UNIVERSITAT POLITÈCNICA DE CATALUNYA

FABADA:
a Fitting Algorithm for Bayesian Analysis of DAta

obtaining molecular structure from diffraction data

Luis Carlos Pardo Soto
Grup de Caracterització de Materials (GCM)

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- Some examples
  - ✓ Analysis of QENS spectra
  - ✓ Model selection using QENS data
  - ✓ Intramolecular structure determination
- Summary and conclusions

Bayes theorem

Likelihood
Probability that our data describes the hypothesis

Prior
Our forehand knowledge

$$P(H \mid D) = \frac{P(D \mid H) \cdot P(H)}{P(D)}$$

T. Bayes.

Posterior
Probability that the hypothesis is true given the experimental data

Evidence
Normalization factor

In our case is very simple...

Bayes theorem

Likelihood
Probability that our data describes the hypothesis

Maximum ignorance Prior

$$P(H \mid D) \propto \frac{P(D \mid H) \cdot P(H)}{P(D)}$$

T. Bayes.

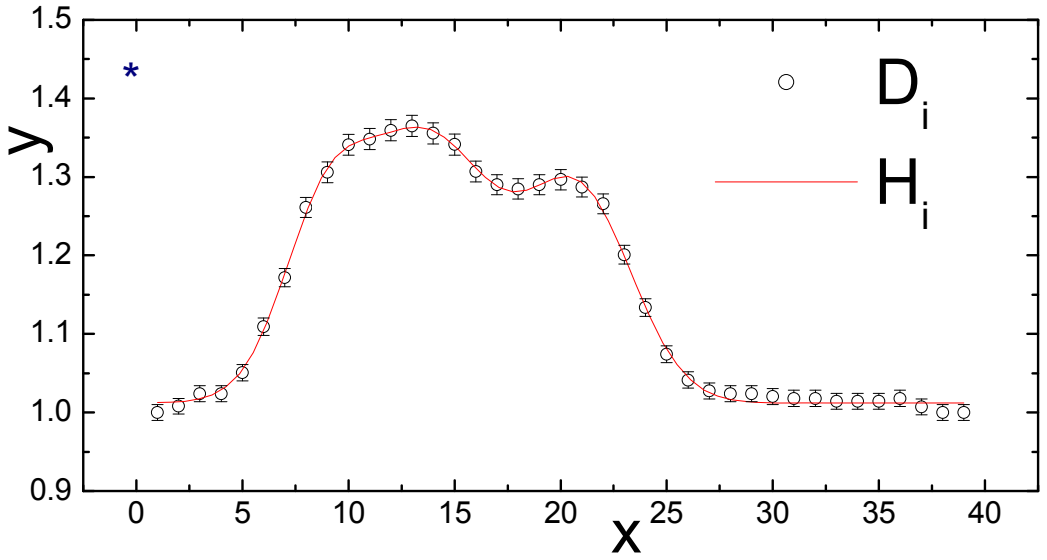
Posterior
Probability that the hypothesis is true given the experimental data

We only care about proportionality

Bayes theorem

$$P(H \mid D) \propto P(D \mid H) \equiv L$$

D_i Data (i=1,n) $H_i\{P_l\}$ Hypothesis (i=1,n) using a parameter set $\{P_l\}$ (l=1,m)



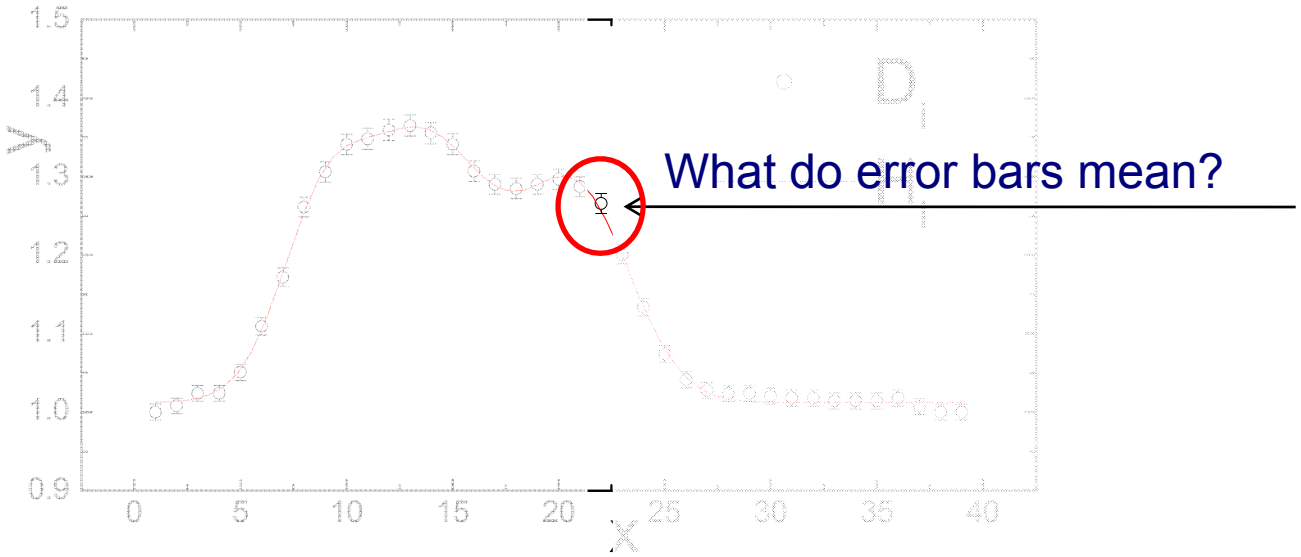
$$P(H_i\{P_l\} \mid D_i) \propto P(D_i \mid H_i\{P_l\})$$

* Figure adapted from „Le petit prince“ A. Saint Exupery (1943)

Bayes theorem

$$P(H \mid D) \propto P(D \mid H)$$

D_i Data (i=1,n) $H_i\{P_l\}$ Hypothesis (i=1,n) using a parameter set $\{P_l\}$ (l=1,m)

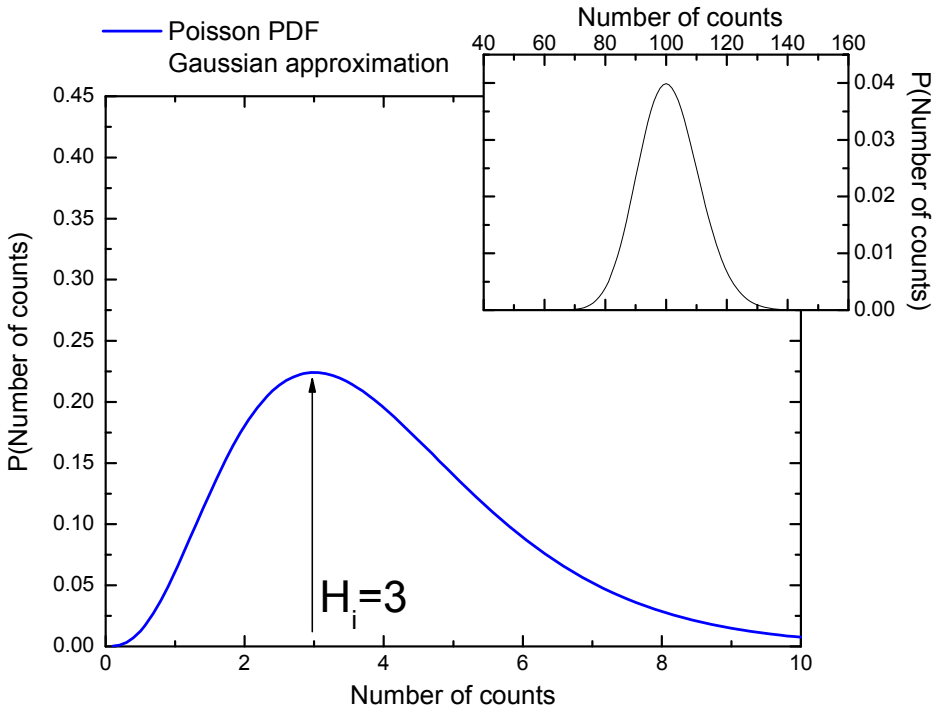


$$P(H_i\{P_l\} \mid D_i) \propto P(D_i \mid H_i\{P_l\})$$

The ubiquitous χ^2

In a counting experiment, if the expected value is H_i measured values will follow a Poisson statistics

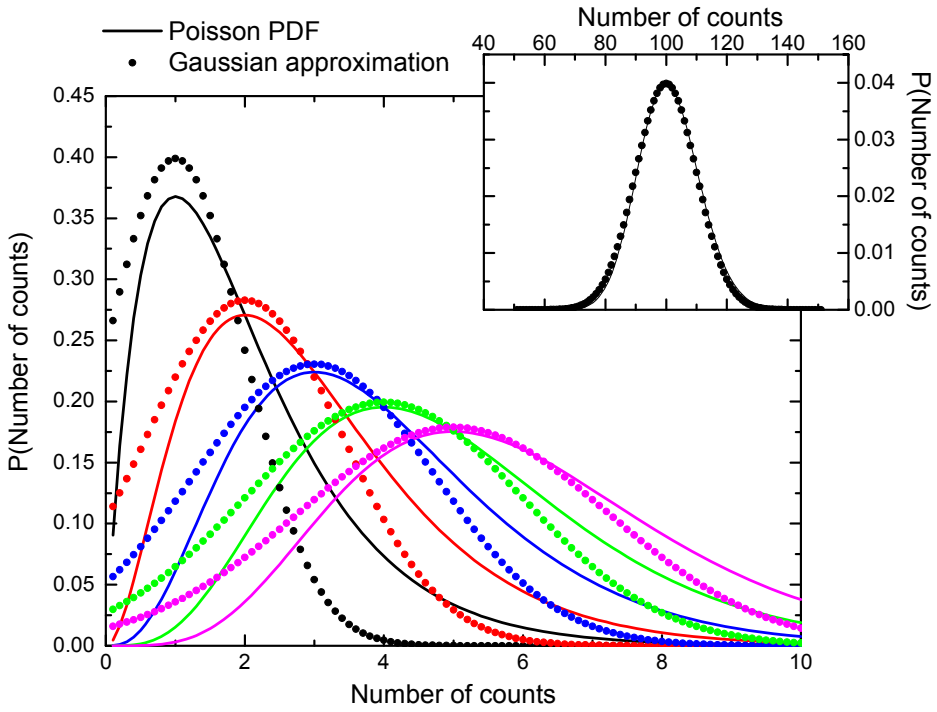
$$P(D_{i=k} \mid H_{i=k}) \propto H_i^{Di=k} e^{-Hi=k}$$



The ubiquitous χ^2

In a counting experiment, if the expected value is H_i measured values will follow a Poisson statistics

$$P(D_{i=k} \mid H_{i=k}) \propto H_i^{Di=k} e^{-Hi=k}$$



If the number of counts is high enough poisson statistics =normal distribution with $\sigma_i = \sqrt{D_i}$

$$P(D_{i=k} \mid H_{i=k}) \propto \exp - \frac{(H_{i=k} - D_{i=k})^2}{2\sigma_{i=k}^2}$$

The error is the square root of the variance $\varepsilon_i = \sigma_i$

Bayes theorem $P(H_i\{P_l\} \mid D_i) \propto P(D_i \mid H_i\{P_l\})$

We now consider all the points i=1,2,... n

$$L = P(D_i \mid H_i\{P_l\}) \propto \prod_{i=1}^{n} \exp - \frac{(H_i - D_i)^2}{2\sigma_i^2}$$

$$= \exp \sum_{i=1}^{n} - \frac{(H_i - D_i)^2}{2\sigma_i^2} = \exp - \frac{\chi^2}{2}$$

Therefore χ^2 is related to the likelihood:

$$\chi^2 \propto -2 \cdot \ln L$$

So, we got the exact meaning on probability bases of χ^2
Let's use it!

Wandering in parameter space χ^2

Similar to Montecarlo simulations, (to compare let's assume that all errors σ_i are equal)

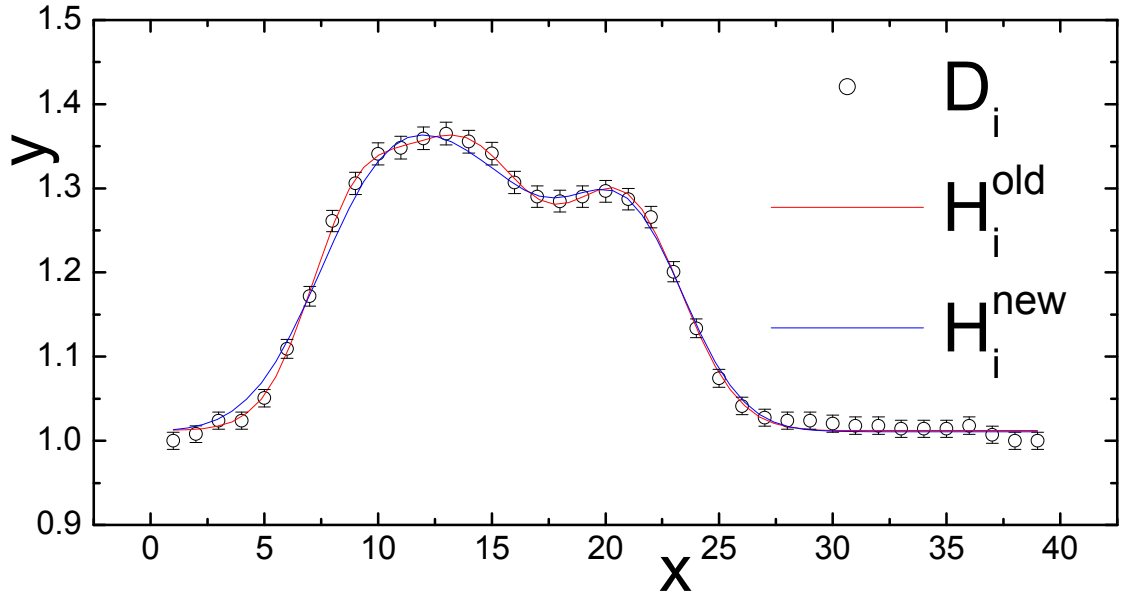
$$\frac{P\left(H_i\{P_l^{new}\}\,|\,D_i\right)}{P\left(H_i\{P_l^{old}\}\,|\,D_i\right)} = \exp -\frac{\sum_{i=1}^{n}\left(H_i^{new}-D_i\right)^2 - \sum_{i=1}^{n}\left(H_i^{old}-D_i\right)^2}{2\sigma^2} = \exp -\frac{\left(\chi^2_{new}-\chi^2_{old}\right)}{2}$$

This term makes the fitting better

$$E \leftrightarrow \sum_{i=1}^{n}\left(H_i^{new}-D_i\right)^2$$

This term allows parameters that make the fitting worst

$$T \leftrightarrow 2\sigma^2$$



This is a general method: it is also used in Reverse MonteCarlo, *and many other cases!*

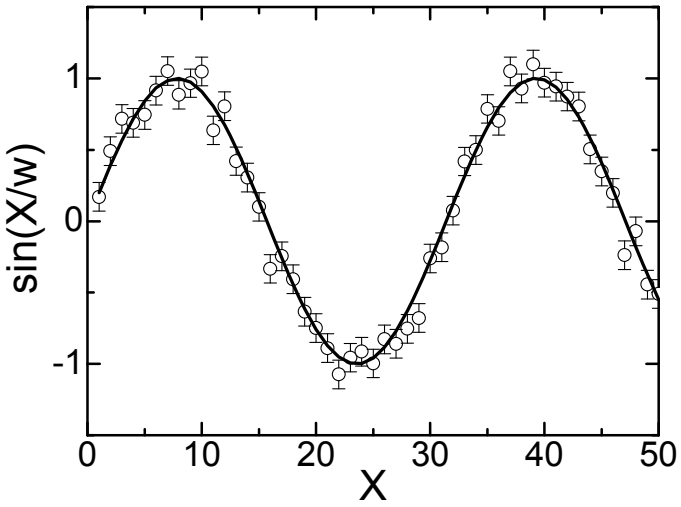
Simulated annealing

a few words on simulated annealing...

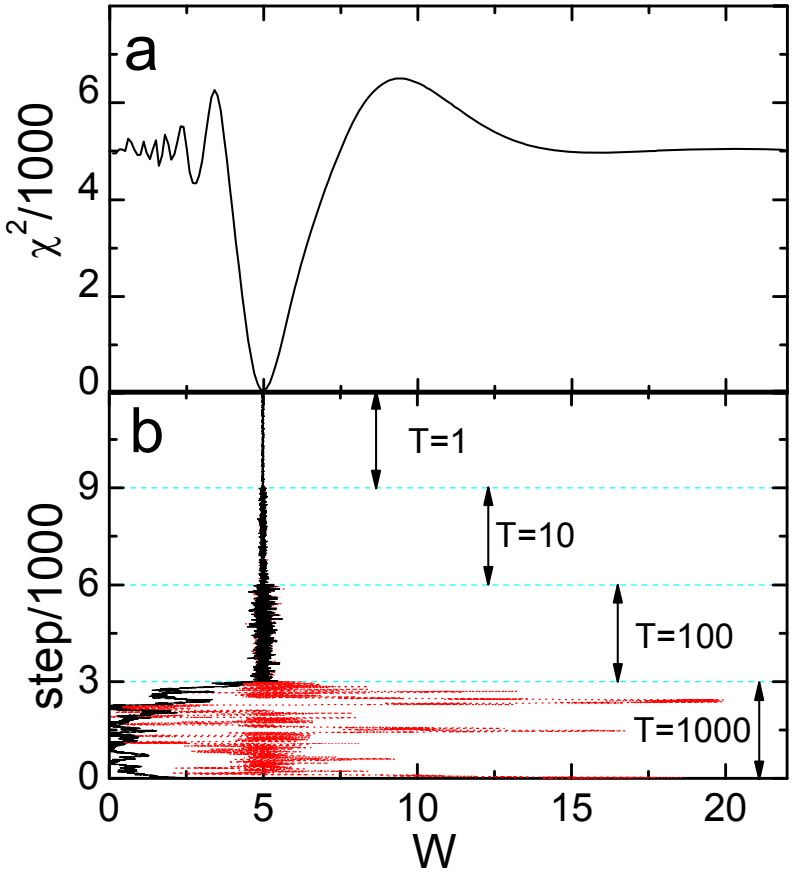
$$\frac{P\left(H_i\{P_l^{new}\}\,|\,D_i\right)}{P\left(H_i\{P_l^{old}\}\,|\,D_i\right)} = \exp-\frac{\left(\chi^2_{new}-\chi^2_{old}\right)}{2T}$$

a test case

$$f(x)=\sin(x/W);\quad W=5$$



„Cooling the fit“

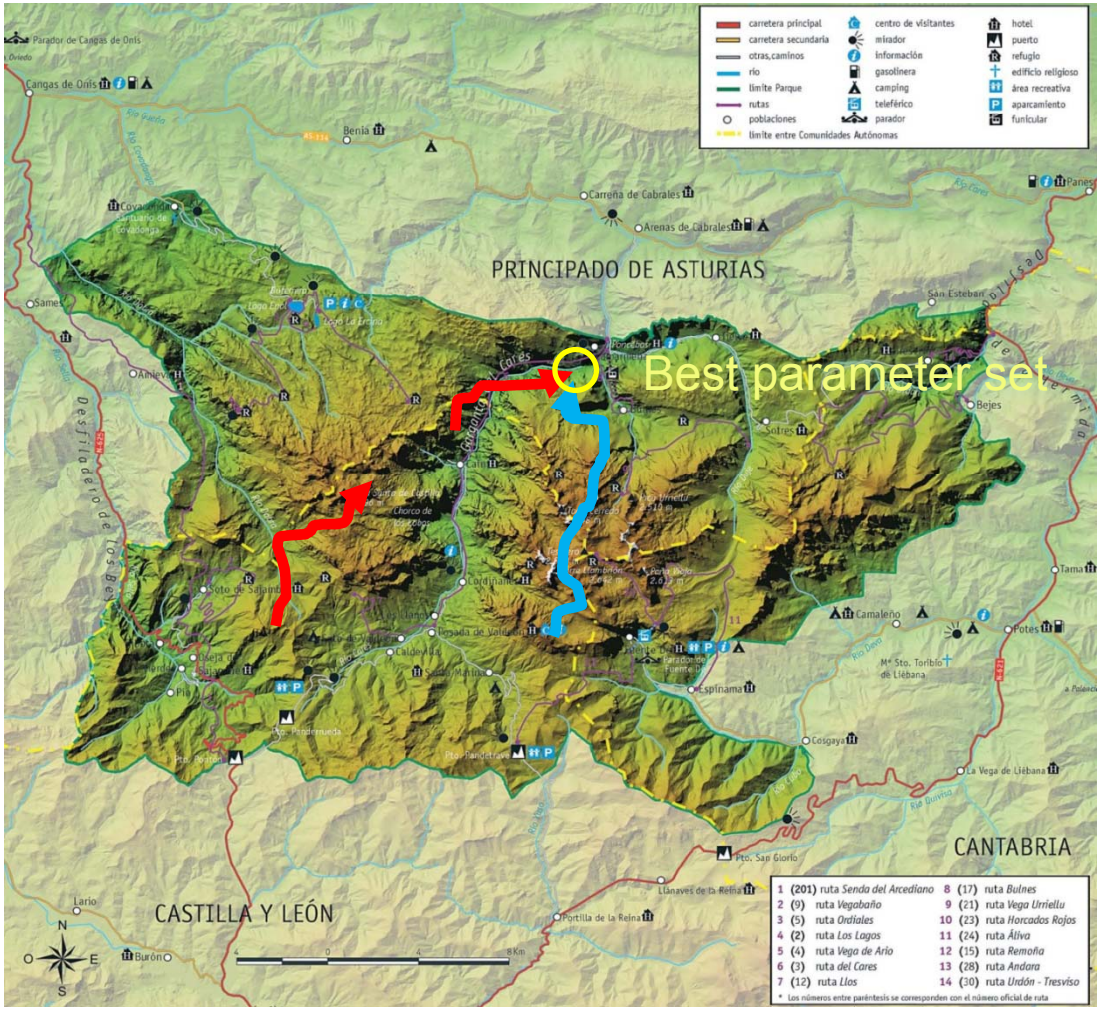


L.C. Pardo et al. Phys. Rev. E. (2011, in press)

- The ubiquitous χ^2
- Advantages of Bayesian analysis
  - The fitting process
  - Parameter estimation
  - Model selection

Fitting in $\chi^2\{P_1\}$ landscape

Classical fiting



Bayessian fiting

It does not get stuck!!!

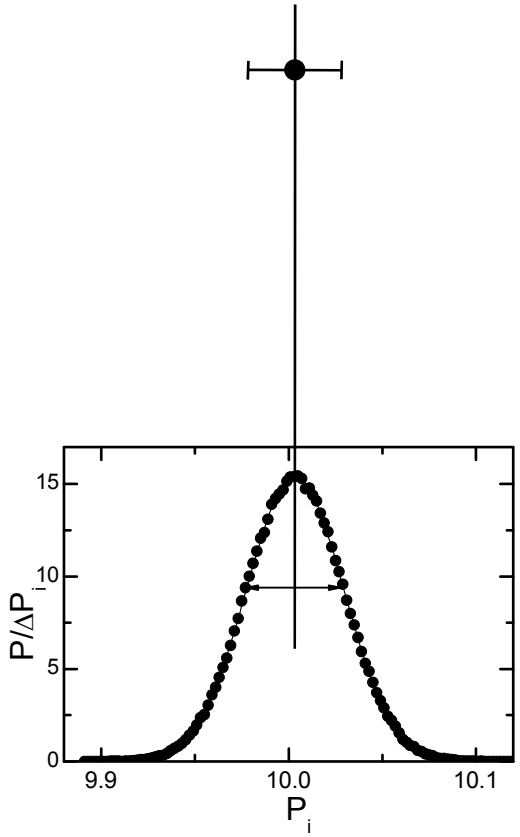
- The ubiquitous χ^2
- Advantages of Bayesian analysis
  - The fitting process
  - Parameter estimation
  - Model selection

Parameter determination

"Classic"

(frequentist)

$$P \pm \delta P$$

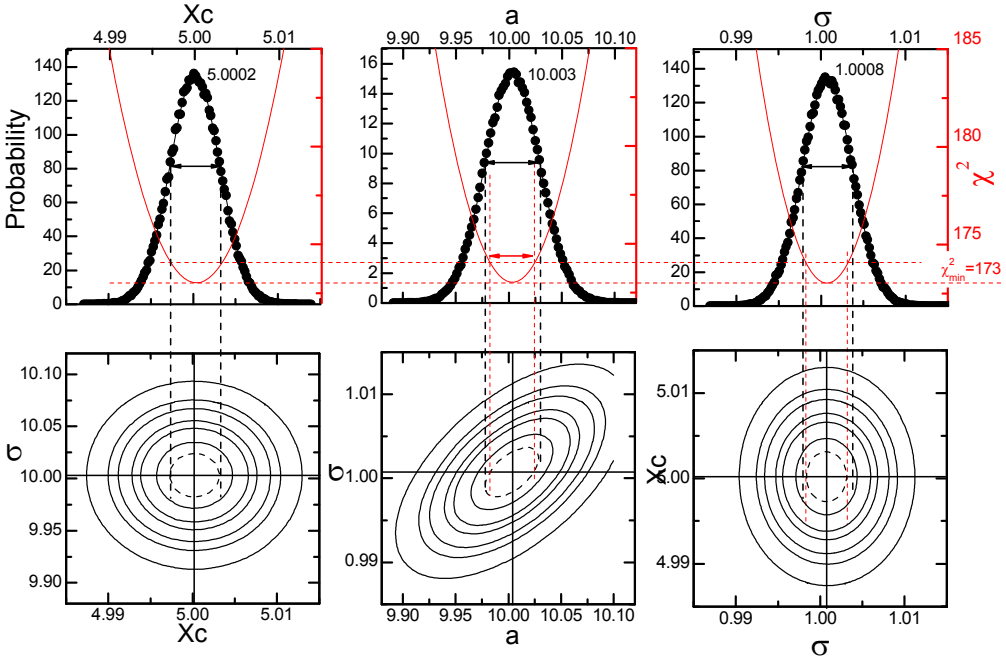


Bayesian

Probability Density Function (PDF)

Parameter determination

Correlation between parameters
are automatically had into account



Fit of a "simple" Gaussian

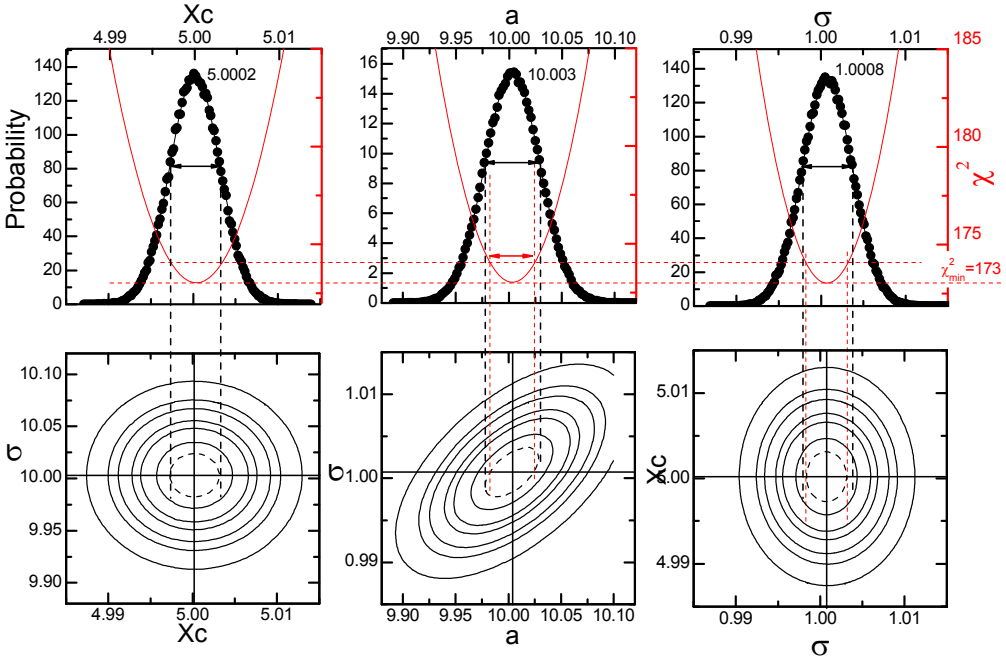
$$f(x) = a \cdot \exp - \frac{(x - x_c)^2}{2\sigma^2}$$

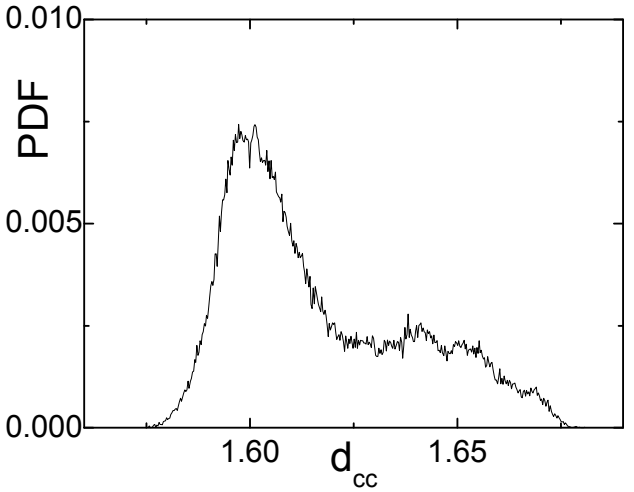
Journal of physics: conference series (in press)

Parameter determination

Correlation between parameters are automatically had into account

No supposition on the minimum geometry is made





Fit of a simple Gaussian

$$f(x) = a \cdot \exp - \frac{(x - x_c)^2}{2\sigma^2}$$

It might be as terrible as this one…

Journal of physics: conference series (in press)

- The ubiquitous χ^2
- Advantages of Bayesian analysis
  - The fitting process
  - Parameter estimation
  - Model selection

Model Selection

Usual methods

- The "guide to the eye" method
- The reduced χ^2 method

$$\chi^2_{red} = \frac{\chi^2}{n-m}$$

n: is the number of points
m: is the number of parameters

This works only if:

✓ There is no correlation between parameters
✓ The PDF in **all** parameters is gaussian
✓ The minimum is not multimodal

Model Selection

Usual methods

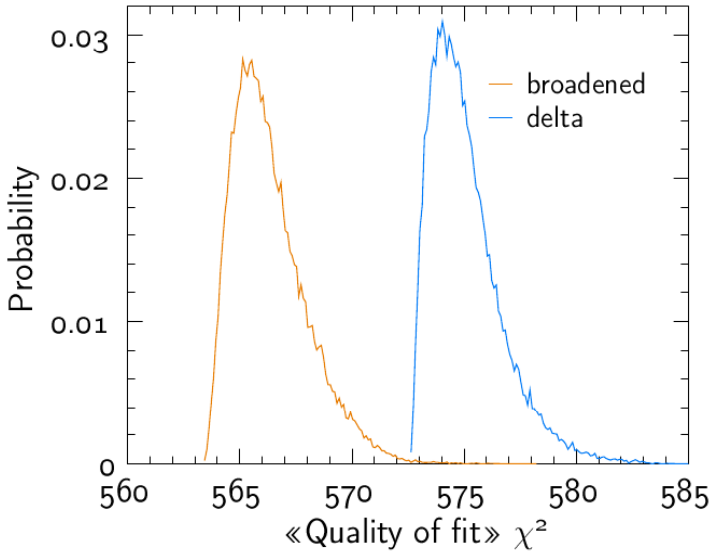
- The "guide to the eye" method
- The reduced χ^2 method

$$\chi^2_{red} = \frac{\chi^2}{n-m}$$

n: is the number of points
m: is the number of parameters

Bayesian method

- Directly compares the PDF related to χ^2



This works only if:

- ✓ There is no correlation between parameters
- ✓ The PDF in **all** parameters is gaussian
- ✓ The minimum is not multimodal

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- Some examples
  - ✓ Analysis of QENS spectra
  - ✓ Model selection using QENS data
  - ✓ Intramolecular structure determination
- Summary and conclusions

Analysis of QENS spectra

Is there any change in the dynamics?

Dynamics transition in trans-dicloroethylene

We analyze TOF spectra with a diffussion+rotation model

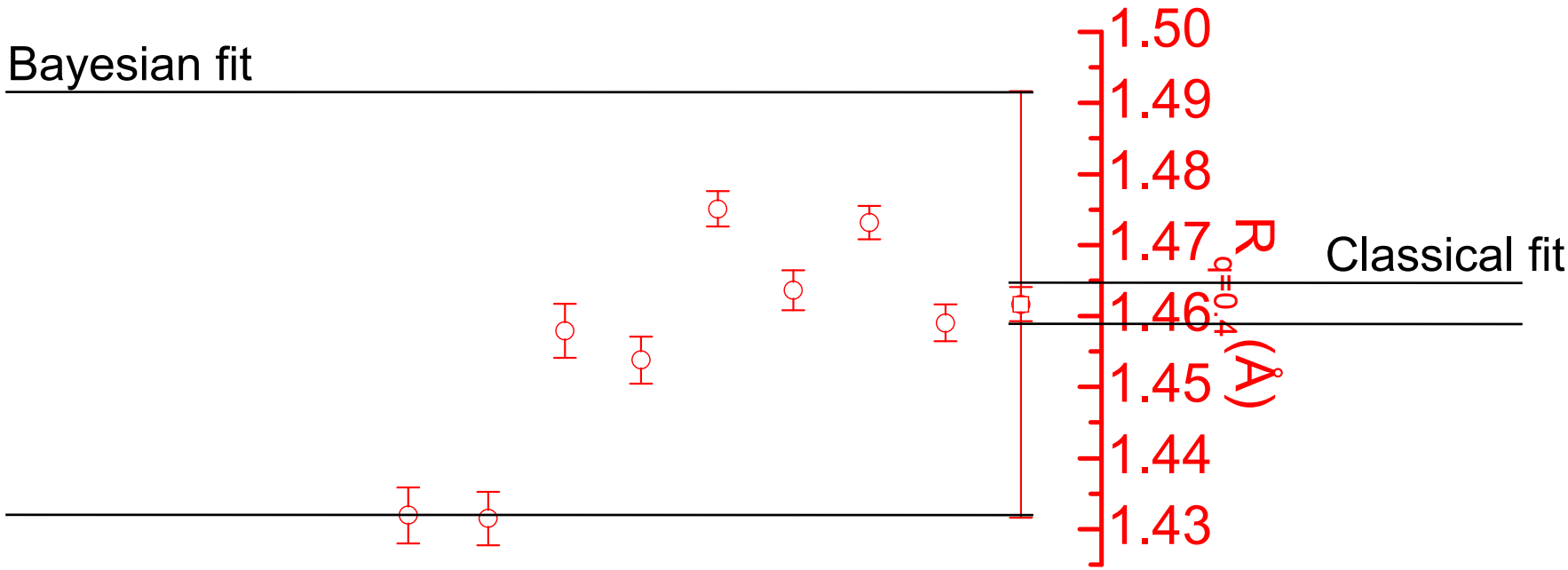
$$S_{rot}(q,\omega) = A_0(qR)\cdot\delta(\omega) + \sum_l A_l(qR)\cdot L_l(w,\gamma = l(l+1)D_r)$$

$$S(q,\omega) = S_{diff}(q,\omega) \otimes S_{rot}(q,\omega) \otimes R(q,\omega)$$

$$S_{diff}(q,\omega) = L(w,\gamma = Dq^2)$$

Rovira-Esteva M. et al. Phys. Rev. B 81(9) 092202 (2010)

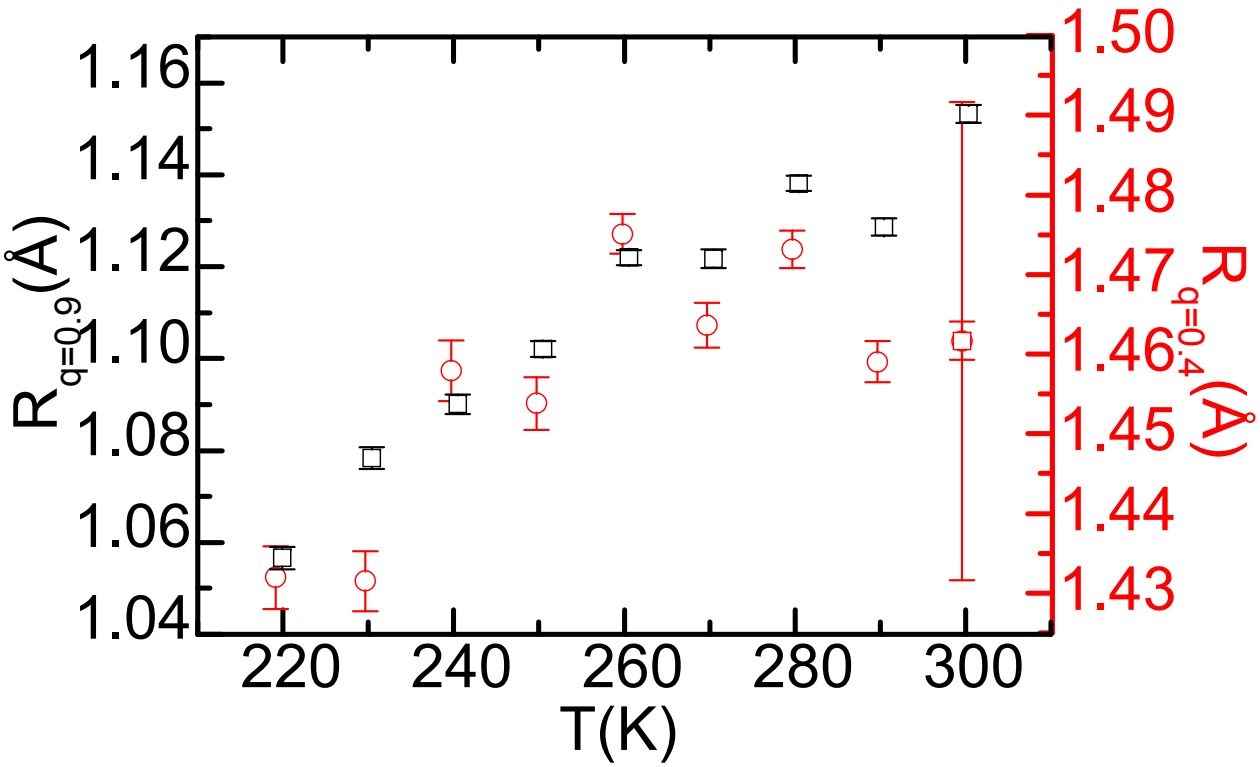
We analyze just one q value



Usually errors are underestimated using classical methods!!

Rovira-Esteva M. et al. Phys. Rev. B 81(9) 092202 (2010)

We analyze just one q value



Rovira-Esteva M. et al. Phys. Rev. B 81(9) 092202 (2010)

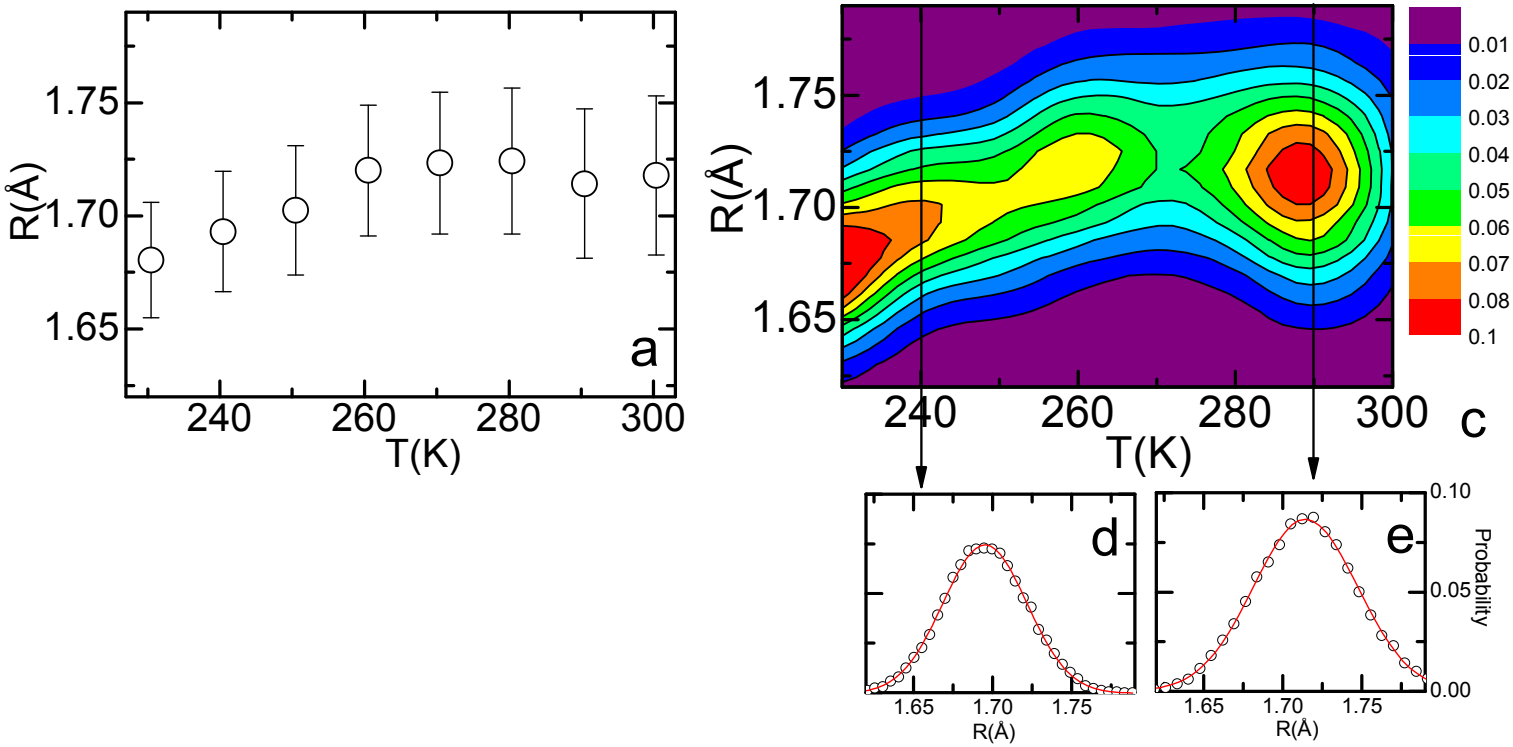
Analysis of QENS spectra

All q values at once

the whole scattering law S(q,ω)

Classical representation

PDF representation

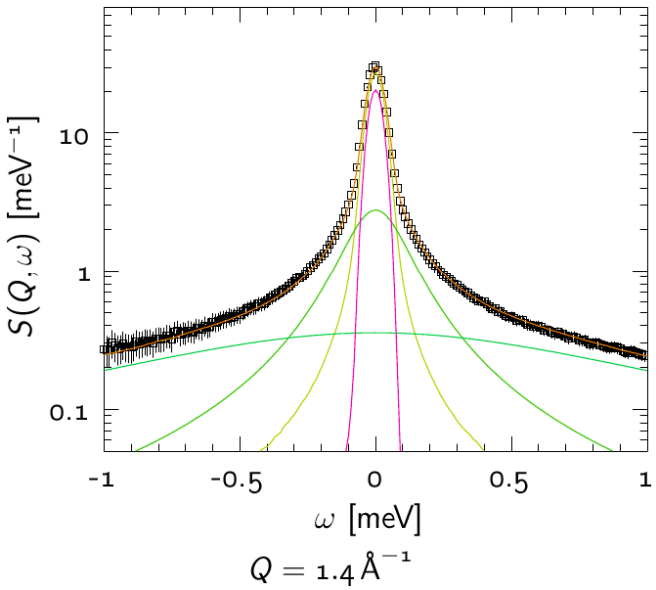


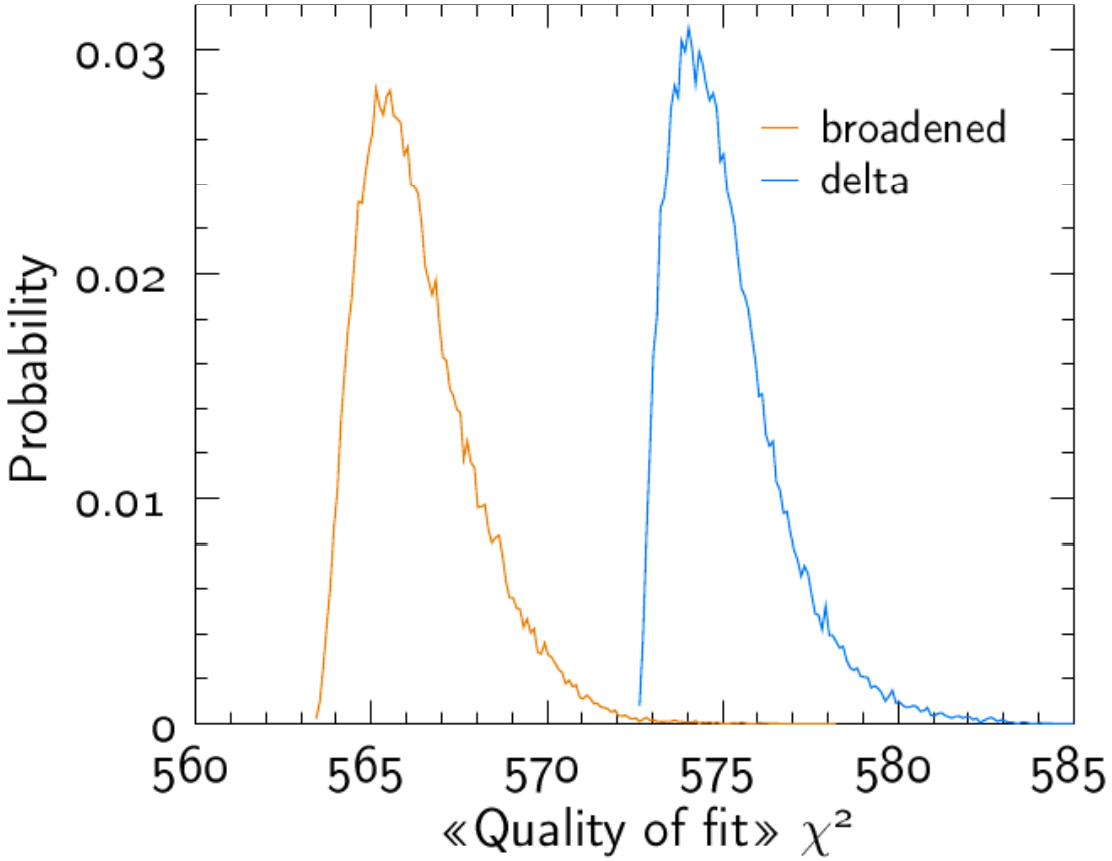
Rovira-Esteva M. et al. Phys. Rev. B 81(9) 092202 (2010)

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- **Some examples**
  - ✓ Analysis of QENS spectra
  - ✓ **Model selection using QENS data**
  - ✓ Intramolecular structure determination
- Summary and conclusions

Motion of phospholipids in the membrane

Is there a broadening? Delta model versus Broadened model



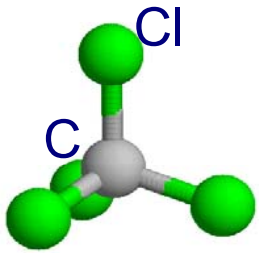


S. Busch et al. J. Am. Chem. Soc. 132(10) 3232 (2010)

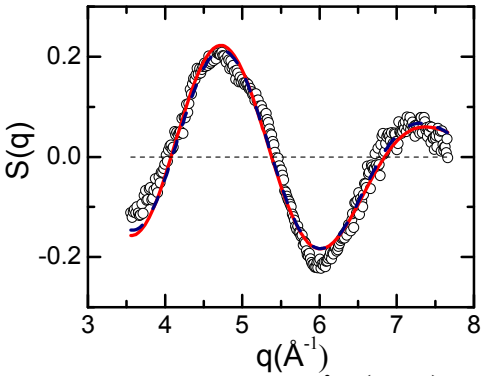
- The ubiquitous χ^2
- Advantages of Bayesian analysis
- **Some examples**
  - ✓ Analysis of QENS spectra
  - ✓ Model selection using QENS data
  - ✓ **Intramolecular structure determination**
- Summary and conclusions

Intramolecular structure

A simple case: CCl_4



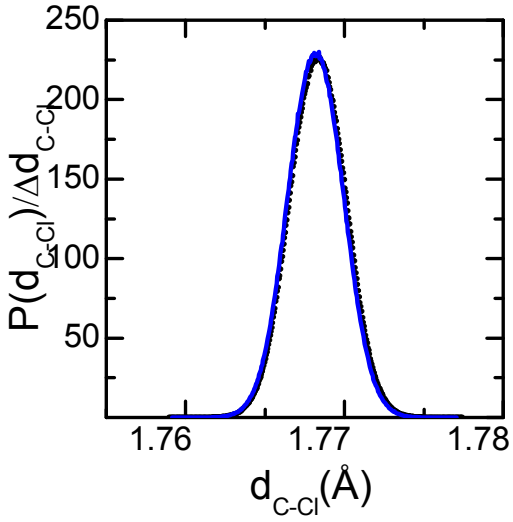
Molecular structure
fit of the high q region of s(q)



$$S(q) = h \cdot \sum_{i,j}^{m} b_i b_j \cdot \frac{\sin(qr_{ij})}{qr_{ij}} \cdot e^{-u_{ij}^2 q^2 / 2}$$

---

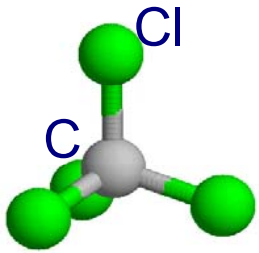
d_{CCl} distance is well defined



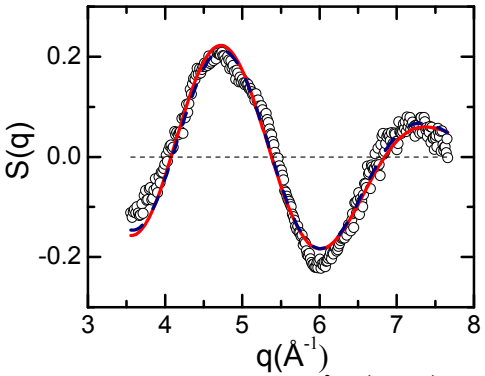
Pardo L.C. Et al. arXiv:0907.3711v3 [physics.data-an]

Intramolecular structure

A simple case: CCl_4



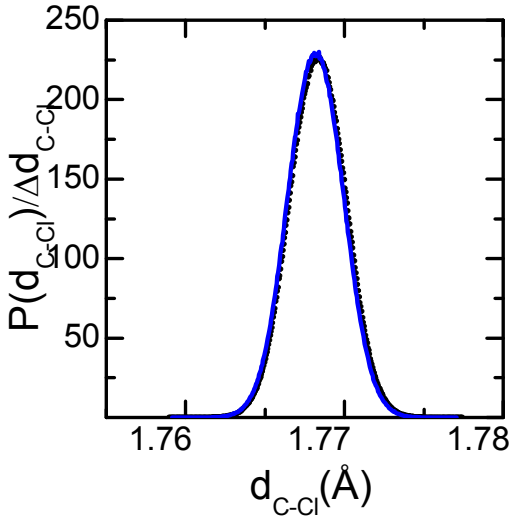
Molecular structure
fit of the high q region of s(q)



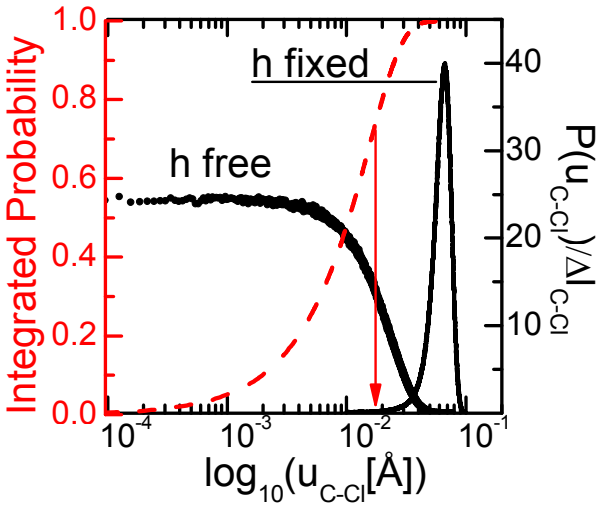
$$S(q) = h \cdot \sum_{i,j}^{m} b_i b_j \cdot \frac{\sin(qr_{ij})}{qr_{ij}} \cdot e^{-u_{ij}^2 q^2 / 2}$$

---

d_{CCl} distance is well defined

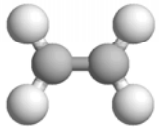


$U_{C\text{-}Cl}$ depends on the scaling factor h



Pardo L.C. Et al. arXiv:0907.3711v3 [physics.data-an]

A more complicated test case: C_2D_4



fit of the high q region of s(q)

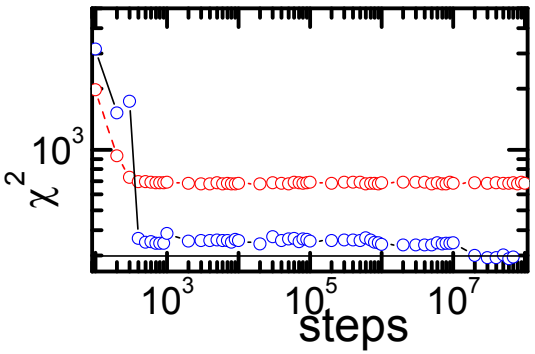
$$S(q) = h \cdot \sum_{i,j}^{m} b_i b_j \cdot \frac{\sin(qr_{ij})}{qr_{ij}} \cdot e^{-u_{ij}^2 q^2 / 2}$$

$\longleftrightarrow$

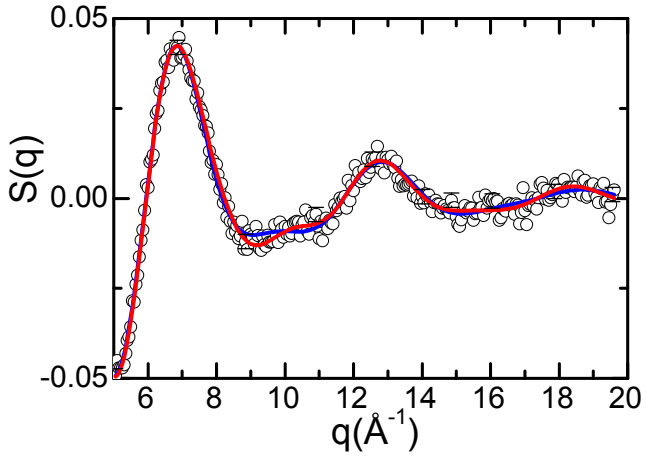
fit of the small r region of G(r)

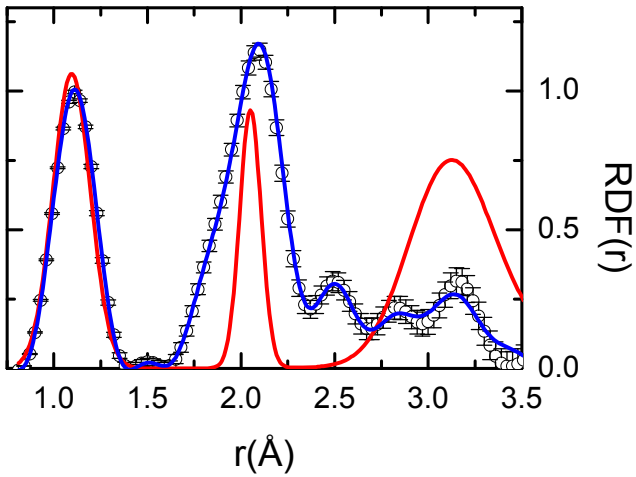
$$G(r) \approx h \cdot \sum_{i,j}^{m} b_i b_j \cdot \exp\left(-\frac{(r - r_{ij})^2}{2u_{ij}^2}\right)$$

Fit using only s(q)



Fit using both s(q) and g(r)





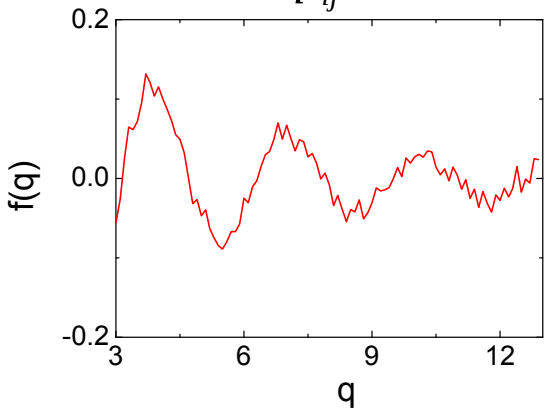
Why is it better fitting both at the same time?

M. Rovira-Esteva et al. Phys. Rev. B 84, 064202 (2011)

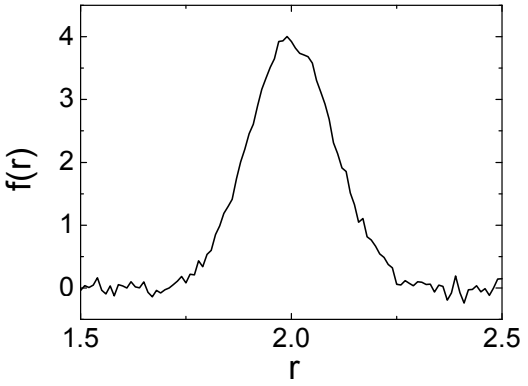
Why is it better fitting both at the same time?

The simplest case (with U_{ij}=0.1, r_{ij}=2):

$$S(q) = \frac{\sin(qr_{ij})}{qr_{ij}} \cdot e^{-u_{ij}^2 q^2 / 2}$$



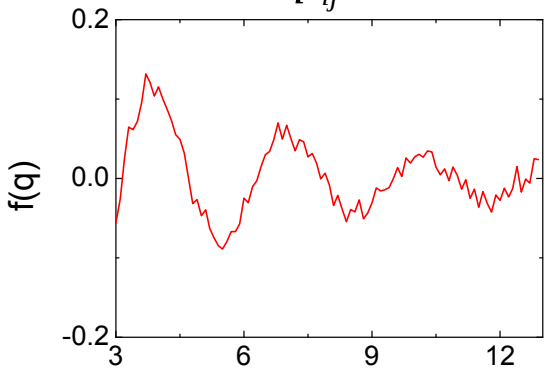
$$RDF(r) = \exp\left(-\frac{(r - r_{ij})^2}{2u_{ij}^2}\right)$$



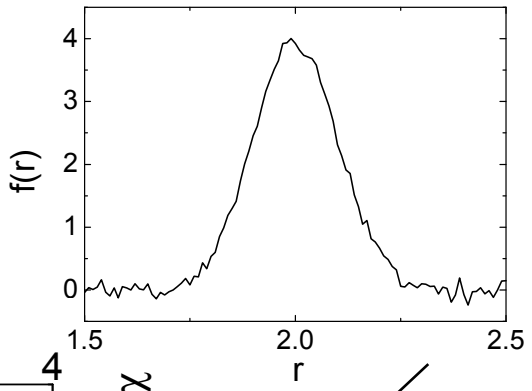
Why is it better fitting both at the same time?

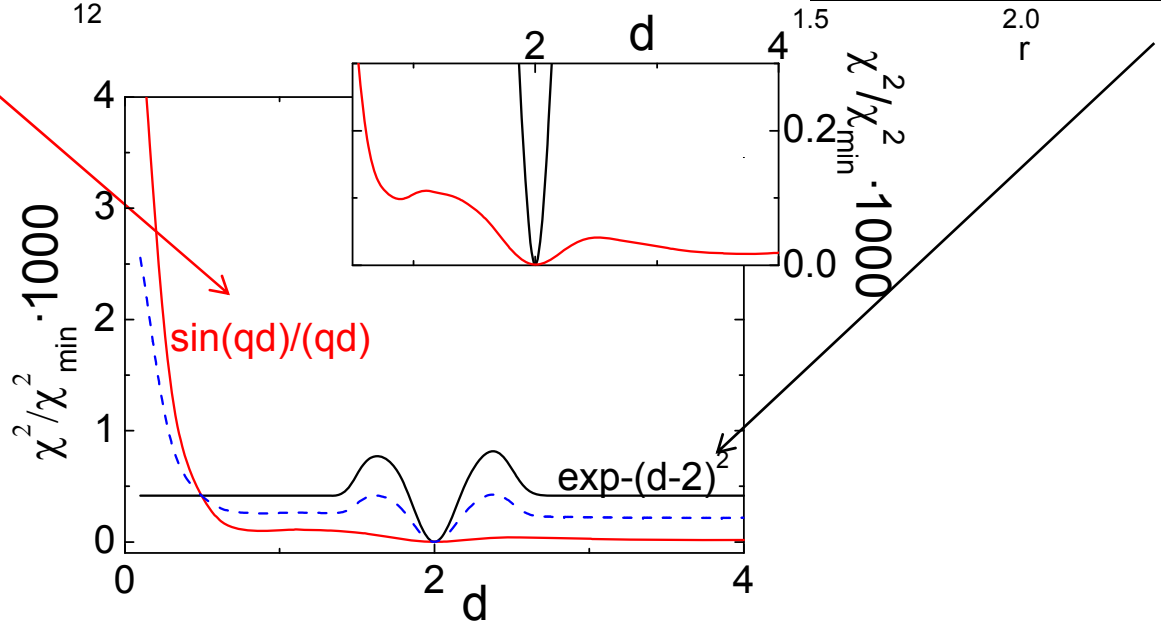
The simplest case (with U_{ij}=0.1, r_{ij}=2):

$$S(q) = \frac{\sin(qr_{ij})}{qr_{ij}} \cdot e^{-u_{ij}^2 q^2 / 2}$$



$$RDF(r) = \exp\left(-\frac{(r-r_{ij})^2}{2u_{ij}^2}\right)$$

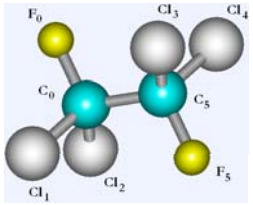




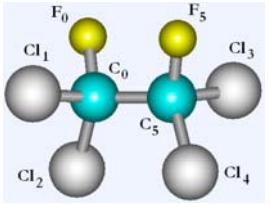
χ^2 landscapes are quite different!!

A really complicated case: $C_2Cl_4F_2$

It has two conformers

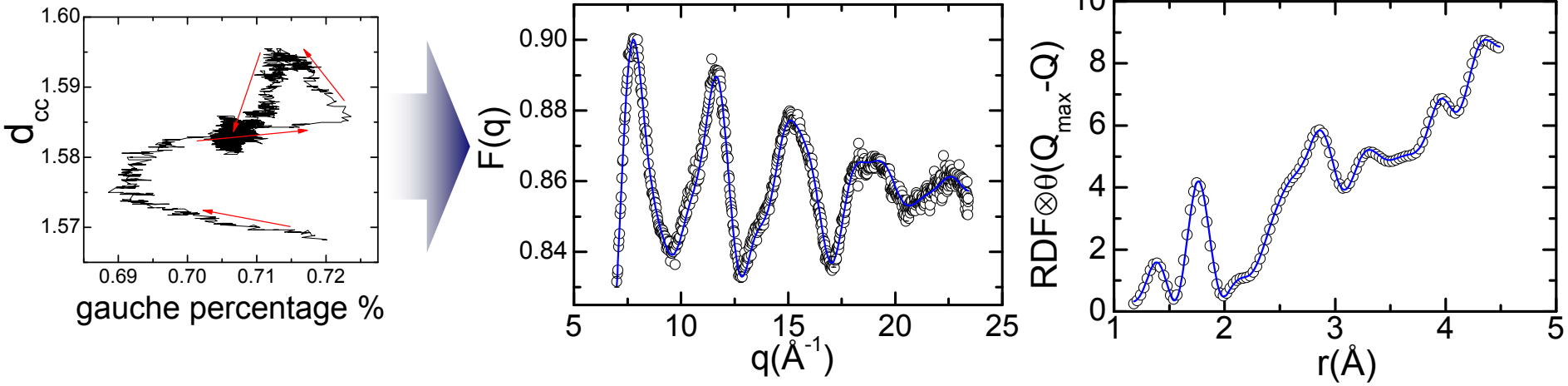


trans



gauche

We have to fit **37 parameters**, and they are not independent:
for example a change in d_{cc} implies changes in the whole molecule



We simply let the program to wander through the parameter space...

M. Rovira-Esteva et al. Phys. Rev. B 84, 064202 (2011)

Summary and conclusions

	Classical fit	Bayesian "Fit"
Fitting process	Only downhill changes of χ^2 are allowed	Uphill changes of χ^2 are allowed
Parameter determination	Parameters and their error $P_i \pm \varepsilon_i$	Probability Distribution Function for P_i
	Correlation between parameters are not automatically taken into account	Correlation between parameters are taken into account
	Multiple minima are invisible	Multiple minima are visible
Model selection	A χ^2 PDF is assumed	No assumption is made on χ^2 PDF

Summary and conclusions

When is it **not** worth to work with Bayesian analysis

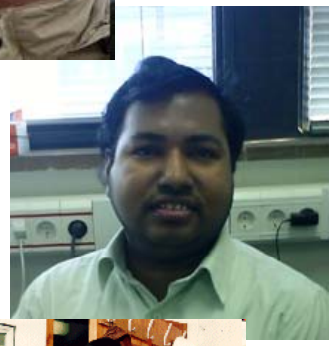
- When you have a simple function, with few parameters
- When parameters can be initialized close to the solution
- When model selection “done by eye” is evident (being m equal!!!)

When is it worth to work with Bayesian analysis

- When your function has too many parameters
- When fit gets stuck every now and then
- When model selection is not evident
- When you have different number of parameters for each model

acknowledgements









We have an open PhD position!

Thank you for your attention

To read more about the examples

- *M. Rovira-Esteva et al. Phys. Rev. B 81(9) 092202 (2010)*
- *M. Rovira-Esteva et al. Phys. Rev. B 84, 064202 (2011)*
- *S. Busch et al. J. Am. Chem. Soc. 132(10) 3232 (2010)*
- *J. C. Martínez et al. J. Phys. Chem. B 114 6099 (2010)*

To read more about FABADA and download it

- ***L.C. Pardo et al. Phys. Rev. E. (2011,** in press)*
- *L.C Pardo et al. J. Phys.: Conf. Ser.(2011, in press)*
- *L. C. Pardo et al. arXiv:0907.3711v3 [physics.data-an]*

- *Download the program, and see these slides: http://gcm.upc.edu/members/luis-carlos/bayesiano*