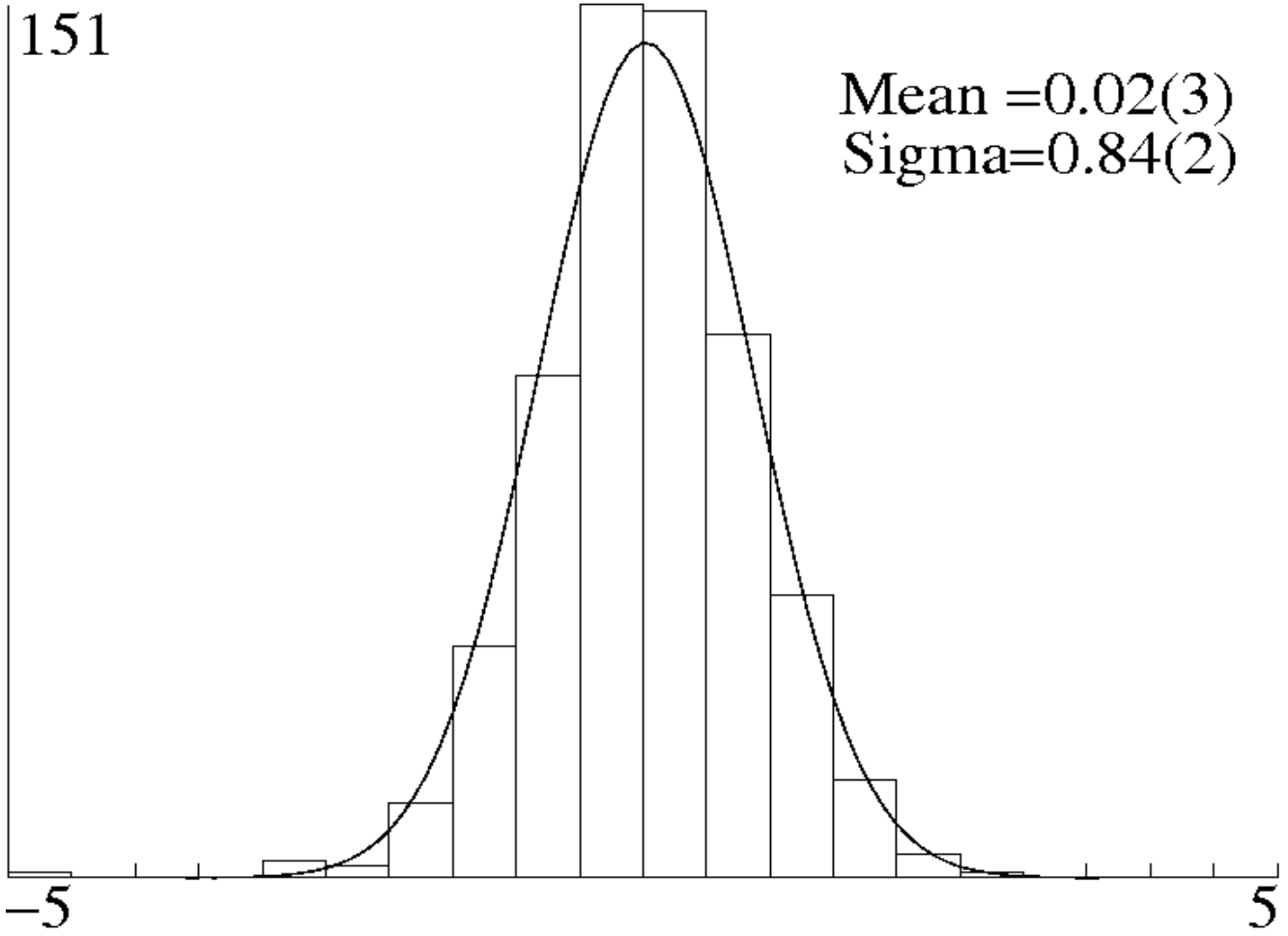
Results for a very good measurement

Example: ammonium bitartrate ($C_4H_9O_6N$), MoKα

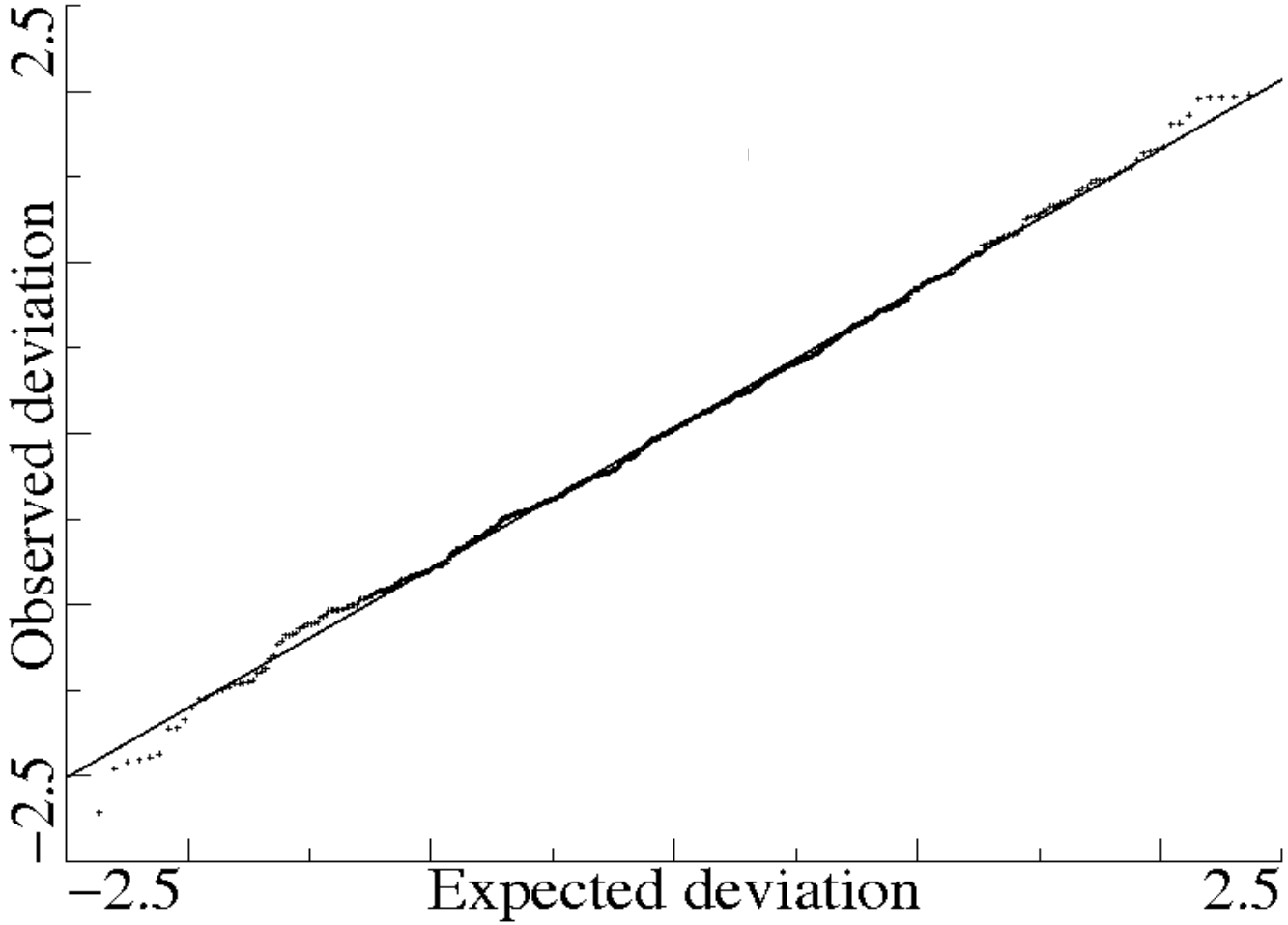
- Refined Flack *x* parameter: -0.1(9)

- Likelihood that the model is inverted: $1x10^{-10}$
- Value of *y*: -0.05(13)

Analysis of the deviations



Normal probability plot



So far...

Implemented in PLATON.

Great results for vast majority of structures.

- Sufficient signal from O, consistent statistics
- PLATON gives NPP slope and correlation
- Correction for slope<1.0 roughly scales σ(y)

But: The accuracy of σ(F) is critical.

- If the correlation is not around 0.999, there is nothing we can do to improve that.
  - Probability is not reliable
  - In extreme cases: assignment not reliable

A troubled example (thank you!)

Example: molecule containing sulphur, MoKα

- Refined Flack *x* parameter: 0.17(14)

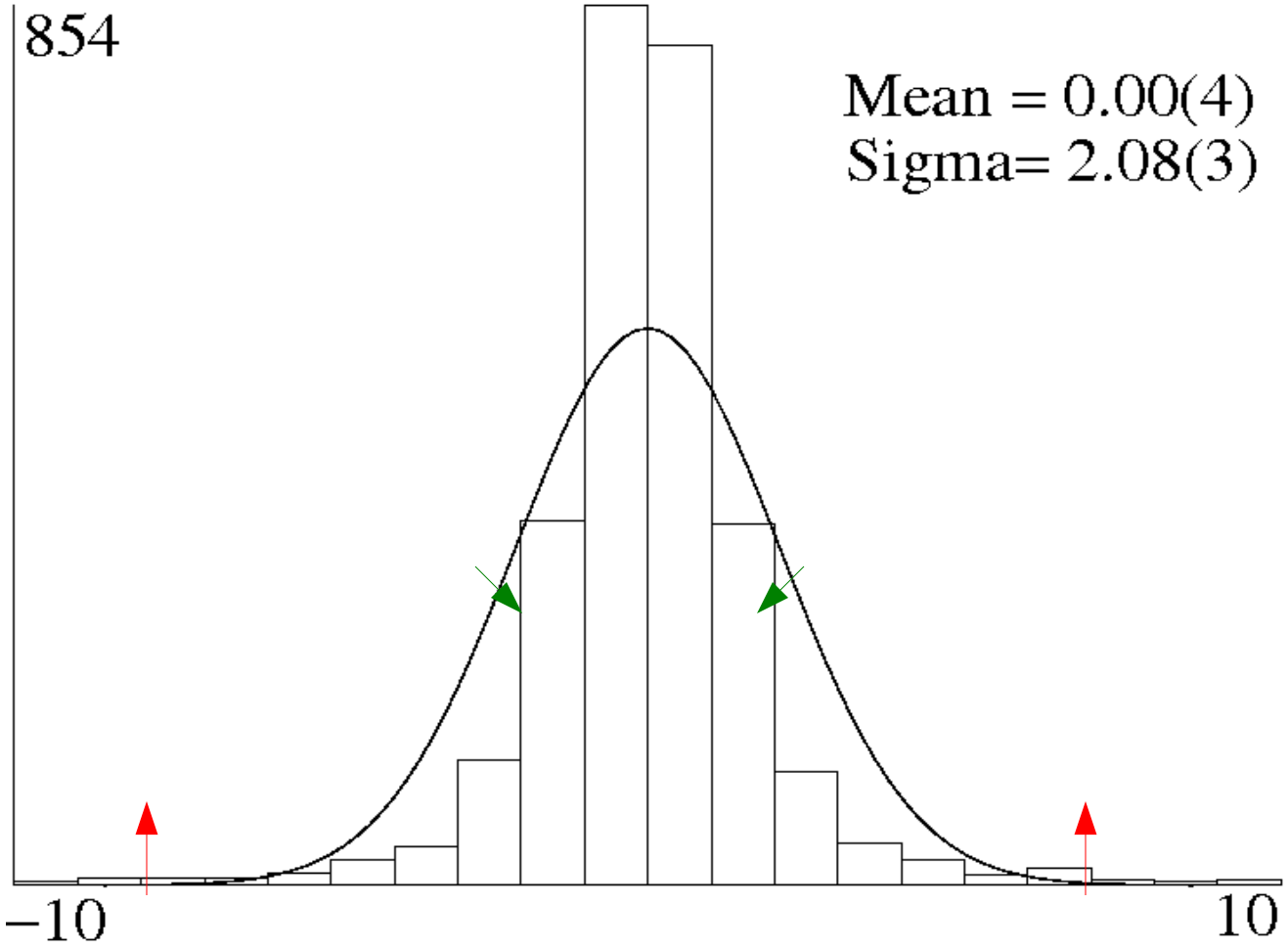
But our “old” analysis gives:

- A big red flag in the Normal Probability Plot
  - Correlation 0.94 < 0.999
  - Slope 1.29 > 1.0

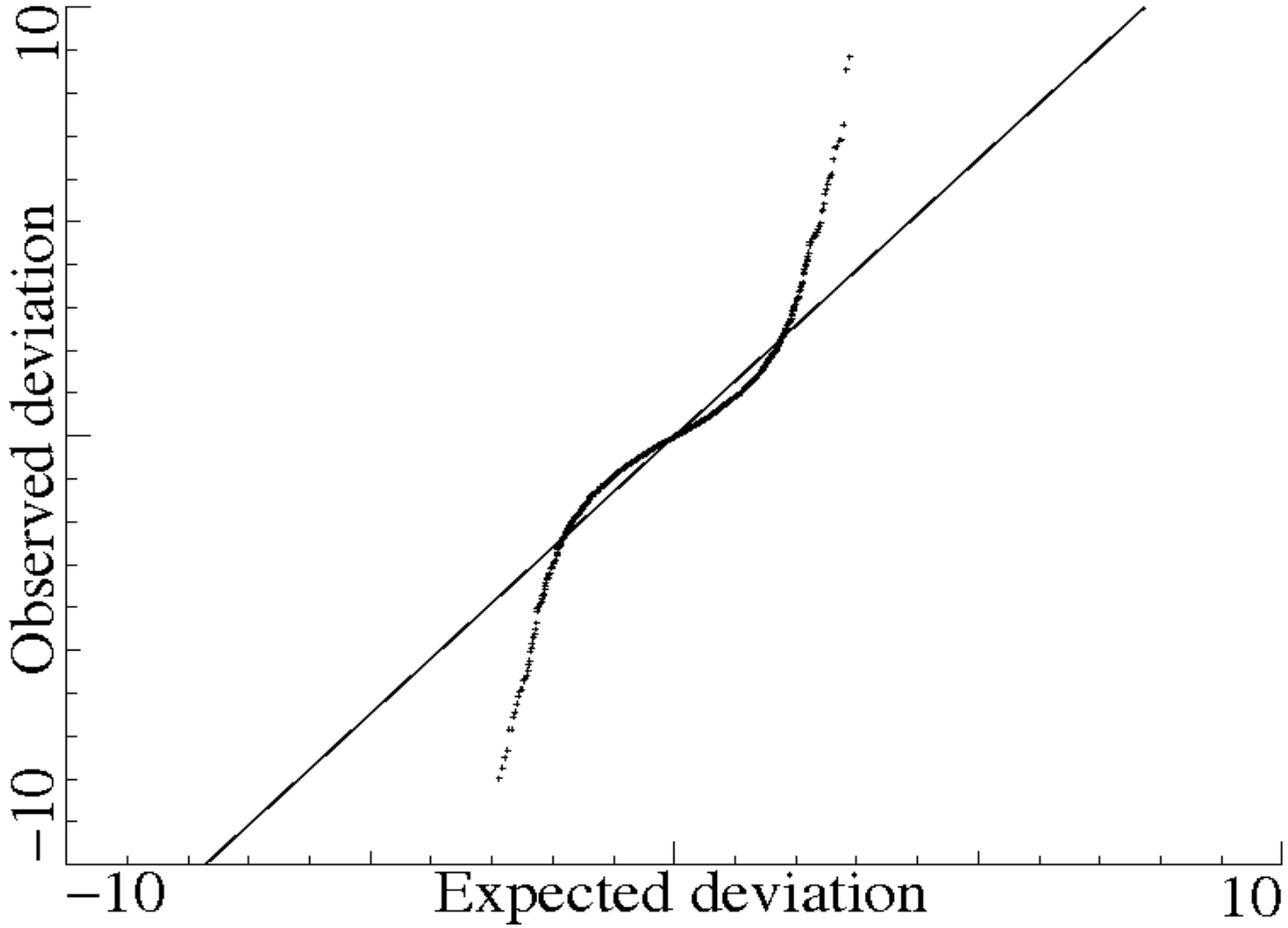
And if we ignore that:

- Most likely an inversion twin
- Value of *y*: 0.42(15)

Error analysis



Normal probability plot



Outlier statistics

One “wrong measurement”, two models: $z_1=10$, $z_2=11$

- Gaussian statistics:
  $p(z_1) = 7\times10^{-23}$, $p(z_2) = 2\times10^{-27}$, ratio > 30000

Broaden the Gaussian view: “Student's t”

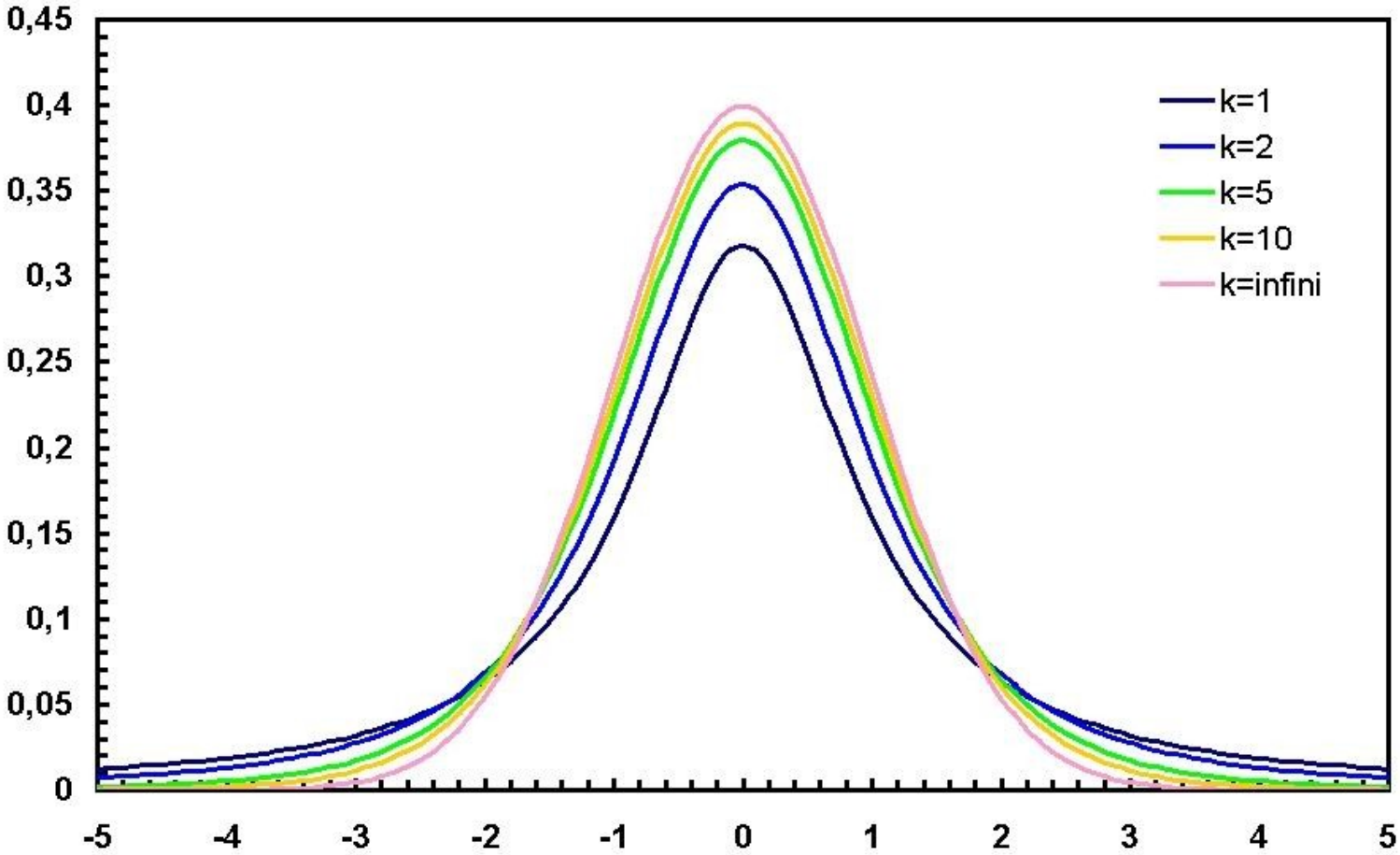


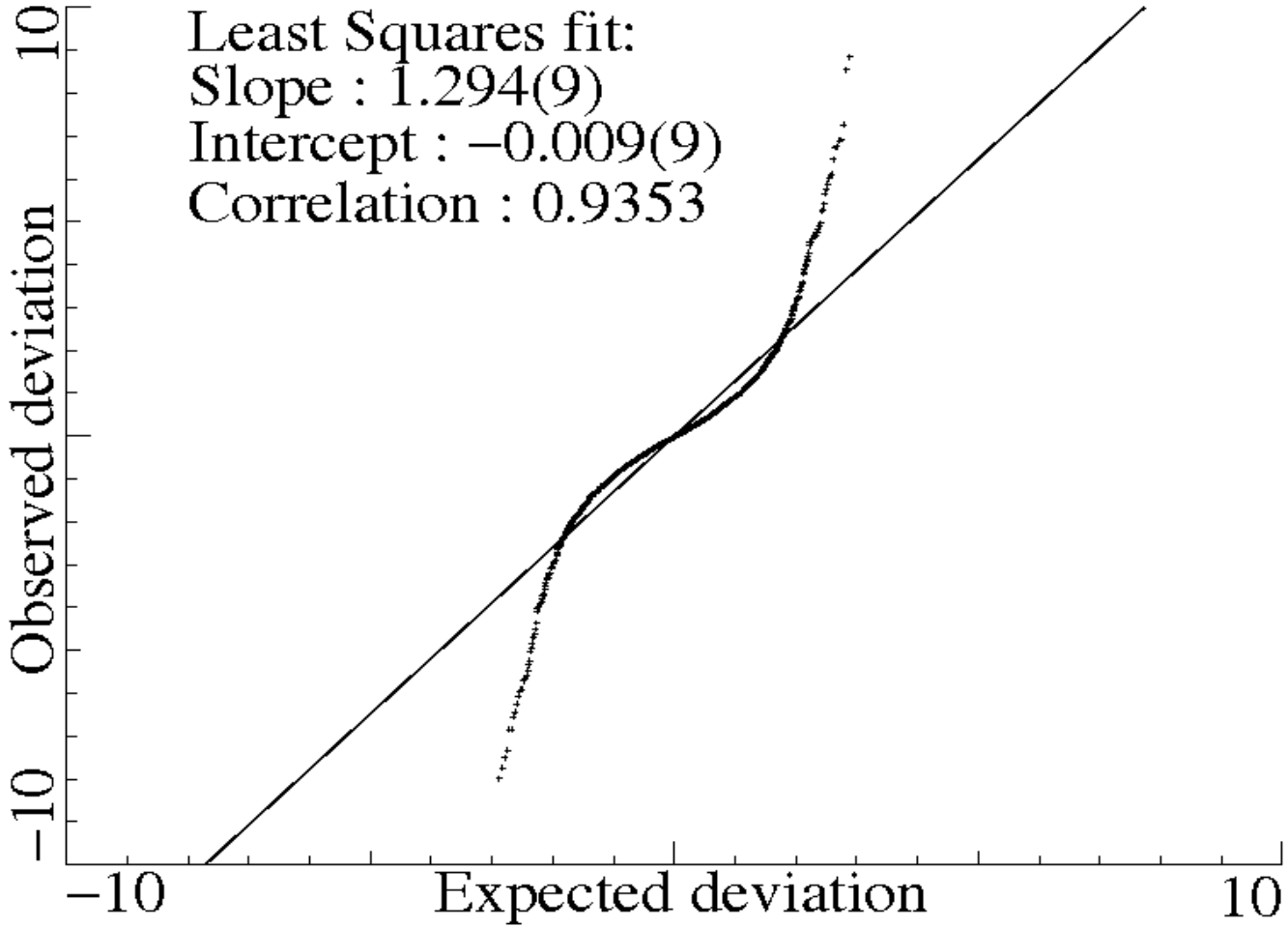
Image by Wikipedia:FR user Thorin

Outlier statistics

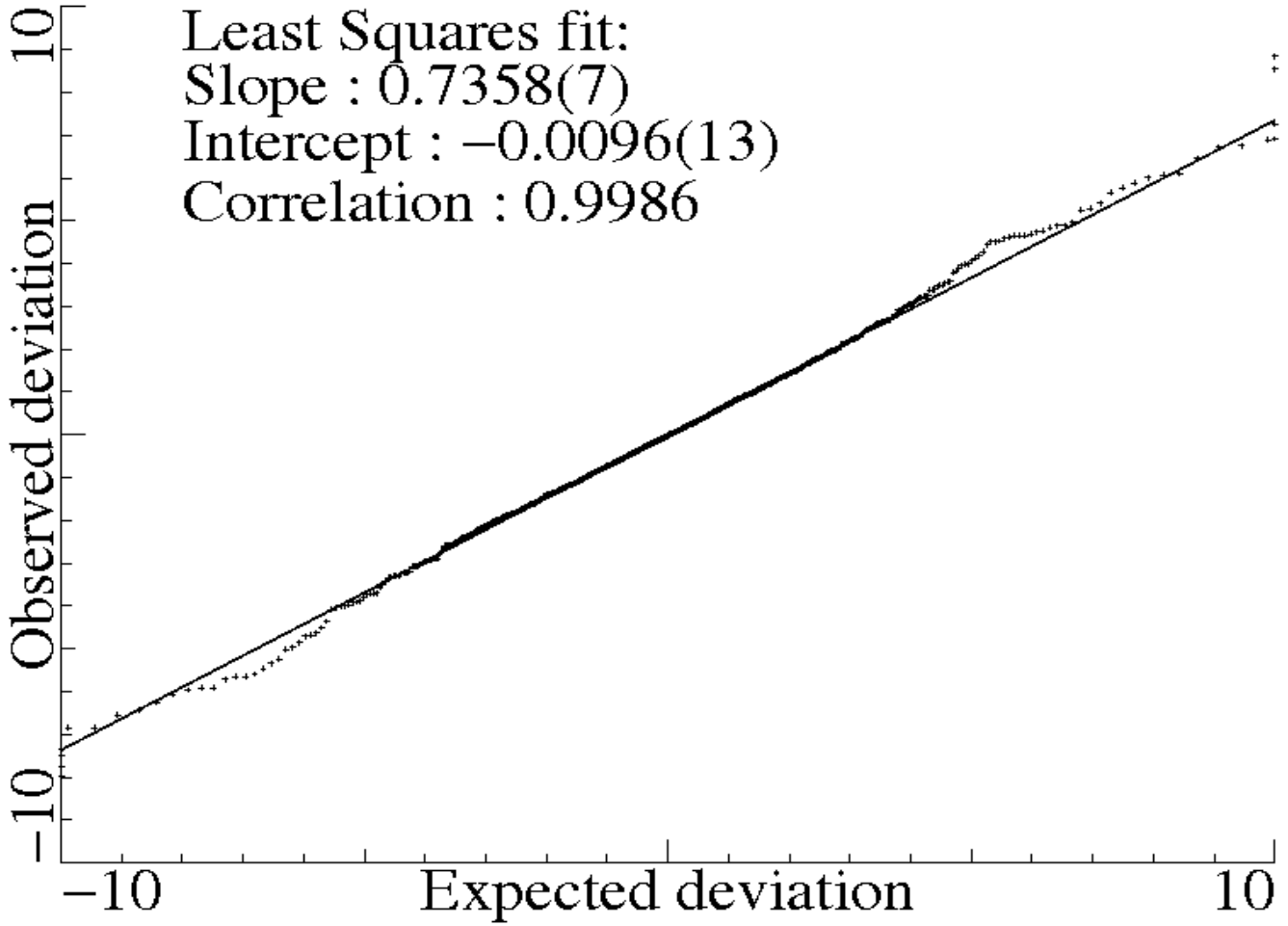
One “wrong measurement”, two models: $z_1=10$, $z_2=11$

- Gaussian distribution:
  $p(z_1) = 7\times10^{-23}$, $p(z_2) = 2\times10^{-27}$, ratio > 30000
- Student's t-distribution, $\nu=10$:
  $p(z_1) = 3\times10^{-9}$, $p(z_2) = 5\times10^{-10}$, ratio = 5
- Student's t-distribution, $\nu=3$:
  $p(z_1) = 3\times10^{-4}$, $p(z_2) = 2\times10^{-4}$, ratio = 1.4

Normal probability plot, revisited

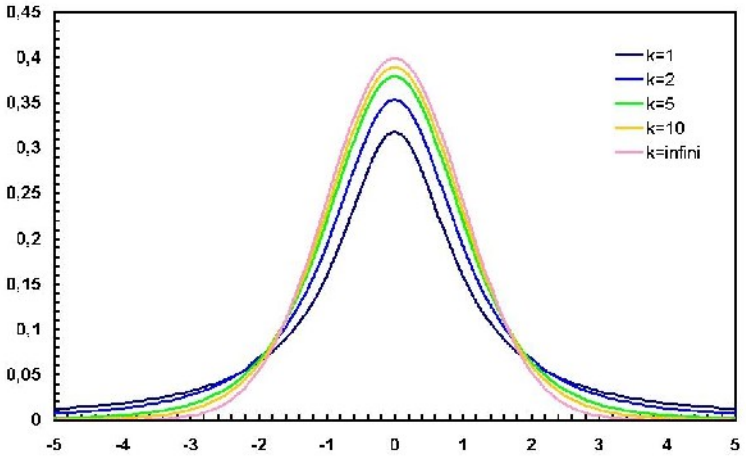


Step 1: Plot using Student's t, ν=2.2



Step 2: The probability distribution

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$



$$p(z,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\ \Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{z^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

Results after moving to Student's t-statistics

Example: molecule containing sulphur, MoKα

- Refined Flack *x* parameter: 0.17(14)

Our optimized analysis gives:

- $\nu = 2.2$
- Probability the model is inverted: 4×10^{-6}
- Value of *y*: 0.10(18)

Experience

A dozen structures tested (3 “red flags”, 9 good):

- Optimized ν values:
  - 2-5 for “red flags”
  - 12-300 for good cases
- Correlation:
  - $0.93 \rightarrow 0.998$; $0.95 \rightarrow 0.991$; $0.995 \rightarrow 0.9996$
  - 2.6x better on average for good structures.
- $\boldsymbol{\sigma}(y)$:
  - Reduced by 7% for bad cases
  - Increased by 1% for good cases

Experience with general racemic mixtures

I, MoKα, f”=1.81, F(000)=1640, $P2_1$

- Flack refinement: x=0.185(9)
- Our analysis: y=0.185(5) [ν=27, r=0.99995]

Pt, MoKα, f”=8.39, F(000)=2631, $P2_12_12_1$

- Flack refinement: x=0.248(5)
- Our analysis: y=0.239(3) [ν=73, r=0.99996]

S, CuKα, f”=0.56, F(000)=1744, $Pca2_1$

- Flack refinement: x=0.346(7)
- Our analysis: y=0.3459(17) [ν=41, r=0.99994]

Can we learn more?

- Standard deviation
  - More relevant than for the y=0 case!
- Calculate relative probability of y=0, y=0.5 or y=1
  - y=0.185(5): p(y=0) = 10^{-258}
  - y=0.239(3): p(y=0) = 10^{-1149}
  - y=0.3459(17): p(y=0.5) = 10^{-1628}
- What else?

Statistical analysis of Bijvoet differences allows absolute structure determination of light atom (CHNO) structures.

Using Student's t-distribution makes results even better.

http://www.absolutestructure.com/

Try it on your structures, and let me know about the results.

Rob.Hooft@Bruker-AXS.nl

http://www.absolutestructure.com/