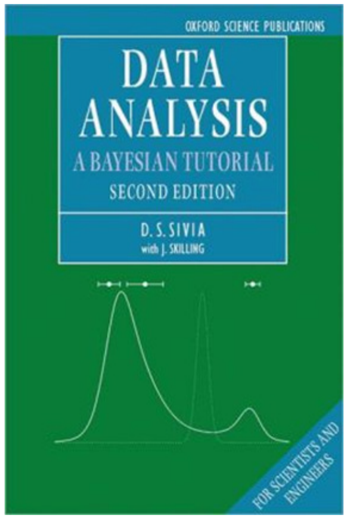
UNIVERSITAT POLITÈCNICA DE CATALUNYA

FABADA:
a Fitting Algorithm for Bayesian Analysis of DAta



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

ISIS 2/07/2013

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- Some examples
  - ✓ Quantitative guide to the eye
  - ✓ Analysis of QENS data
  - ✓ Intramolecular structure determination
- Robustness and simulated annealing
- Summary and conclusions

PRL 104, 060201 (2010) PHYSICAL REVIEW LETTERS week ending 12 FEBRUARY 2010

Why are Nonlinear Fits to Data so Challenging?

Mark K. Transtrum,* Benjamin B. Machta,† and James P. Sethna‡
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA
(Received 22 September 2009; published 10 February 2010)

In other words

- Why do they get stuck?
- Why do software thinks that it has finished to fit?
- How sure am I about the results?

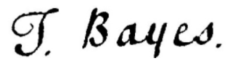
Why need a tool to efficiently explore the χ^2 landscape?

Bayes theorem

Likelihood
Probability that our data describes the hypothesis

Prior
Our forehand knowledge

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



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

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

T. Bayes.

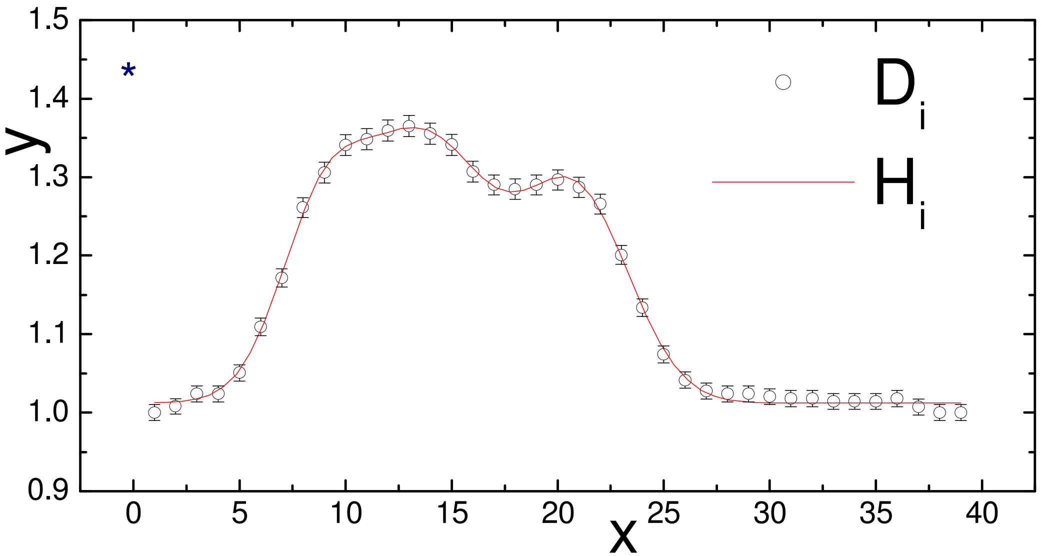
Posterior
Probability that the hypothesis is true given the experimental data

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)



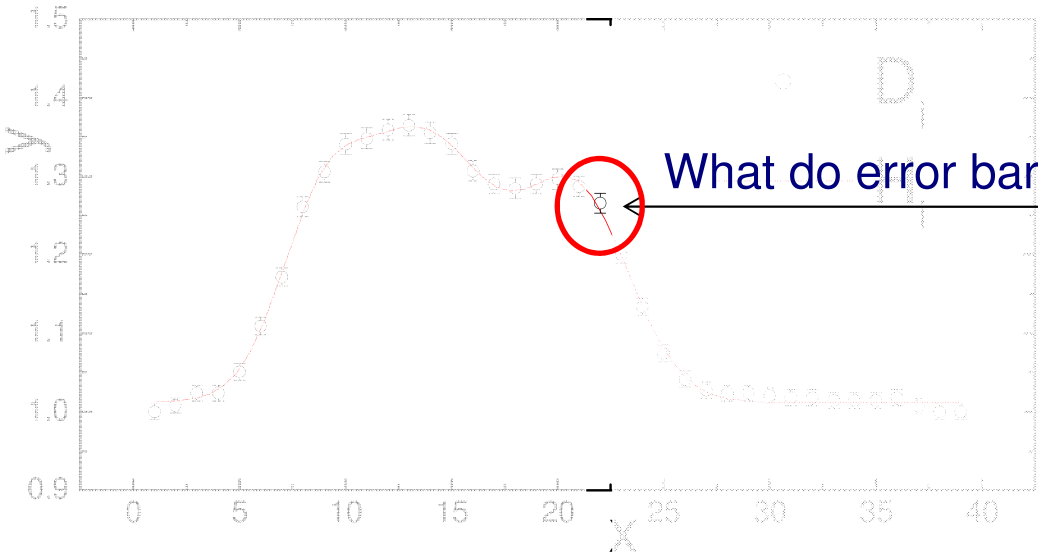
$$P(H_i\{P_l\} \mid D_i) \propto P(D_i \mid H_i\{P_l\})$$

* 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)

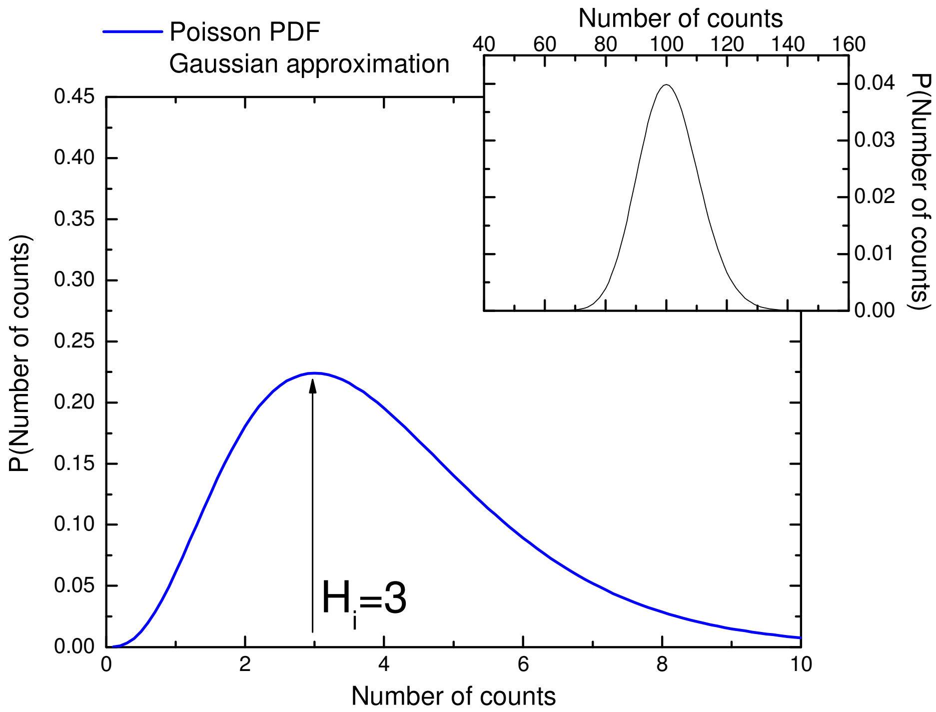


$$P(H_i\{P_l\} \mid D_i) \propto P(D_i \mid H_i\{P_l\})$$

The ubiquitous χ^2

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

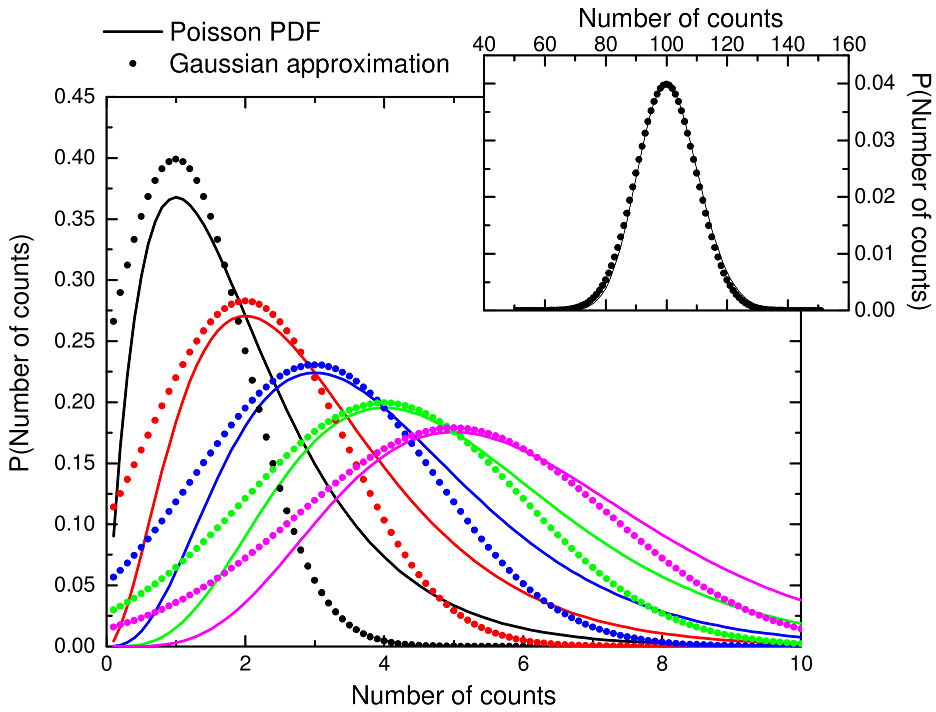
$$P(D_{i=k} \mid H_{i=k}) \propto H_i^{Di=k} e^{-Hi=k}$$



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:

$$\boxed{\chi^2 \propto -2 \cdot \ln L}$$

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)

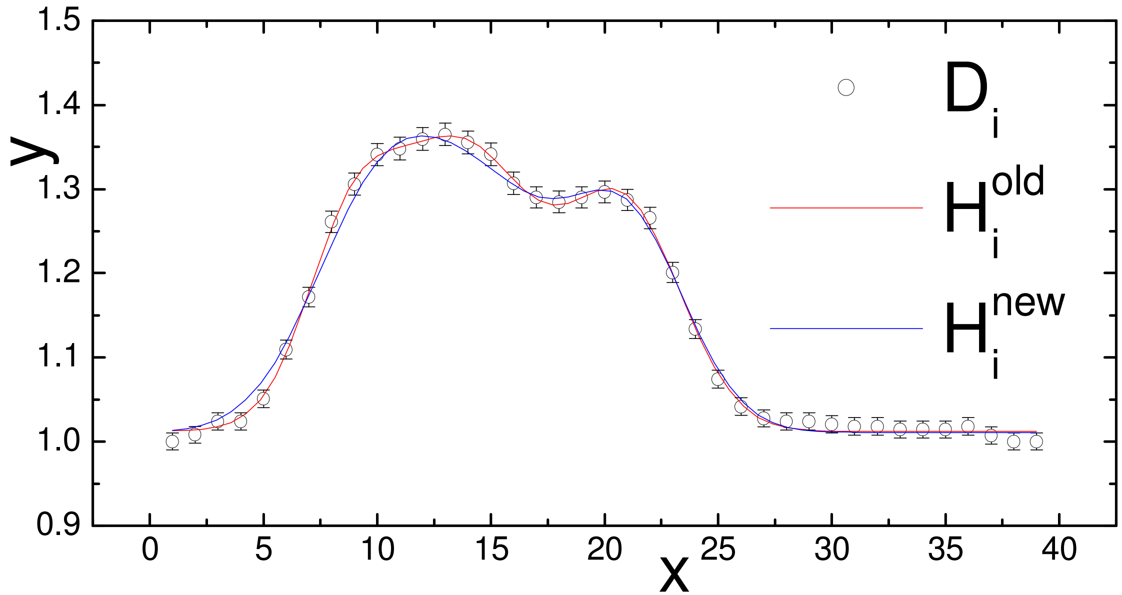
$$\frac{P\left(H_i\{P_l^{new}\}\mid D_i\right)}{P\left(H_i\{P_l^{old}\}\mid D_i)\right)} = \exp - \frac{\sum_{i=1}^{n}\left(H_i^{new} - D_i\right)^2 - \sum_{i=1}^{n}\left(H_i^{old} - D_i\right)^2}{2\sigma^2} = \exp - \frac{\left(\chi_{new}^2 - \chi_{old}^2\right)}{2}$$

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

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



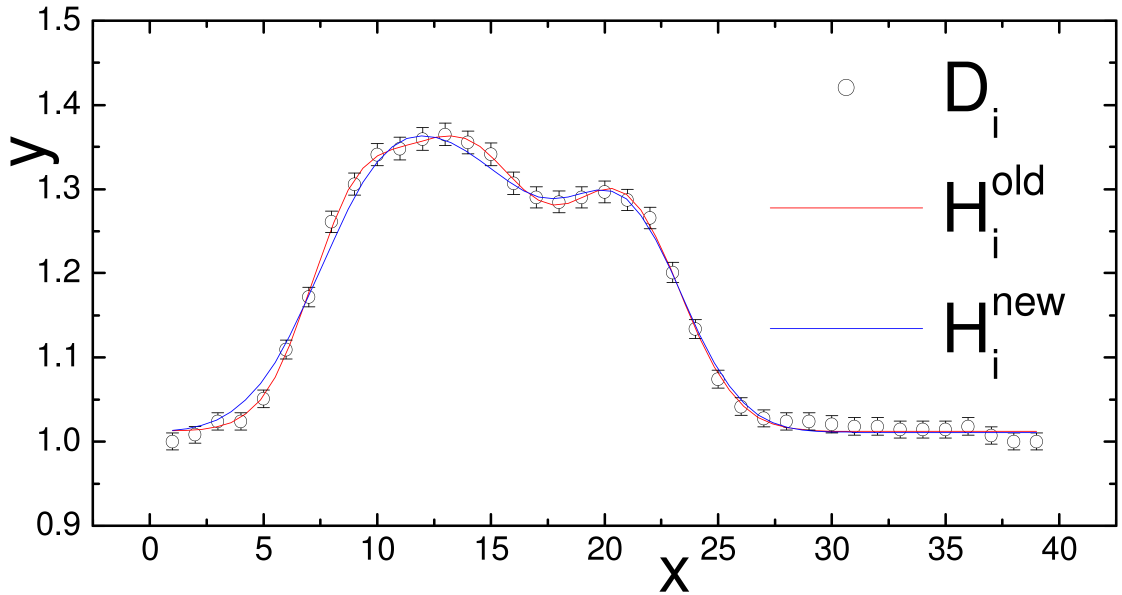
Wandering in parameter space χ^2

What does FABADA do?

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

step	*a1*	*c1*	*sigma1*	*chisq*
100	*9.806162E-01*	*5.169688E+01*	*5.290837E+00*	*4.267798E+03*
200	*9.999395E-01*	*5.180603E+01*	*5.411015E+00*	*4.265669E+03*
300	*1.015084E+00*	*5.172740E+01*	*5.411015E+00*	*4.267409E+03*
400	*9.944127E-01*	*5.171415E+01*	*5.395535E+00*	*4.265770E+03*
500	*1.005000E+00*	*5.180460E+01*	*5.440930E+00*	*4.266081E+03*

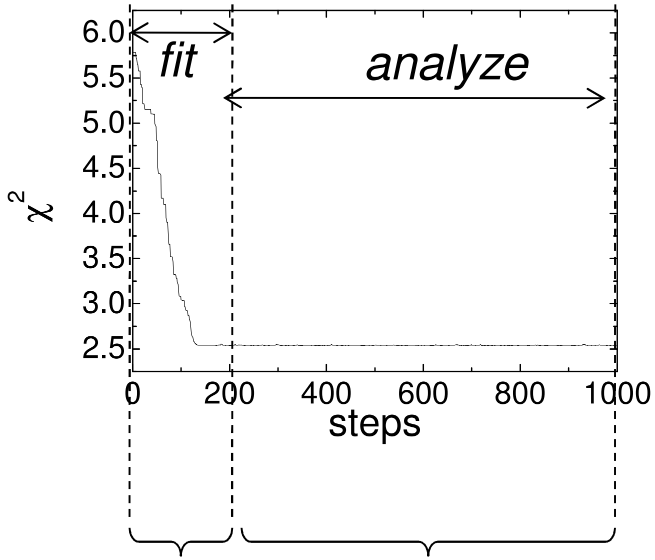
.
.
.



It "simply" generates a Markov Chain with parameters and χ^2 supported by the data

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- Some examples
  - Quantitative guide to the eye
  - Analysis of QENS data
  - Intramolecular structure determination
- Robustness and simulated annealing
- Summary and conclusions

Three main advantadges

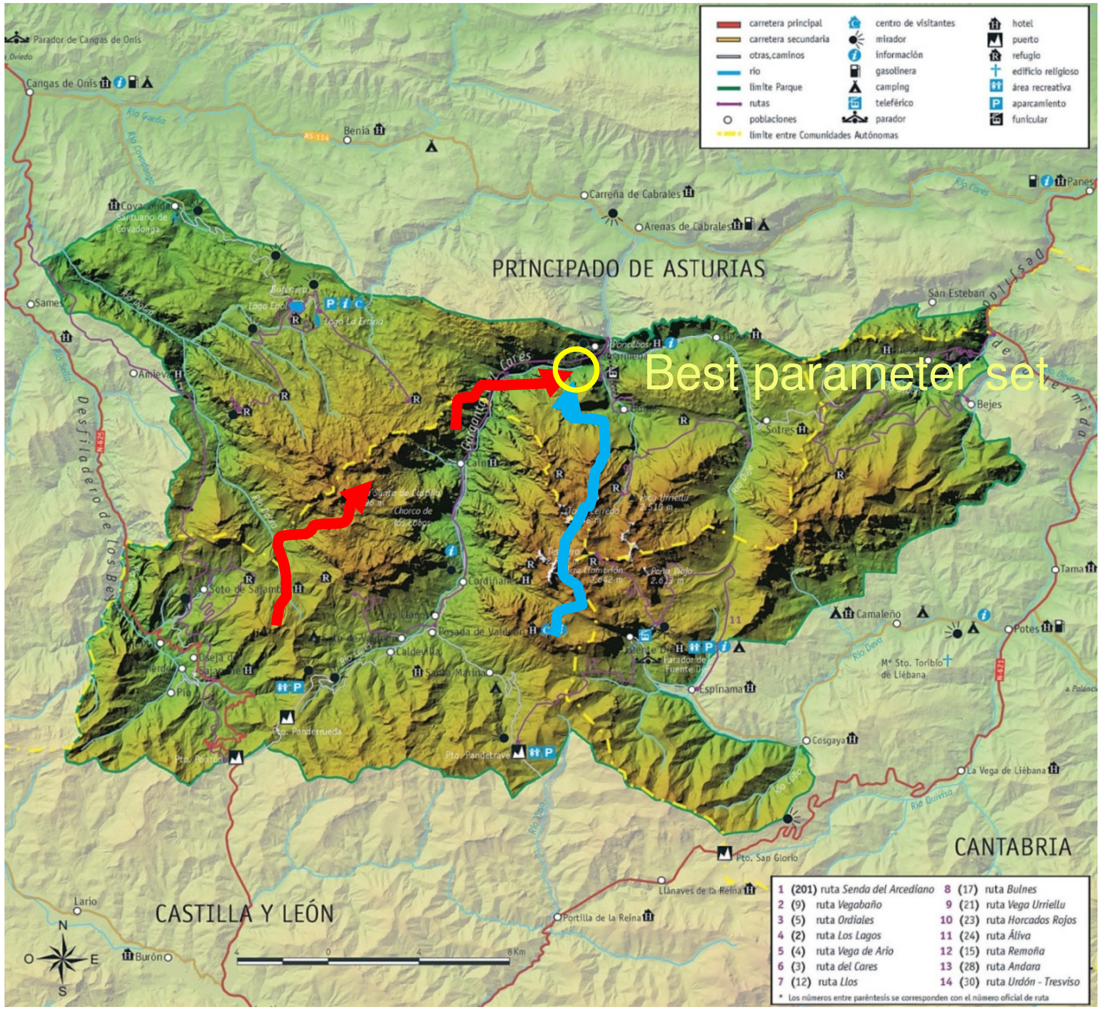


- The fitting process

- Parameter estimation
- Model selection

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

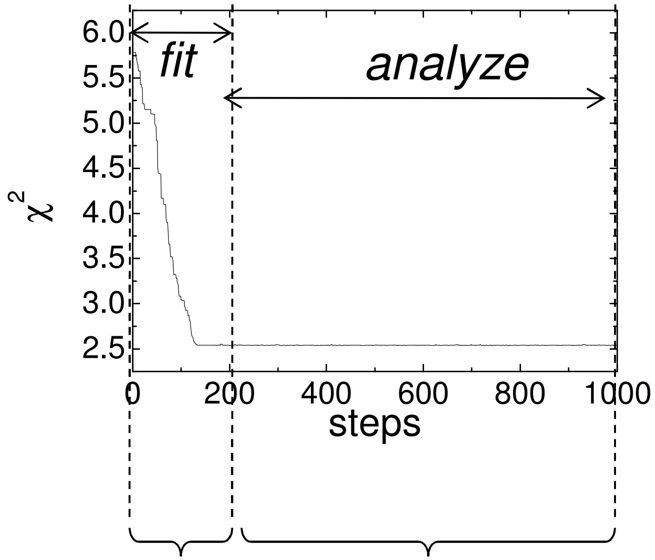
Classical fiting



Bayessian fiting

It does not get stuck!!!

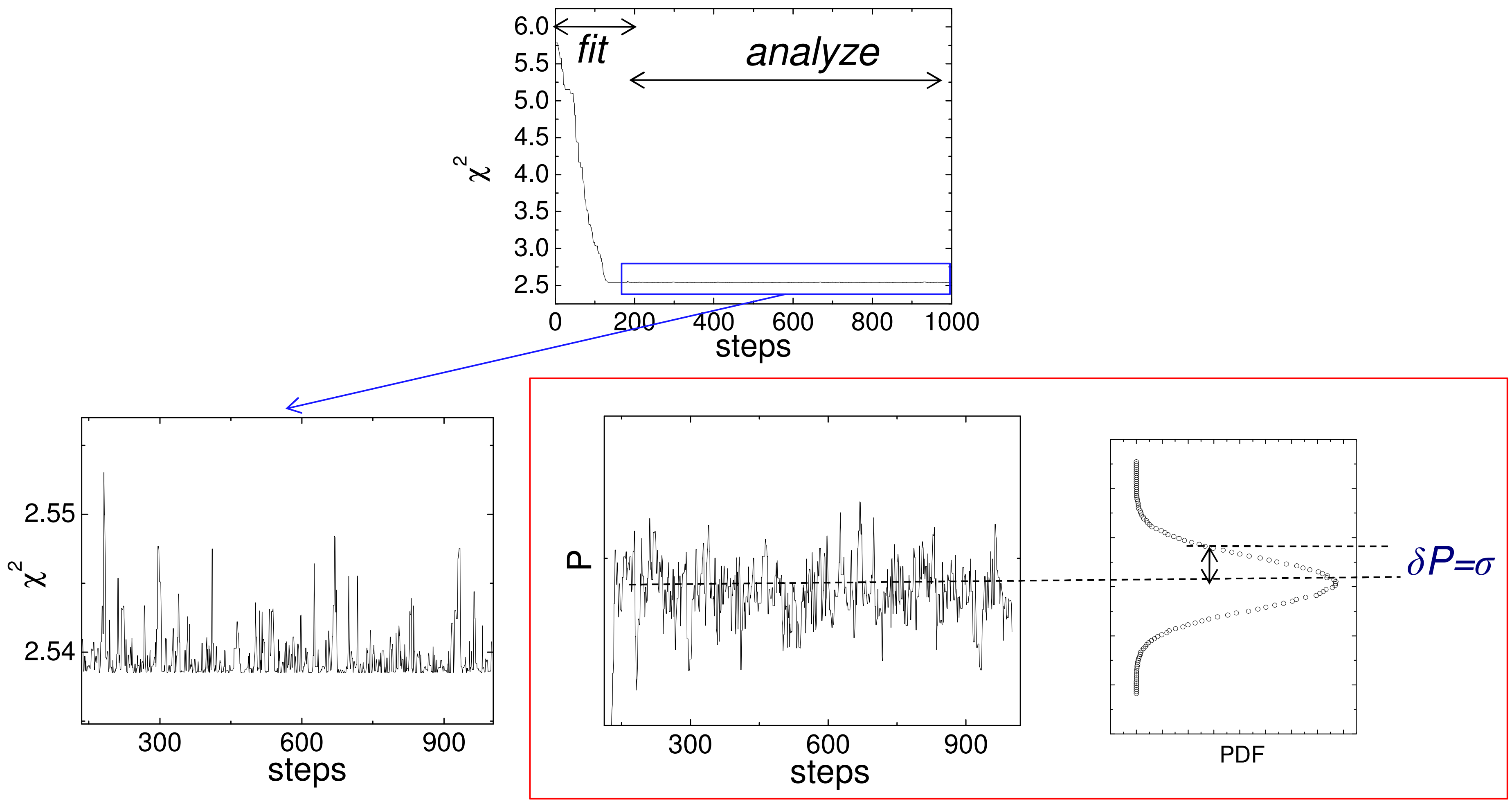
Three main advantadges



- The fitting process

- Parameter estimation
- Model selection

Parameter estimation



Parameter estimation:
Parameters are obtained as PDF's not as $P \pm \delta P$

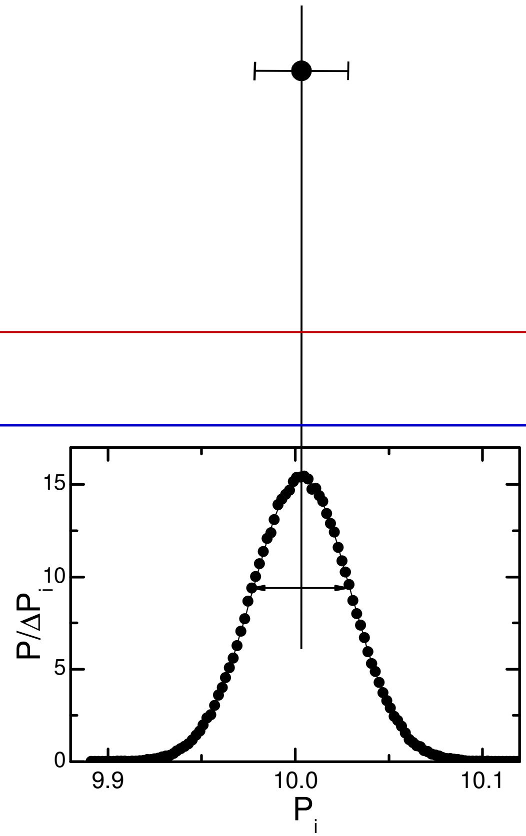
Parameter determination

"Classic"

(frequentist)

$$P \pm \delta P$$

Bayesian

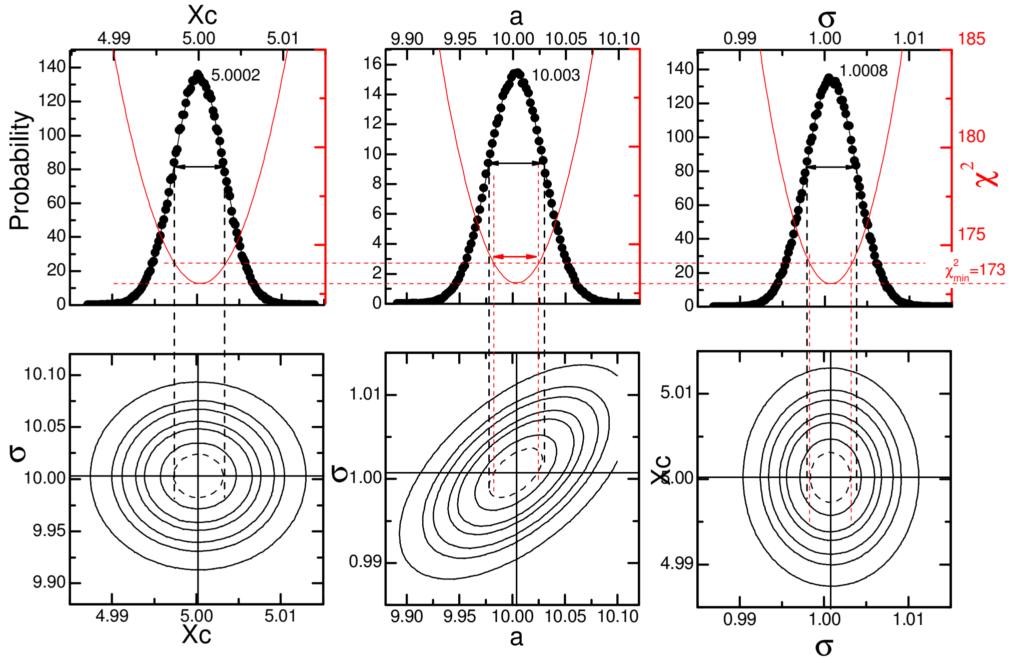


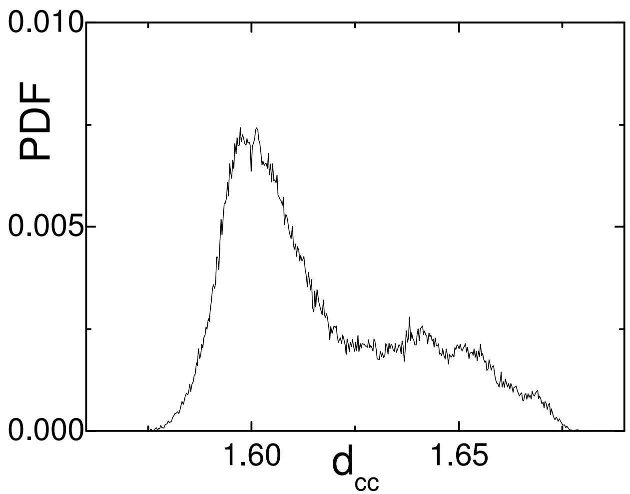
Probability Density Function
(PDF)

Parameter determination

Correlation between parameters are automatically had into account

No supposition on the minimum geometry is made





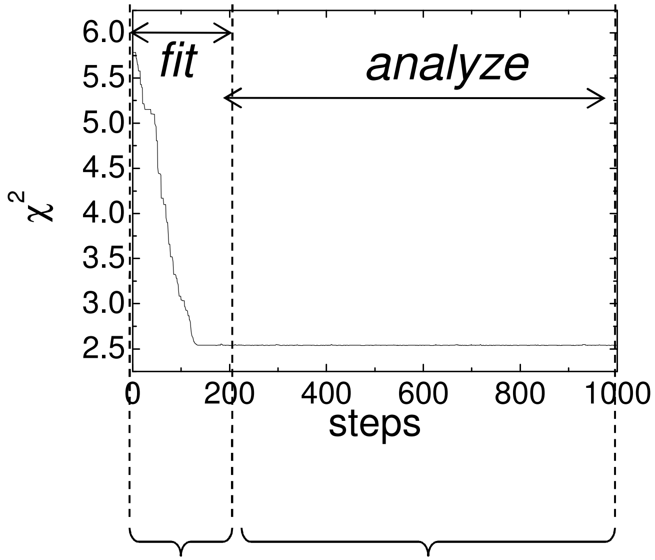
Fit of a "simple" Gaussian

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

It might be as terrible as this one…

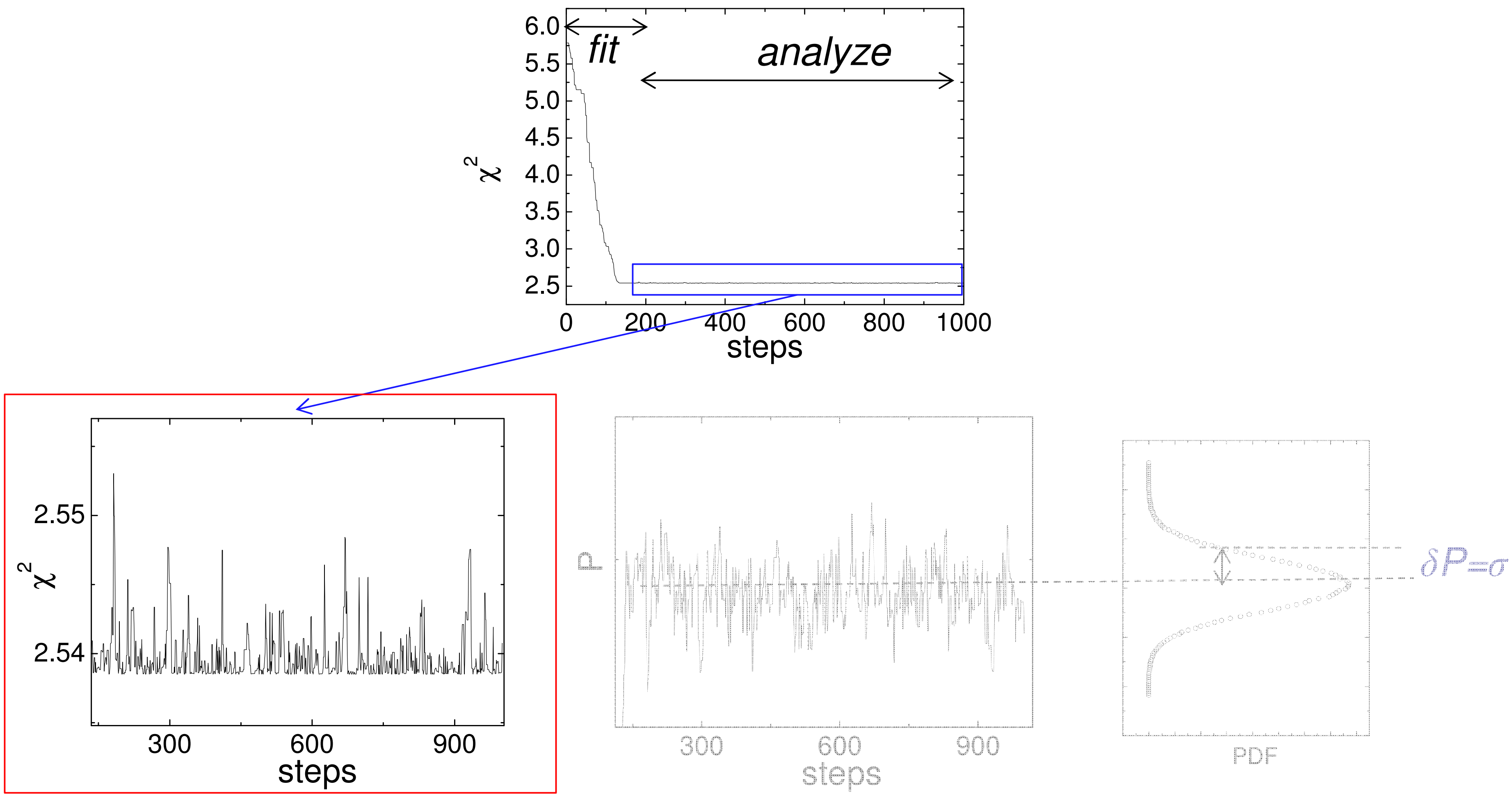
J. Phys.: Conf. Ser. 325, 012006 (2011)

Three main advantadges



- The fitting process
- Parameter estimation
- Model selection

The fitting process



Model selection:
all possible combinations of parameters are investigated and their χ^2 calculated

Model Selection

Usual methods

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

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

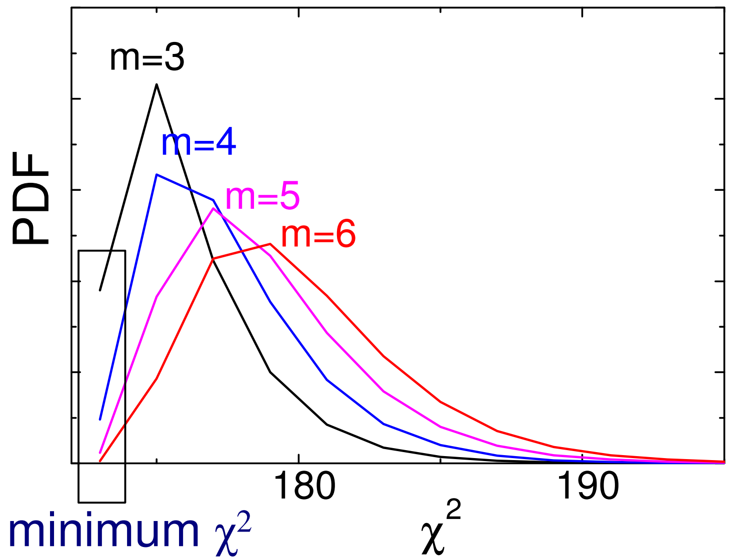
n: is the number of points
m: is the number of parameters

This works only if:

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

Bayesian method

- Directly compares the PDF related to χ^2

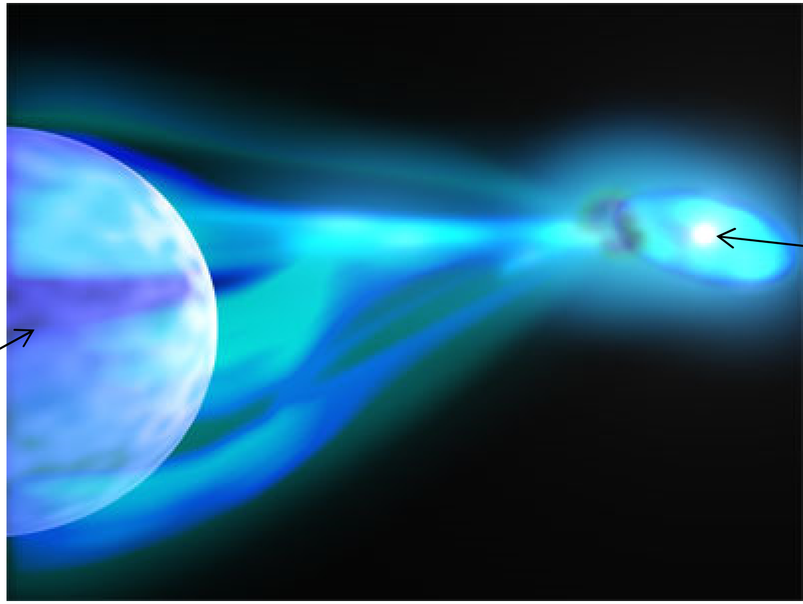


an increasing number of parameters broadens the χ^2 PDF

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- Some examples
  - ✓ Quantitative guide to the eye
  - ✓ Model selection using QENS data
  - ✓ Intramolecular structure determination
- Summary and conclusions

A far away example…

Main Sequence star



Neutron star
degenerate matter

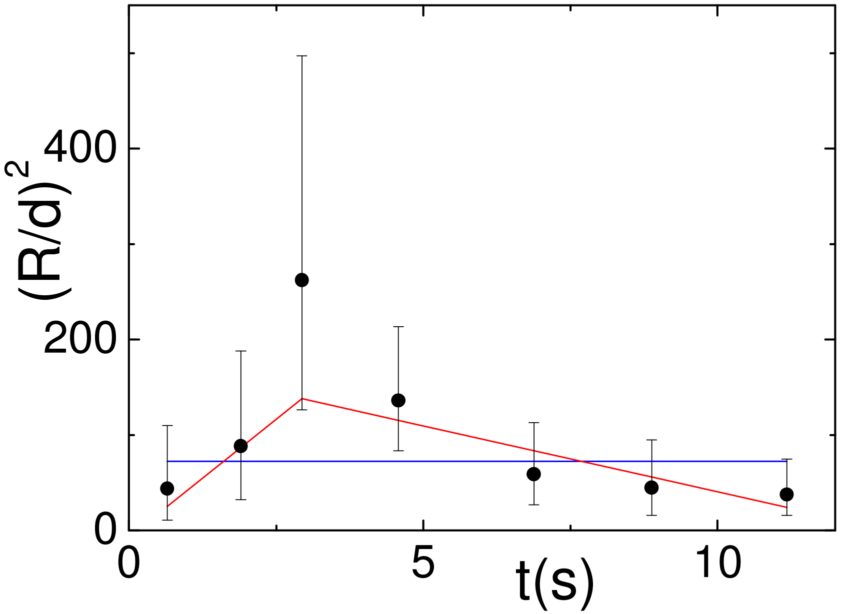
Two models: does the neutron star envelop expand?

YES: radiation pressure"=" escape velocity
you know the mass! (you have a paper!)

NO: you don't know the mass! *(Game over!)*

G. Sala et al. The Astrophysical Journal (2012)

A far away example…



Two models: does the neutron star envelop expand?

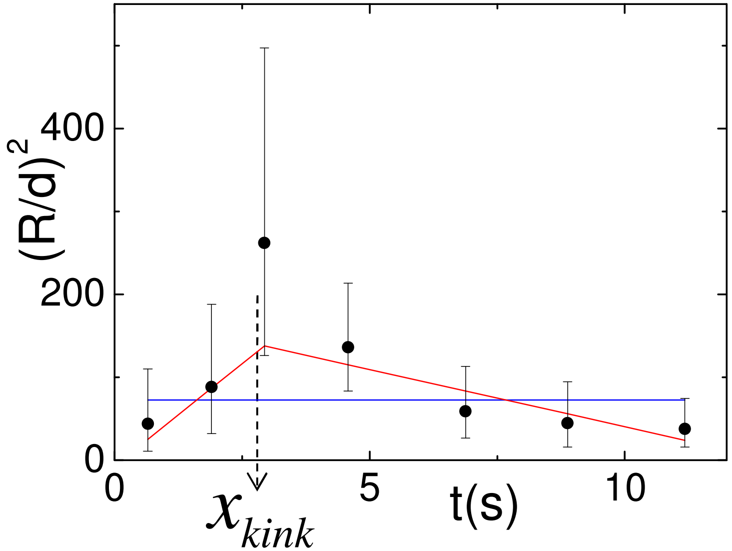
YES: Eddington luminosity: radiation pressure"=" escape velocity
you know the mass! (you have a paper!)

NO: you don't know the mass! *(Game over!)*

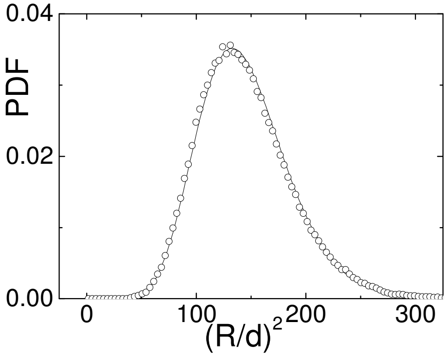
G. Sala et al. The Astrophysical Journal (2012)

A Bayesian Dream...

(Levenberg-Marquart does not work)



Poisson statistics



$$\chi^2 = -2\sum_{i=1}^{n}\left[\left(D_i \ln\frac{H_i}{D_i} - H_i + D_i\right)\right]$$

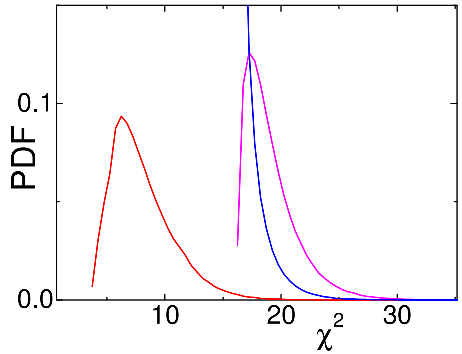
Non differentiable model

$$y_1 = a_1 + b_1 x$$

$$a_2 = a_1 + (m_1 - m_2)\, x_{kink}$$

$$y_2 = a_2 + b_2 x$$

Model Selection



YES!!

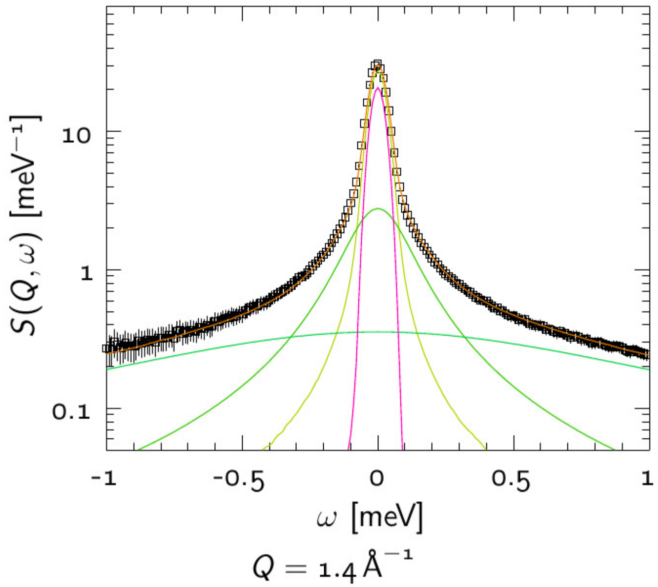
you know the mass

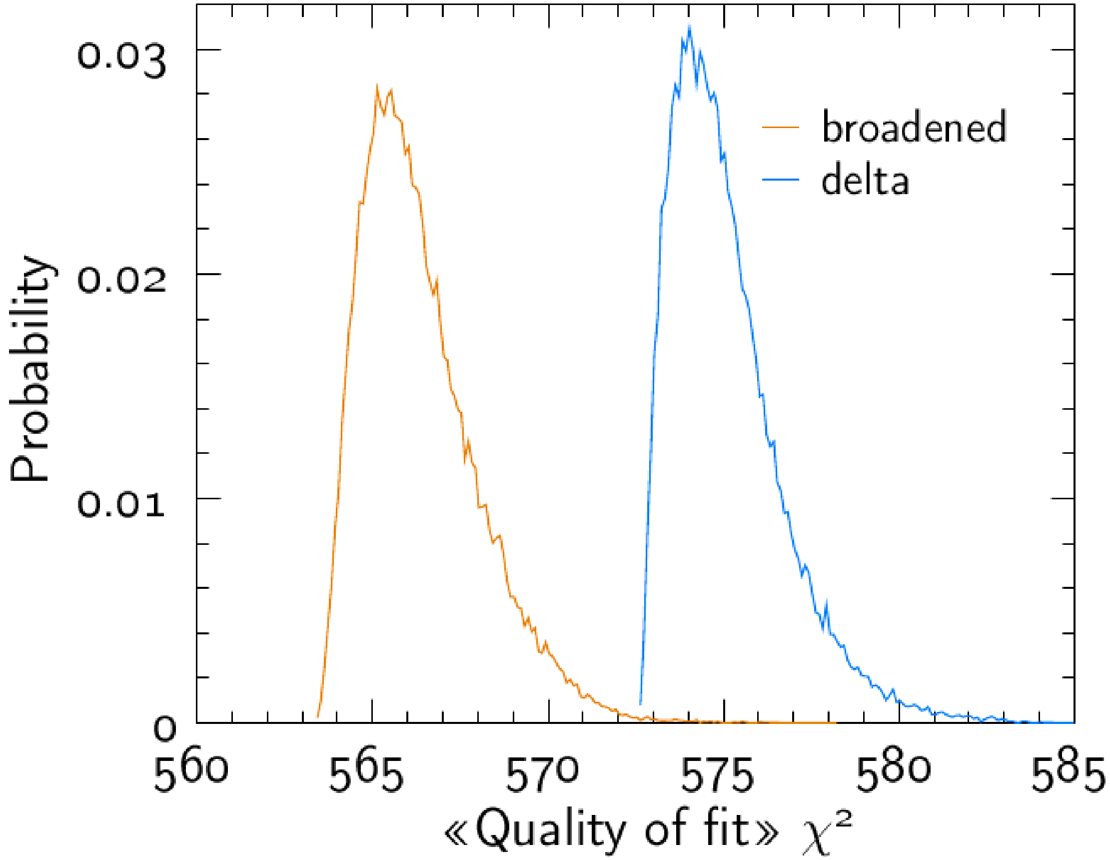
G. Sala et al. The Astrophysical Journal (2012)

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- Some examples
  - Quantitative guide to the eye
  - Analysis of QENS data
  - Intramolecular structure determination
- Summary and conclusions

Motion of phospholipids in the membrane

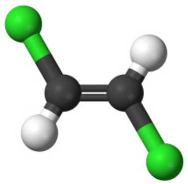
Is there a broadening? Delta model versus Broadened model





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

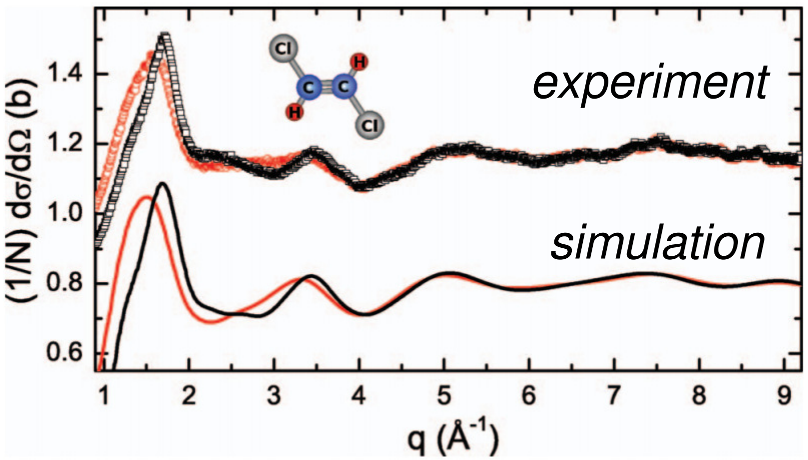
Analysis of QENS spectra

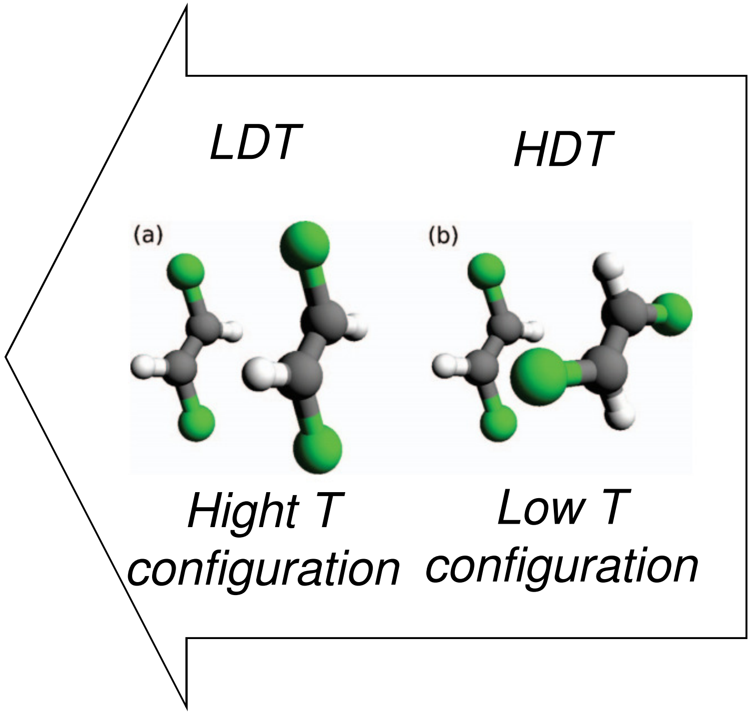


Is there a liquid-liquid phase transition in trans-dicloroethylene?

- Reported changes in density
- Reported changes in dynamics (NMR, infrarred spc, viscosity)

There is a change in the structure





Is there any change in the dynamics?

M. Rovira-Esteva et al. J. Chem. Phys. 136, 124514 (2012)
M. Rovira-Esteva et al. Phys. Rev. B 81(9) 092202 (2010)

Is there any change in the dynamics?

Dynamics transition in trans-dicloroethylene

We analyze TOF spectra with a diffussion+rotation model

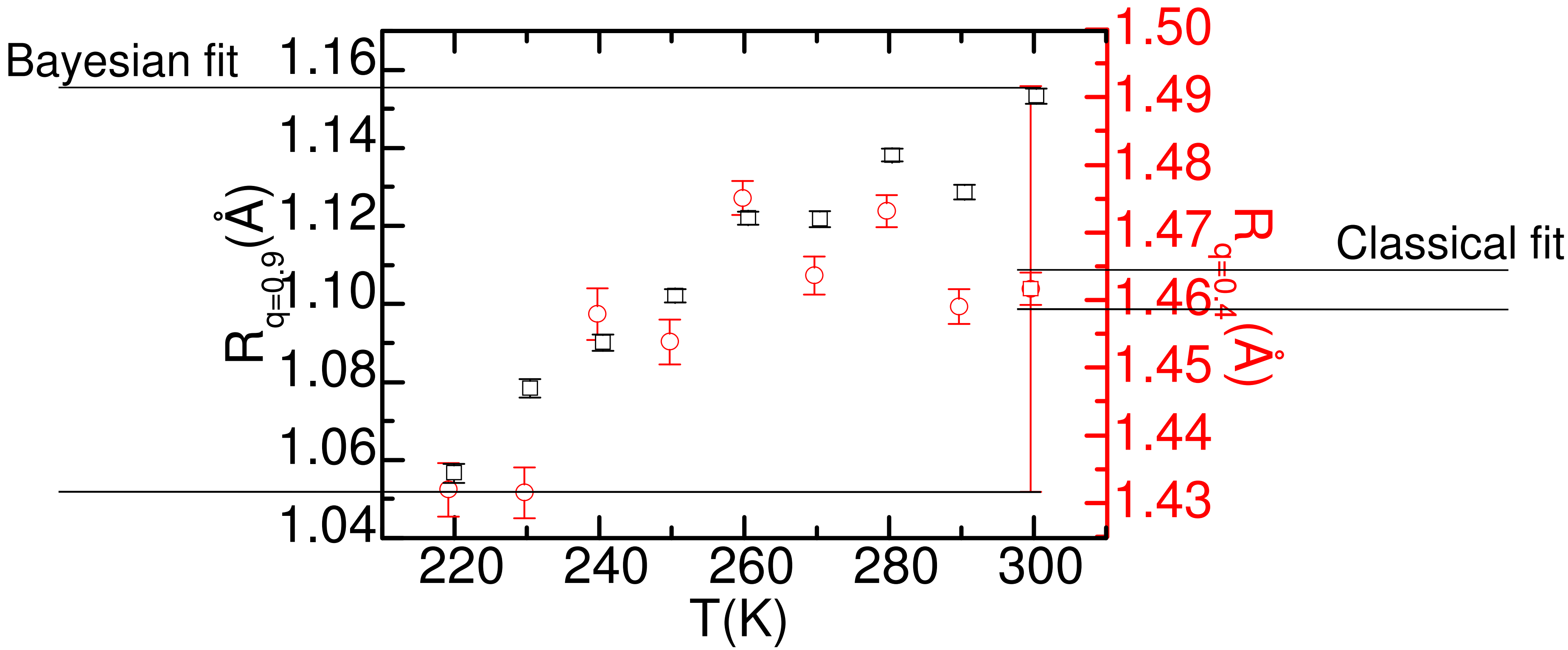
$$S_{rot}(q,\omega) = A_0(qR)\cdot\delta(\omega) + \sum_l A_l(qR)\cdot L_l(w, \gamma = l(l+1)D_r)$$

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

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

M. Rovira-Esteva et al. J. Chem. Phys. 136, 124514 (2012)
M. Rovira-Esteva et al. Phys. Rev. B 81(9) 092202 (2010)

We analyze just one q value

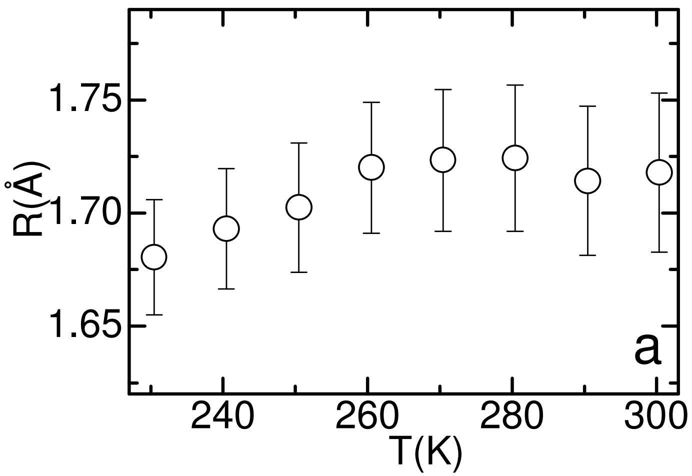


M. Rovira-Esteva et al. J. Chem. Phys. 136, 124514 (2012)
M. Rovira-Esteva et al. Phys. Rev. B 81(9) 092202 (2010)

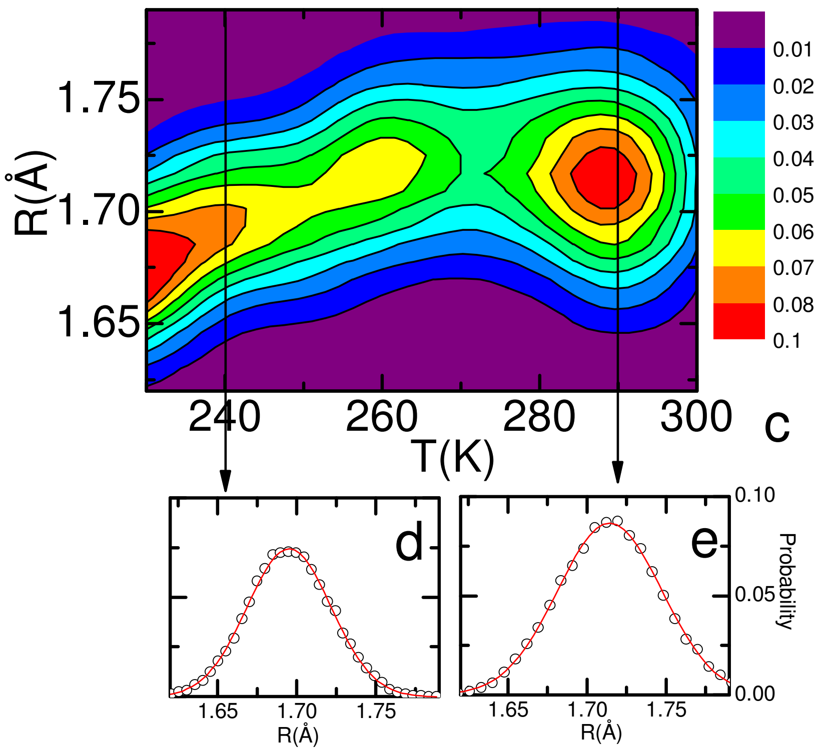
All q values at once!!

i.e. the whole scattering law

Classical representation



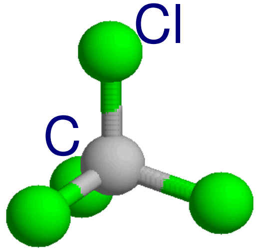
PDF representation



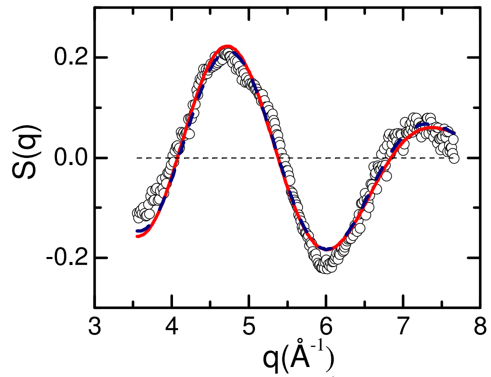
M. Rovira-Esteva et al. J. Chem. Phys. 136, 124514 (2012)
M. Rovira-Esteva et al. Phys. Rev. B 81(9) 092202 (2010)

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- Some examples
  - Quantitative guide to the eye
  - Model selection using QENS data
  - Intramolecular structure determination
- Summary and conclusions

A simple case: CCl_4



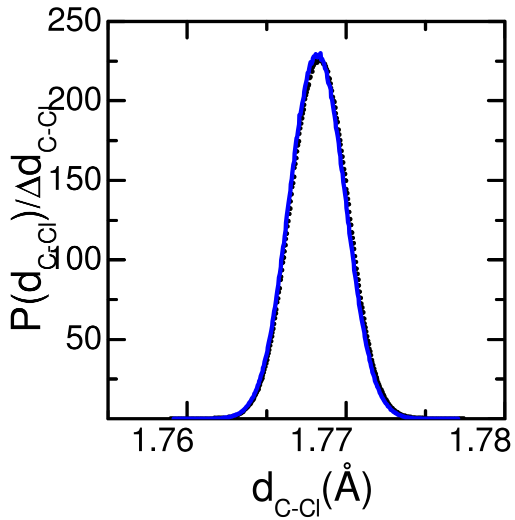
Molecular structure
fit of the high q region of s(q)



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

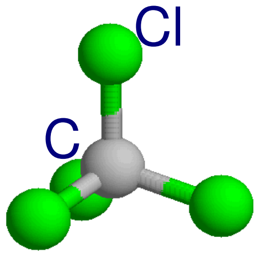
---

d_{CCl} distance is well defined

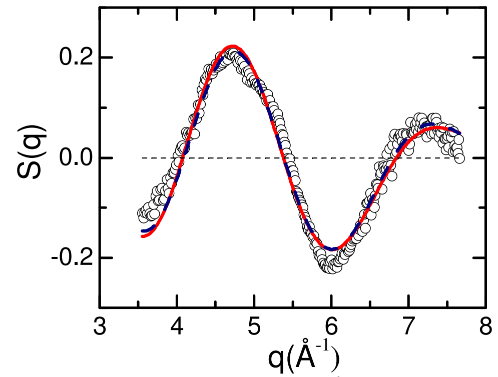


Pardo L.C. Et al. arXiv:0907.3711v3 [physics.data-an]

A simple case: CCl_4



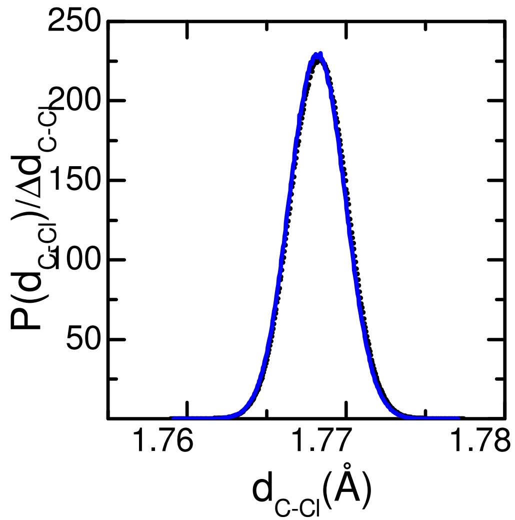
Molecular structure
fit of the high q region of s(q)



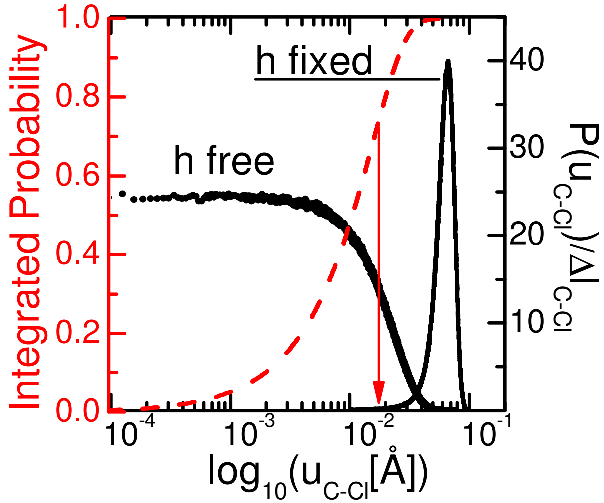
$$S(q) = h \cdot \sum_{i,j}^{m} b_i b_j \cdot \frac{\sin(qr_{ij})}{qr_{ij}} \cdot e^{-u_{ij}^2 q^2 / 2}$$

---

d_{CCl} distance is well defined

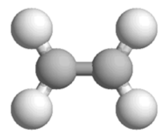


U_{C-Cl} depends on the scaling factor h



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)

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

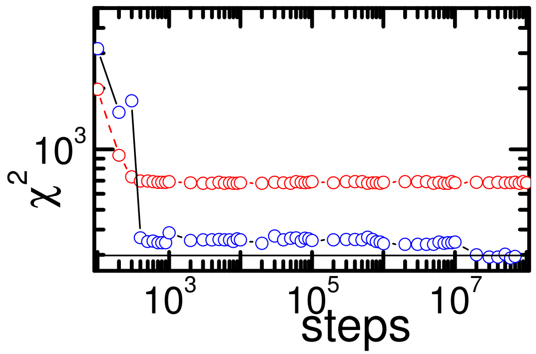
fit of the small r region of G(r)

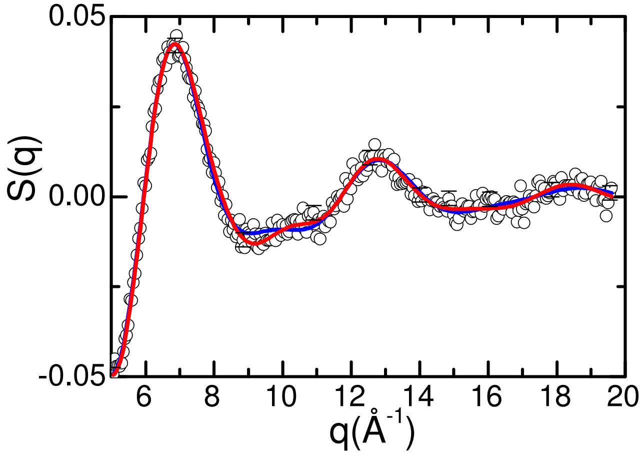
$$G(r) \approx h \cdot \sum_{i,j}^{m} b_i b_j \cdot \exp\left(-\frac{(r - r_{ij})^2}{2u_{ij}^2} \right)$$

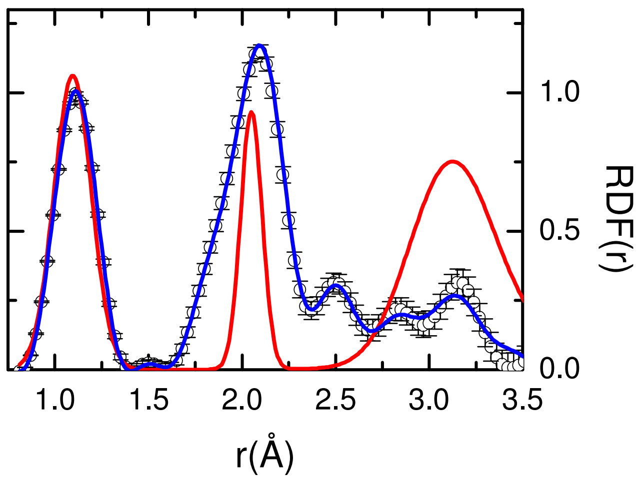
$\longleftrightarrow$

Fit using only s(q)

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







Why is it better fitting both at the same time?

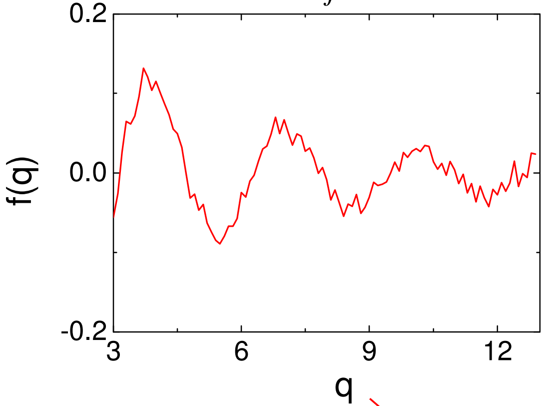
M. Rovira-Esteva et al. Phys. Rev. B 84, 064202 (2011)

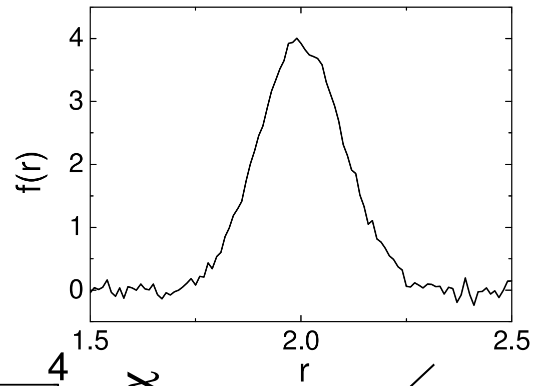
Why is it better fitting both at the same time?

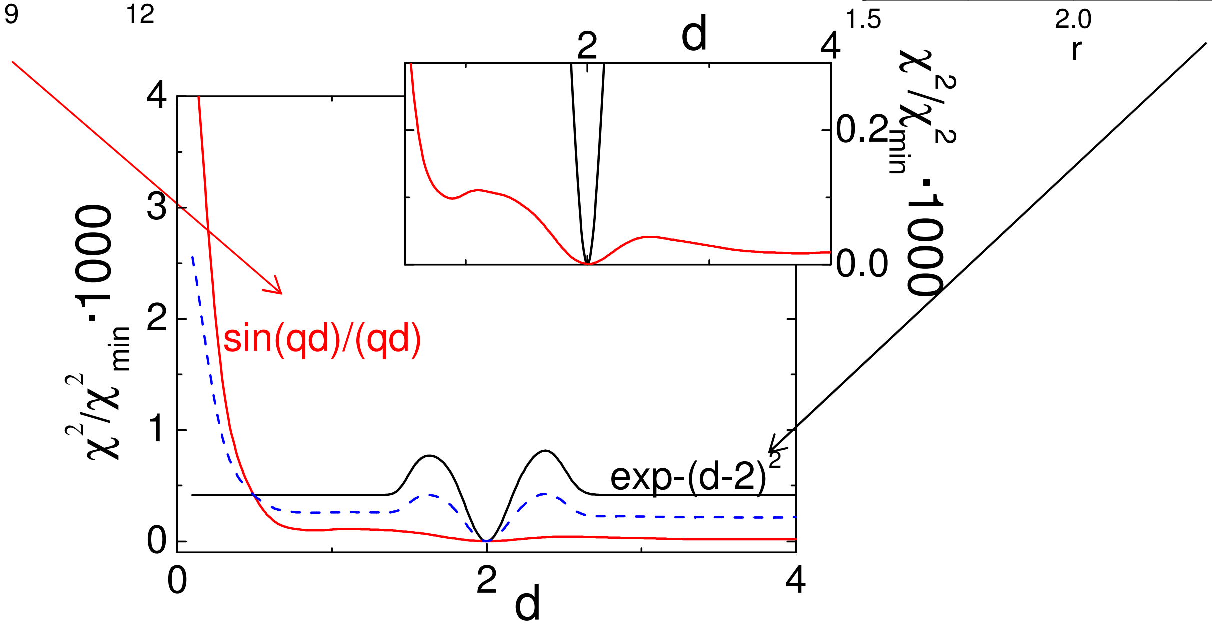
The simplest case (with U_{ij}=0.1, r_{ij}=2):

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

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



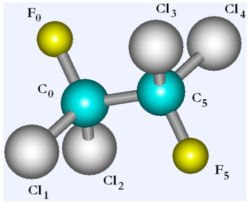




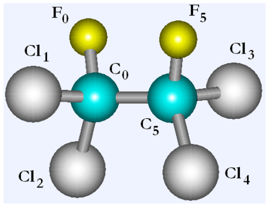
χ^2 landscapes are quite different!!

A really complicated case: $C_2Cl_4F_2$

It has two conformers



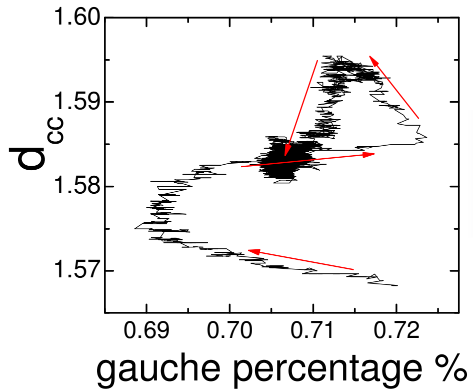
trans

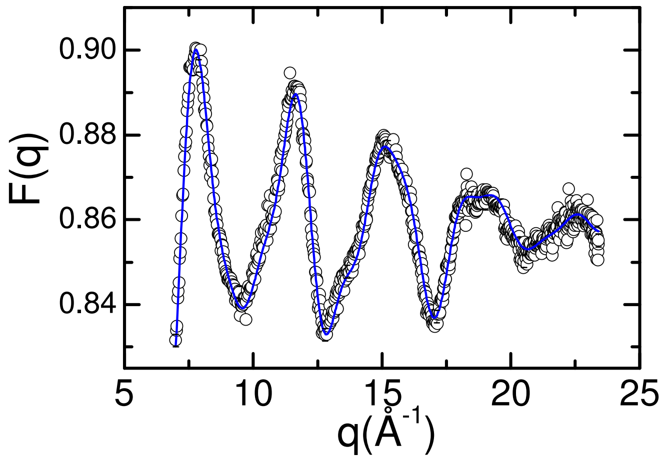


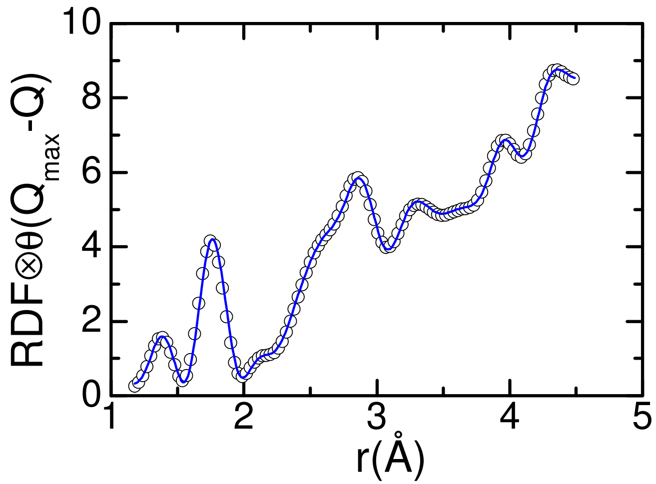
gauche

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)

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- Some examples
  - Quantitative guide to the eye
  - Analysis of QENS data
  - Intramolecular structure determination
- Robustness and simulated annealing
- Summary and conclusions

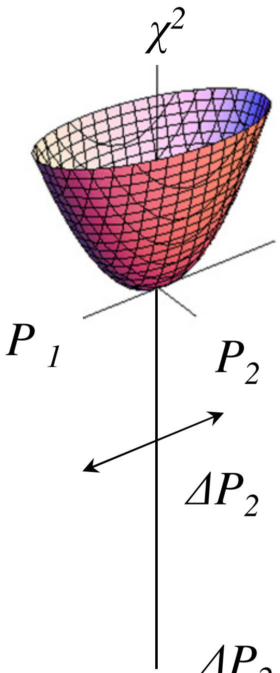
Adaptive Markov Chain Montecarlo

$$\Delta P_{new} = \Delta P_{old} \frac{R_i}{R_{i,desired}}$$

step	a1	c1	sigma1	chisq
100	9.806162E-01	5.169688E+01	5.290837E+00	4.267798E+03
200	9.999395E-01	5.180603E+01	5.411015E+00	4.265669E+03

ΔP is optimized!

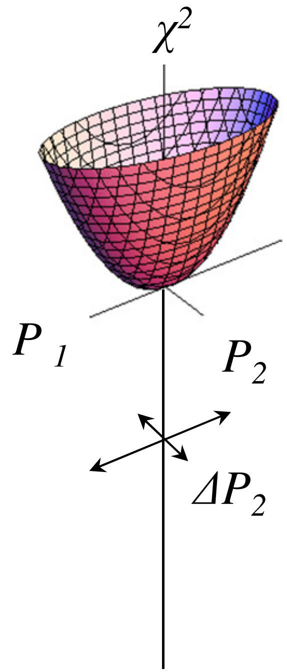
To optimize de number of accepted moves



ΔP_2 too mall: always accepted
but no moving

ΔP_2 too big: moving a lot
but not being accepted

To explore the χ^2 landscape
in all directions



$\Delta P_2 < \Delta P_1$

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

Adaptive Markov Chain Montecarlo

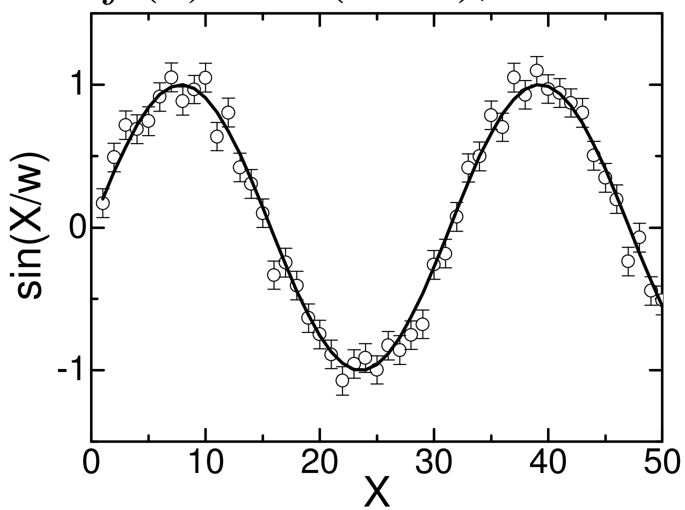
$$\Delta P_{new} = \Delta P_{old} \frac{R_i}{R_{i,desired}}$$

+ Simulated annealing...

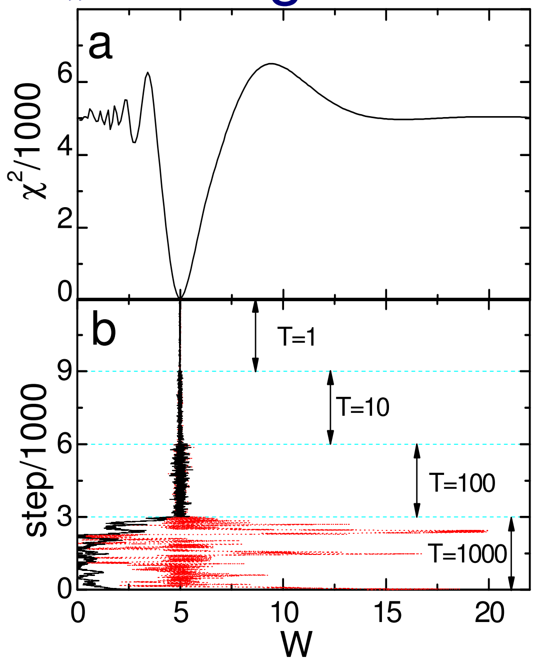
$$\frac{P\left(H_i\{P_l^{new}\} \mid D_i\right)}{P\left(H_i\{P_l^{old}\} \mid D_i\right)} = \exp - \frac{\left(\chi^2_{new} - \chi^2_{old}\right)}{2T}$$

= a test case

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

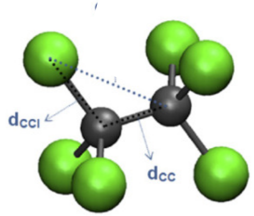


„Cooling the fit“

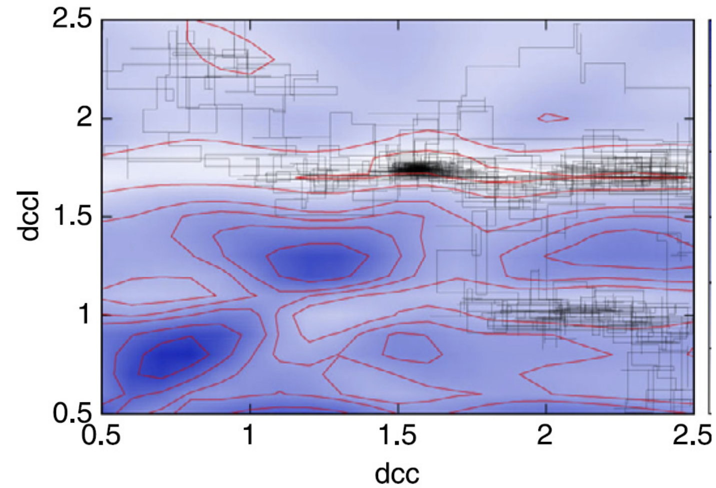


L.C. Pardo et al. Phys. Rev. E. (2011)

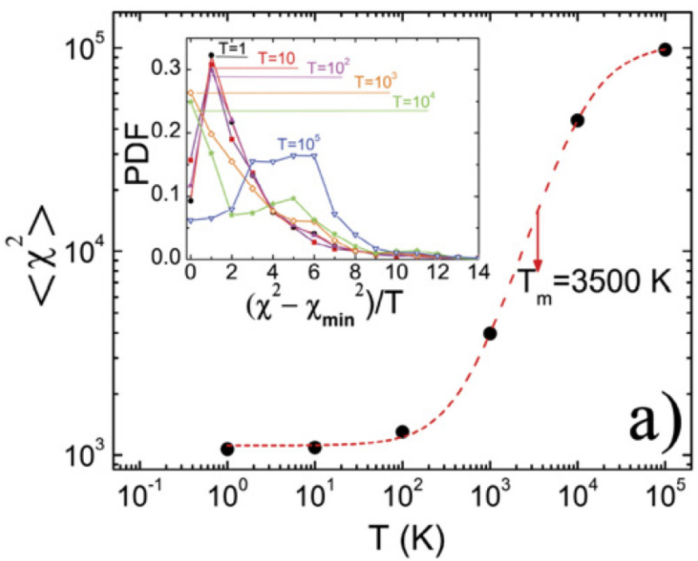
The structure of C_2Cl_6

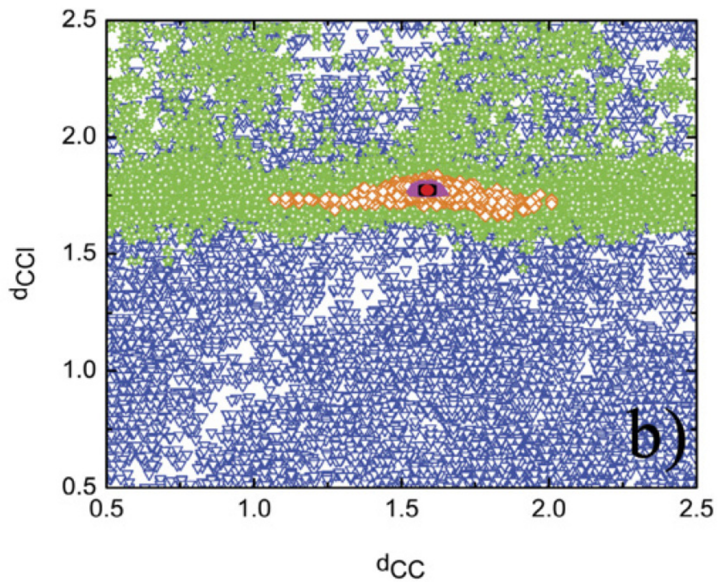


χ^2(dcc,dccl) landscape



„Melting“ the fit





Aristizabal A. H. et al. J. Phys.: cond. Matt. RMC special issue

Just a word about priors:

From some paper (that I trust) I know $P \pm \delta P$

Can I include this information in the fit?

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

$$P(H \mid D) \propto \exp - \frac{\chi^2}{2} \cdot \boxed{\exp - \frac{(P - P_i)^2}{2\delta P^2}}$$

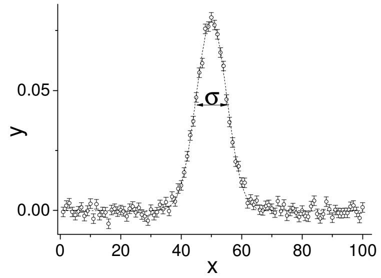
$$\boxed{\chi^2_{+prior} = \sum_{i=1}^{n} \frac{(H_i - D_i)^2}{\sigma_i^2} + \sum_{i=1}^{n_p} \frac{(P - P_i)^2}{\delta P^2}}$$

Aristizabal A. H. et al. J. Phys.: cond. Matt. RMC special issue

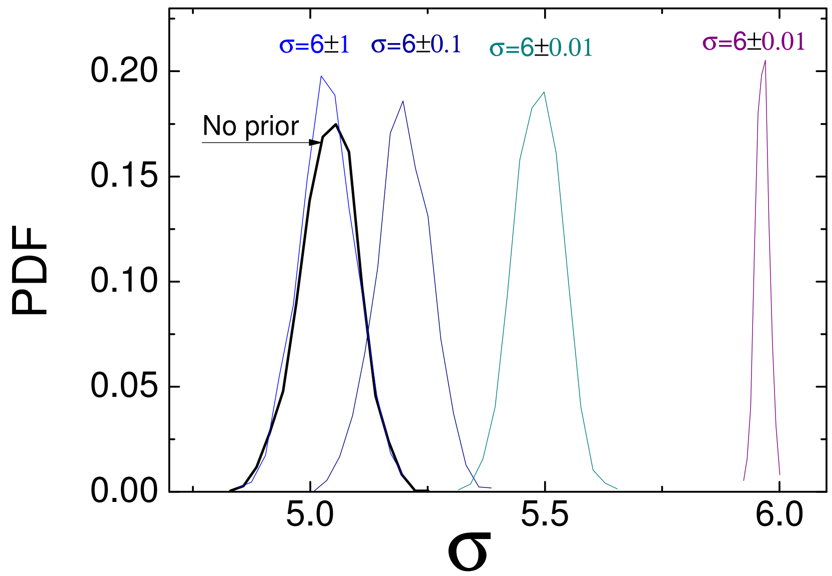
Just a word about priors:

From some paper (that I trust) I know $P \pm \delta P$

Can I include this information in the fit?



Fitting a gaussian with σ=5, and P=6±δP



Aristizabal A. H. et al. J. Phys.: cond. Matt. RMC special issue

- The ubiquitous χ^2
- Advantages of Bayesian analysis
- Some examples
  - Quantitative guide to the eye
  - Analysis of QENS data
  - Intramolecular structure determination
- Robustness and simulated annealing
- Summary and conclusions

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 Density 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

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

UNIVERSITAT POLITÈCNICA DE CATALUNYA

acknowledgements

Muriel Rovira-Esteva (GCM-UPC)

Alessandro Vispa (GCM-UPC)

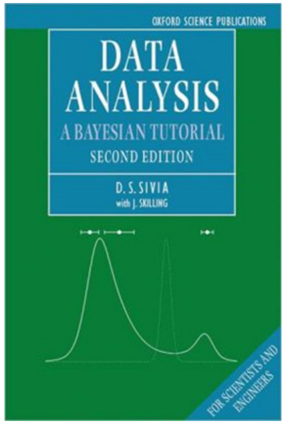
Andres Henao (GCM / UPC)

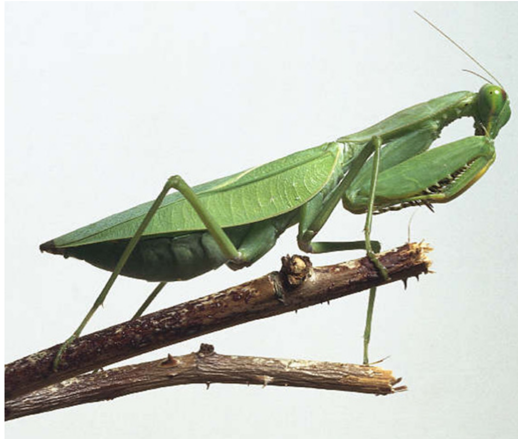
Gabriel Cuello (ILL / D4)

Sebastian Busch (University of Oxford)

Glòria Sala Cladellas (GAA-UPC)

Thank you for your attention





To read more about the examples

Astrophysics
G. Sala et al. The Astrophysical Journal (2012)

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

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

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

Dielectric spectroscopy
J. C. Martínez et al. J. Phys. Chem. B 114 6099 (2010)

To read more about FABADA and download it

Aristizabal A. H. et al. Journal of Physics: condensed matter. Special issue: Reverse Monte Carlo 2012

L.C. Pardo et al. Phys. Rev. E. 84, 046711 (2011)

L.C Pardo et al. J. Phys.: Conf. Ser. 325, 012006 (2011)

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