

# A new bootstrap-based fitting method

**Paolo Pedroni**

INFN-Sezione di Pavia, Italy

**In collaboration with B. Pasquini and S. Sconfietti - University and INFN - Pavia**



Uncertainty Quantification  
(UQ) at the Extremes  
(ISNET-6)

8-12 October 2018  
TU Darmstadt

# SUMMARY

## ➤ Introduction

- ✓ The Physics Problem: Fit of experimental data on Real Compton Scattering off protons
- ✓ Definition of the physics parameters to be fitted: Static and Dynamical polarizabilities

## ➤ Fitting Procedure

- ✓ Problems in the existing data set
- ✓ New fit approach: simplex + bootstrap
- ✓ A cross check: static polarizabilities
- ✓ Dynamical polarizabilities
- ✓ A further check of the fitting procedure

## ➤ Conclusions

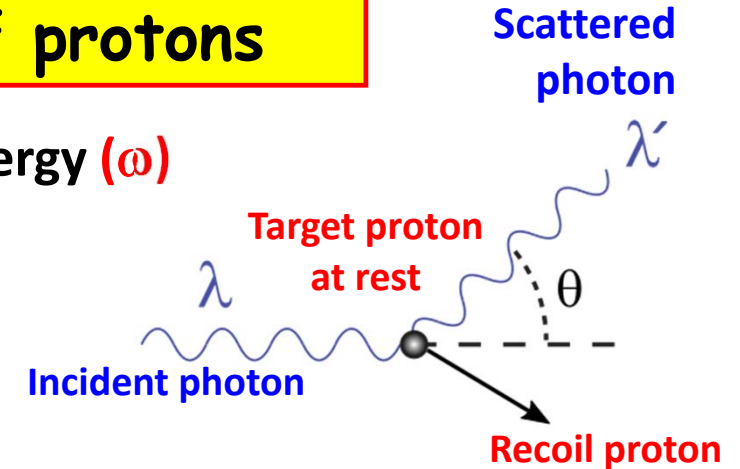
# Real Compton Scattering off protons

Expansion of the effective Hamiltonian in incident photon energy ( $\omega$ )

0th order  $\longrightarrow$  charge, mass  
 1st order  $\longrightarrow$  magnetic moment

} «point-like» nucleon  
 (Born terms)

2nd order  $\longrightarrow$  2 scalar polarizabilities      Baldin's sum rule:  $(\alpha_{E1} + \beta_{M1}) \sim \text{known}$



$$H_{eff}^{(2)} = -4\pi \left[ \frac{1}{2} \alpha_{E1} \vec{E}^2 + \frac{1}{2} \beta_{M1} \vec{H}^2 \right]$$

3rd order  $\longrightarrow$  4 spin (vector) polarizabilities      only one direct measurement      P. Martel et al, PRL 114, 112501 (2015)

$$H_{eff}^{(3)} = -4\pi \left[ \frac{1}{2} \gamma_{E1E1} \vec{\sigma} \cdot (\vec{E} \times \dot{\vec{E}}) + \frac{1}{2} \gamma_{M1M1} \vec{\sigma} \cdot (\vec{H} \times \dot{\vec{H}}) - \gamma_{M1E2} E_{ij} \sigma_i H_j + \gamma_{E1M2} H_{ij} \sigma_i E_j \right]$$

$$E_{ij} = \frac{1}{2} (\nabla_i E_j + \nabla_j E_i)$$

$$H_{ij} = \frac{1}{2} (\nabla_i H_j + \nabla_j H_i)$$

$$\begin{aligned} \gamma_0 &= -\gamma_{E1E1} - \gamma_{E1M2} - \gamma_{M1M1} - \gamma_{M1E2} \\ \gamma_\pi &= -\gamma_{E1E1} - \gamma_{E1M2} + \gamma_{M1M1} + \gamma_{M1E2} \end{aligned} \quad (\approx \text{known})$$

# How to extract (dynamical) scalar polarizabilities from data ?

The method used  $\Rightarrow$  Dispersion relations (DRs)

( $\omega < 150$  MeV)

- At low energies, contribution of the 4 spin polarizabilities  $\gamma_{E1E1}, \gamma_{M1M1}, \gamma_{M1E2}, \gamma_{E1M2}$  is «small» (they can be approximately treated as constant)
- Only  $\alpha_{E1}, \beta_{M1}$  be fitted to the data
- Up to now  $\alpha_{E1}, \beta_{M1}$  have been treated as constant (not dependent on  $\omega$ ) (static polarizabilities)
- Dynamical (energy-dependent) polarizabilities  $\alpha_{E1}(\omega), \beta_{M1}(\omega)$  can be defined

(static polarizabilities)

$$\begin{aligned} \alpha_{E1-DYN}^{DR}(\omega) &= f_{\alpha}(\alpha_{E1}, \beta_{M1}, \alpha_{E1,\nu}, \beta_{M1,\nu}) + g_{\alpha}(\gamma_i) + h_{\alpha}(\text{any other term}) \\ \beta_{M1-DYN}^{DR}(\omega) &= f_{\beta}(\alpha_{E1}, \beta_{M1}, \alpha_{E1,\nu}, \beta_{M1,\nu}) + g_{\beta}(\gamma_i) + h_{\beta}(\text{any other term}) \end{aligned}$$

Low Energy Expansion (LEX) up to  $\omega^5$

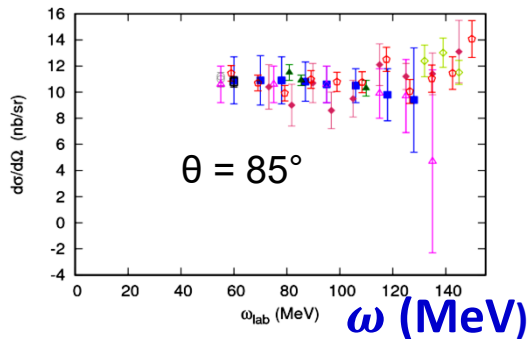
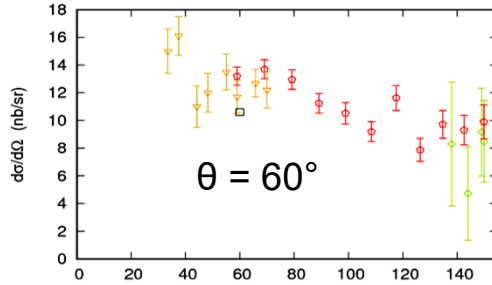
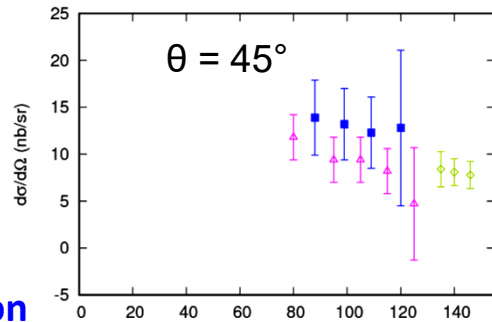
2 new additional parameters to be fitted

Evaluated using measured values and/or model calculations

# The proton data base

( $\omega < 150$  MeV)

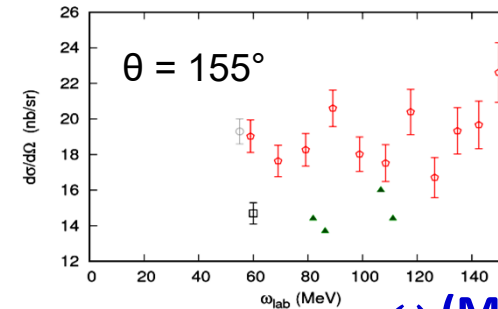
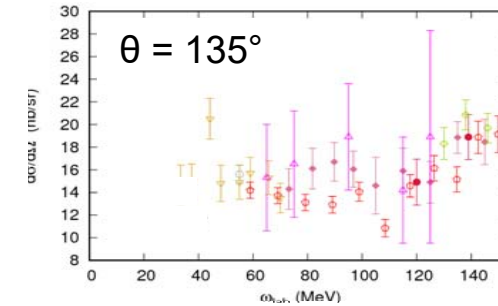
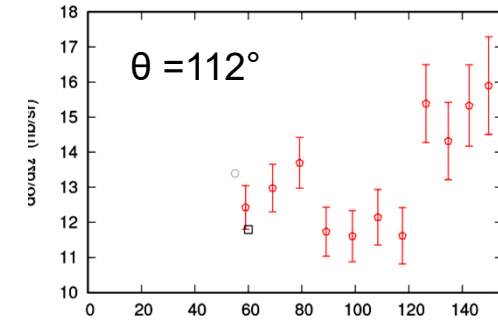
Differential  
Cross section  
(nb/sr)



| Symbol | Set    | Ref               |
|--------|--------|-------------------|
|        | GOL 60 | Goldansky et al.  |
|        | OdL 01 | Olmos de León     |
|        | HAL 93 | Hallin et al.     |
|        | HYM 59 | Hyman et al.      |
|        | PUG 67 | Pugh et al.       |
|        | FED 91 | Federspiel et al. |
|        | BER 61 | Bernardini et al. |
|        | BAR 74 | Baranov et al.    |
|        | OXL 58 | Oxley             |
|        | MAC 95 | MacGibbon et al.  |

(Only) 150 points

*(Half of the Spartans that  
King Leonidas led to the  
Battle of Thermopylae in  
480 BC)*



Differential  
Cross section  
(nb/sr)

$\omega$  (MeV)

⇒ **Poor quality of the data set** ( ... a difficult experiment to perform ...)

**Large statistical -and systematic- errors ; possible inconsistencies between subsets**

## Complications

**3-parameter fit**  $(\alpha_{E1} - \beta_{M1}) ; \alpha_{E1,\nu} ; \beta_{M1,\nu} \rightarrow (\alpha_{E1} + \beta_{M1})$  from Baldin's sum rule

**Standard gradient** (Newton) method to find the minimum of the “ $\chi^2$  function”  
using first and second derivatives (only statistic errors taken into account)

```
MINUIT WARNING IN HESSE  
===== MATRIX FORCED POS-DEF BY ADDING  
0.13727E-01 TO DIAGONAL.
```

**Too high correlations** between fitted parameters!



**VERY low sensitivity** of the data to dynamical polarizabilities

**NO WAY** to find a “reliable” minimum and to evaluate “reliable” (even if large) errors

**SIMPLEX** method (a purely geometrical search; derivatives are not required) was used  
 (“However, it gives no reliable information about parameter errors, no information whatsoever  
 about parameter correlations” -MINUIT manual) in combination with **parametric bootstrap**

# Parametric bootstrap

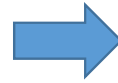
known (purely kinematical variables)

True distribution of the experimental data

$$Y_i \sim p(x_i, \mu_i)$$

known (is the experimental resolution)

unknown true value of  $d\sigma/d\Omega(x_i)$



$$Y_i^* \sim p(x, \mu_i = y_i)$$

known measured value

A model for the distribution of the experimental data

## 2.2 Parametric Simulation

In the previous section we pointed out that theoretical properties of  $T$  might be hard to determine with sufficient accuracy. We now describe the sound practical alternative of repeated simulation of data sets from a fitted parametric model, and empirical calculation of relevant properties of  $T$ .

Suppose that we have a particular parametric model for the distribution of the data  $y_1, \dots, y_n$ . We shall use  $F_\psi(y)$  and  $f_\psi(y)$  to denote the CDF and PDF respectively. When  $\psi$  is estimated by  $\hat{\psi}$  — often but not invariably its maximum likelihood estimate — its substitution in the model gives the *fitted model*, with CDF  $\hat{F}(y) = F_{\hat{\psi}}(y)$ , which can be used to calculate properties of  $T$ , sometimes exactly. We shall use  $Y^*$  to denote the random variable distributed according to the fitted model  $\hat{F}$ , and the superscript  $*$  will be used with  $E$ ,  $\text{var}$  and so forth when these moments are calculated according to the fitted distribution. Occasionally it will also be useful to write  $\hat{\psi} = \psi^*$  to emphasise that this is the parameter value for the simulation model.

A.C.Davidson, D.V.Hinkley

Bootstrap Methods and Their Applications

Cambridge University Press, 1997

## Sampling using parametric bootstrap

- The simplest case: only gaussian statistical errors

$$y_i^* \sim \text{Gauss}(y_{i,\text{exp}}, \sigma_{i,\text{exp}}^2) ; i = 1, \dots, N_{\text{tot}} \quad \text{Gaussian distributed statistical errors}$$



Differential cross section value of the i-th point

After each bootstrap cycle

- Simplex algorithm: minimization of  $\chi_{\text{boot}}^2(\theta) = \sum_{I=1}^{N_{\text{tot}}} \left( \frac{y_i^* - t_i(\theta)}{\sigma_{i,\text{exp}}} \right)^2$

final  $\hat{\theta}^*$  parameters are stored in histograms.

- Their probability distribution can then empirically found after a sufficiently large number of cycles

## Sampling using parametric bootstrap

➤ **A more complicated case: inclusion of a systematic error (common scale factor)**

«standard method»  $\chi_{\text{mod}}^2 = \sum_{I=1}^{N_{\text{subset}}} \left( \frac{\alpha \cdot y_{i,\text{exp}} - t_i(\theta)}{\alpha \cdot \sigma_{i,\text{exp}}} \right)^2 + \left( \frac{\alpha - 1}{\sigma_{\text{sys}}} \right)^2$

G .D'Agostini, NIM A 346, 306 (1994)

**Drawbacks:**

- **One additional normalization factor per data subset is needed**
- **Systematic errors may also vary within a given subset**
- **Is the minimum value still chi-squared distributed ? (correlations)**
- **are errors on the final  $\hat{\theta}$  values still Gaussian ?**

«bootstrap method»  $Y_{j,\text{boot}}^{\text{SYS}} = \xi_j^{\text{Subset}} \cdot Y_{j,\text{boot}} ; \quad j = 1, \dots, N_{\text{TOT}}^{\text{Subset}}$

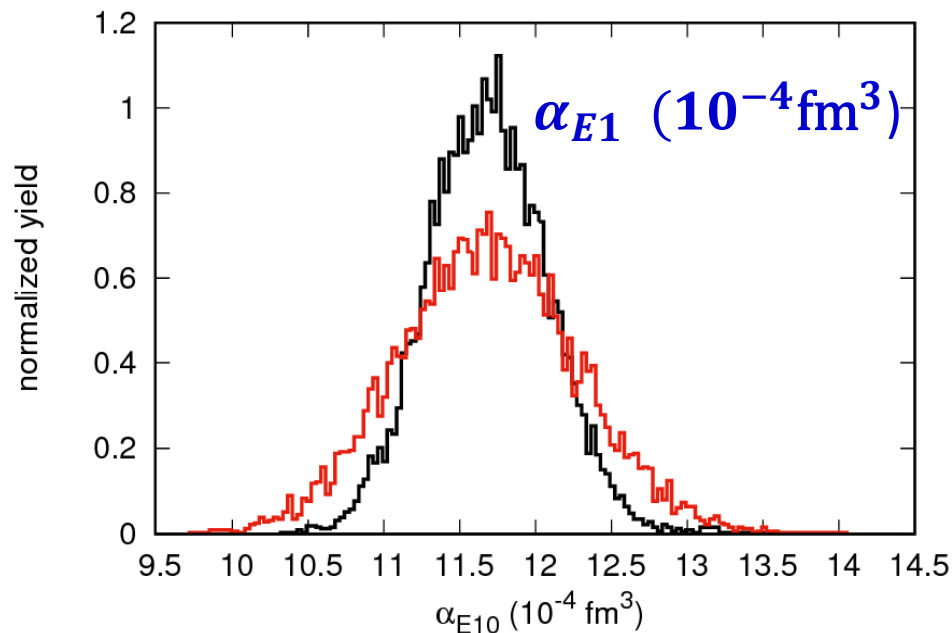
**At every bootstrap cycle the systematic error is sampled independently for each subset**

$$\xi \sim P\left(1, \sigma_{\text{SYS}}^2\right)$$

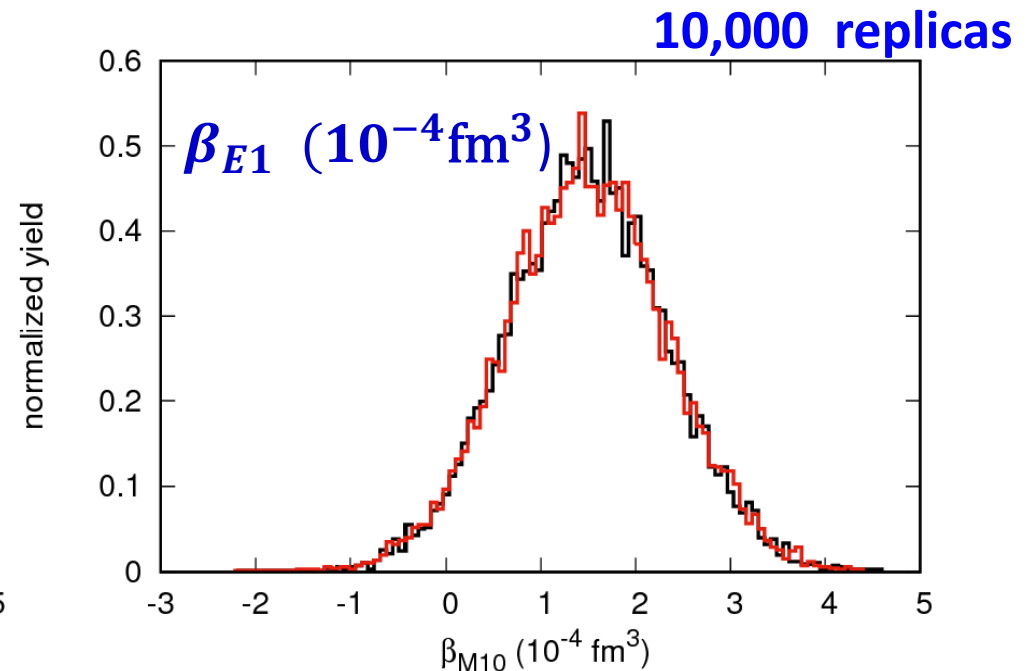
**(uniform, normal, .... distribution)**

## A check: fit only scalar static polarizabilities

2-parameter fit  $\alpha_{E1} ; \beta_{M1}$



———— Systematics OFF  
expected Gaussian shape



———— Systematics ON  
broadening due to systematics

Probability distributions given by our technique (**not** assumed a priori)

The combination of statistical and systematic errors may give non-gaussian distributions

# “Equivalent $\chi^2$ ” value and expected probability distribution from bootstrap

“Equivalent  $\chi^2$ ”  
value

$$(\chi^2_{equiv})_{\min} = \frac{1}{\nu} \sum_{I=1}^{N_{tot}} \left( \frac{y_{i,\text{exp}} - t_i(\hat{\theta}_{boot})}{\sigma_{i,\text{exp}}} \right)^2 \approx 1.4$$

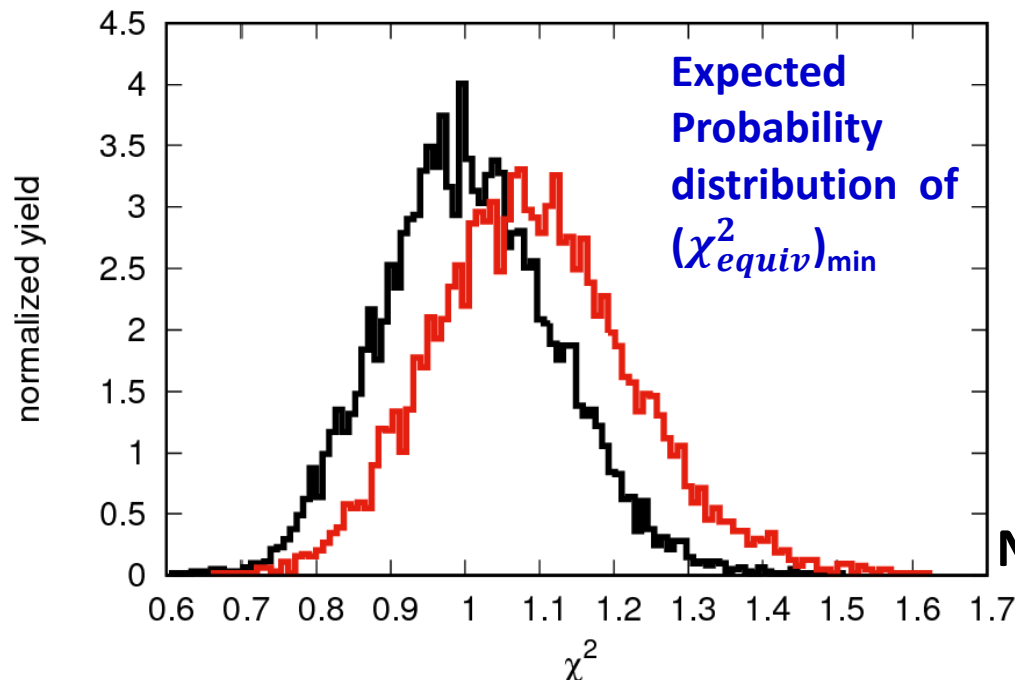
Similar values with/out  
systematic errors

If  $T(\hat{\theta})$  is the  
correct model

$$Y^*_{theo} \sim p(x, \mu = T(\hat{\theta}_{boot}))$$

$$(\chi^2_{theo})_{\min} = \frac{1}{\nu} \sum_{I=1}^{N_{tot}} \left( \frac{y^*_{i,theo} - t_i(\hat{\eta})}{\sigma_{i,\text{exp}}} \right)^2$$

Another  
bootstrap  
sample can be  
generated to  
empirically find  
this distribution



Static polarizabilities fit

— Systematics OFF  
“true”  $\chi^2$  distribution

— Systematics ON



Not a “true”  $\chi^2$  distribution (due to correlations among  
points given by systematics ?)

## Comparison: bootstrap vs gradient

1-parameter fit  $(\alpha_{E1} - \beta_{M1}) \rightarrow (\alpha_{E1} + \beta_{M1})$  from Baldin's sum rule

|                     | $\alpha_{E1}$  | $\beta_{M1}$  |              |
|---------------------|----------------|---------------|--------------|
| GRADIENT SYS OFF    | $11.8 \pm 0.2$ | $2.0 \mp 0.2$ | } Model: DRs |
| BOOTSTRAP SYS OFF   | $11.8 \pm 0.2$ | $2.0 \mp 0.2$ |              |
| BOOTSTRAP SYS ON    | $11.8 \pm 0.3$ | $2.0 \mp 0.3$ |              |
| PDG "Best Estimate" | $11.2 \pm 0.4$ | $2.5 \mp 0.4$ |              |

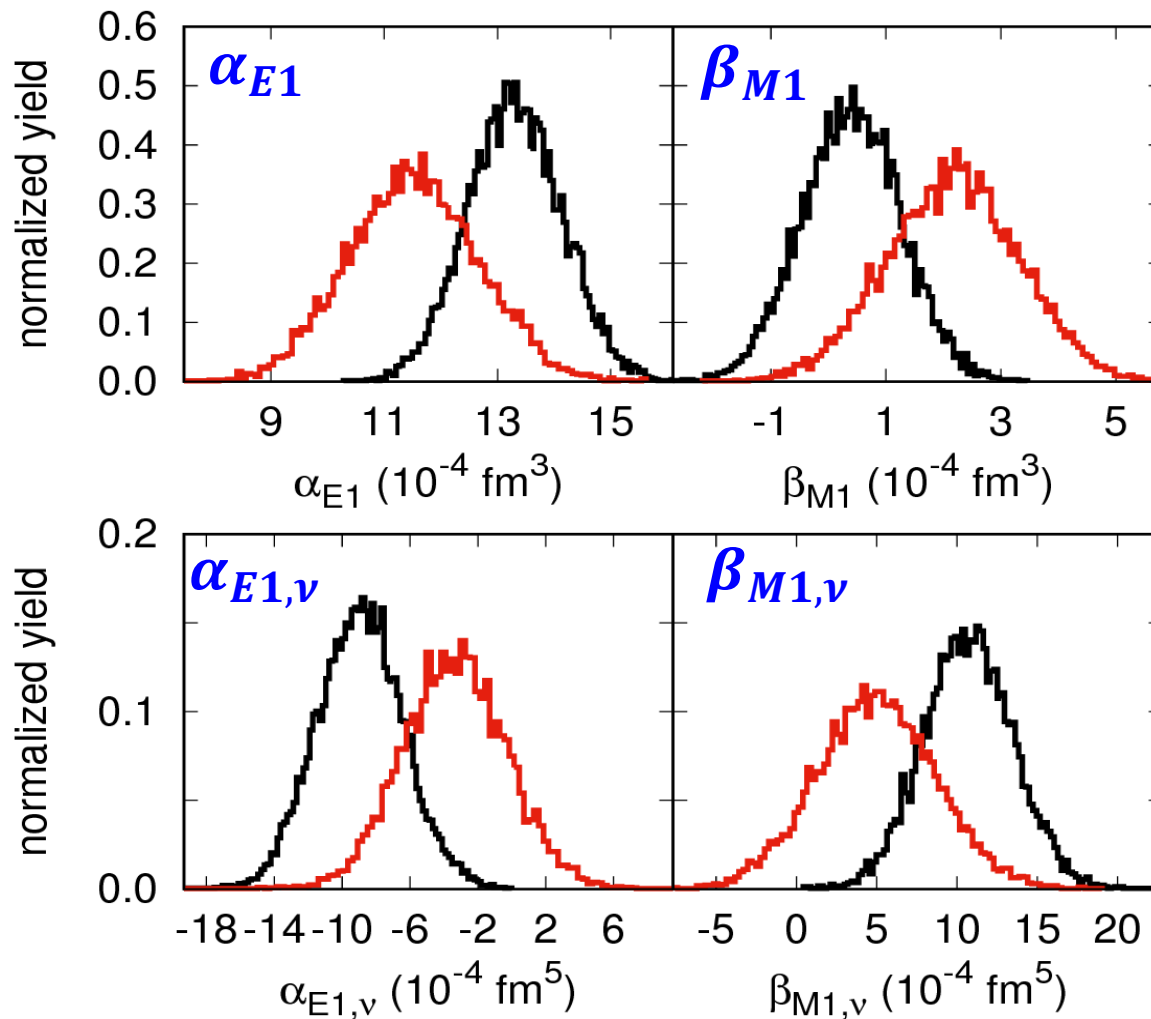
Systematic errors enlarge the error band of the polarizabilities  
(an analysis to detect possible outliers is underway)

## Bootstrap and dynamical polarizabilities

**3-parameter fit**  $(\alpha_{E1} - \beta_{M1}) ; \alpha_{E1,\nu} ; \beta_{M1,\nu}$

- ✓ Baldin's sum rule  $\rightarrow (\alpha_{E1} + \beta_{M1})$
- ✓ Systematic errors ON
- ✓ FULL data set (150 data)
- ✓ TAPS data subset (55 data)  $\rightarrow$  O. De Leon et al., EPJA 10, 207 (2001)
- ✓ Errors on Baldin's sum rule and  $\gamma_\pi$  included in the procedure

## Dynamical polarizabilities: fit distributions



10,000 replicas

**FULL data set**

**TAPS data set**

Probability distributions given by our technique (**not** assumed a priori)

The combination of statistical and systematic errors may give non-gaussian distributions

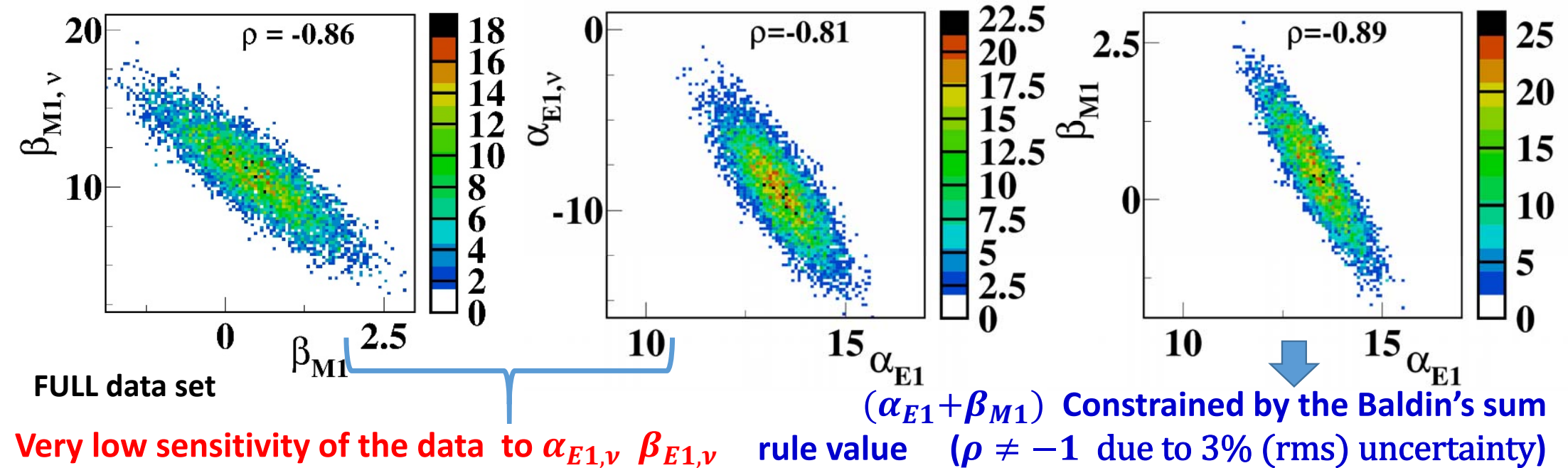
# Dynamical polarizabilities: fit results

B. Pasquini et al,  
PRC 98, 015204  
(2018)

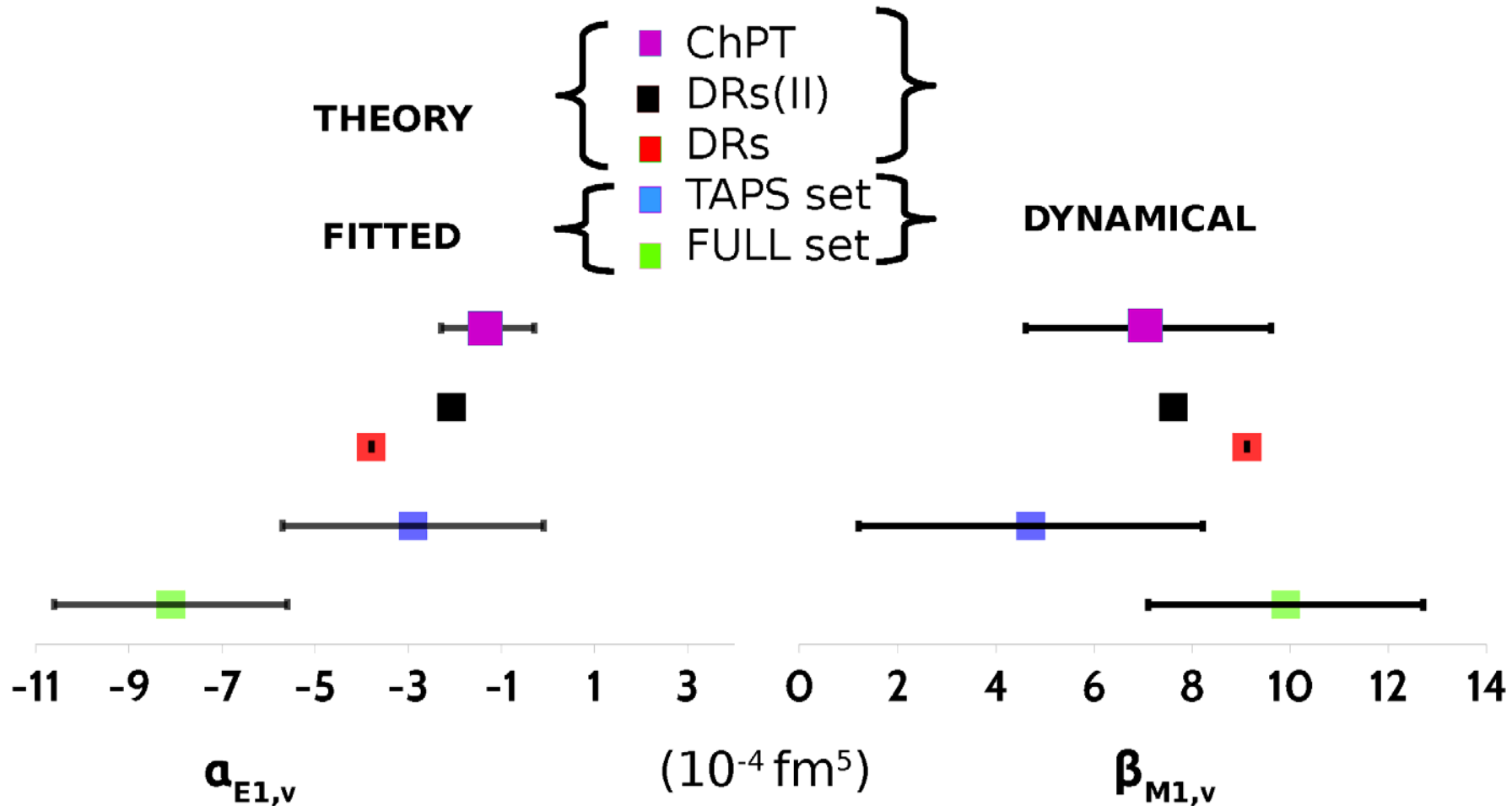
|                   |                        | FULL           | TAPS           |
|-------------------|------------------------|----------------|----------------|
| $\alpha_{E1}$     | $(10^{-4}\text{fm}^3)$ | $13.3 \pm 0.8$ | $11.6 \pm 1.1$ |
| $\alpha_{E1,\nu}$ | $(10^{-4}\text{fm}^5)$ | $-8.8 \pm 2.5$ | $-3.2 \pm 3.1$ |
| $\beta_{M1}$      | $(10^{-4}\text{fm}^3)$ | $0.4 \mp 0.9$  | $2.2 \mp 1.1$  |
| $\beta_{M1,\nu}$  | $(10^{-4}\text{fm}^5)$ | $10.8 \pm 2.8$ | $5.1 \pm 3.7$  |

Quite strong dependence on  
data set (maybe due to outliers or to  
different covered angular regions ?)

Very strong correlations between the  
fit parameters



## Dynamical polarizabilities: comparisons



■ V. Lensky et al., EPJC 75, 604 (2015)

■ B. Pasquini et al., PRC 76, 015203 (2007)

■ B. Holstein et al., PRC 61, 034316 (2000)

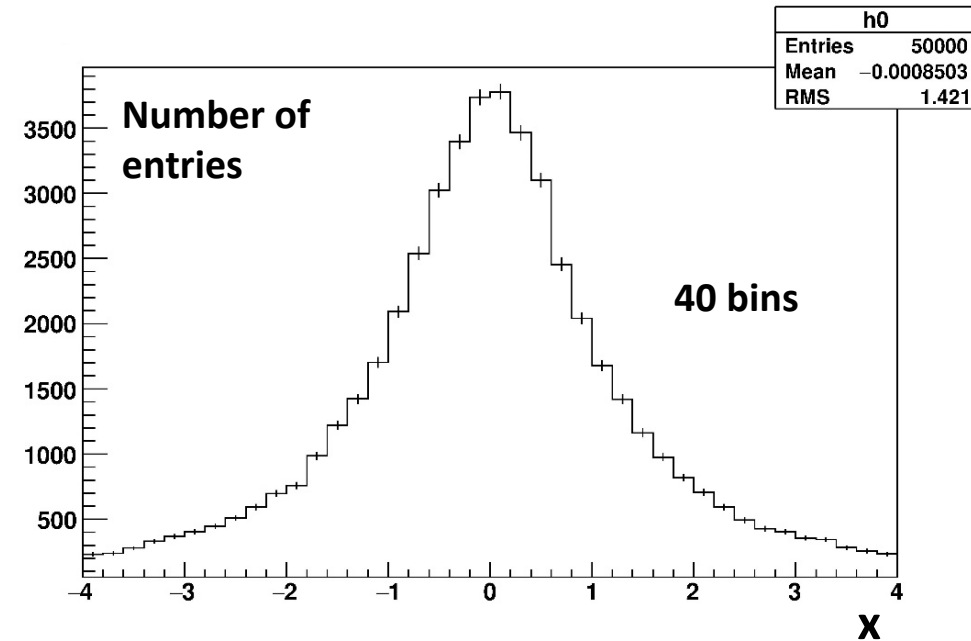
## A further check using a toy model

### ➤ Only statistical errors

Simulated (50000) experimental data generated from the Standard Cauchy Distribution (SCD)

- $\theta_0$  normalisation factor
- $\theta_1$  peak position = 0
- $\theta_2$  FWHM = 1

$$T(x) = \theta_0 \cdot \frac{\theta_2}{(x - \theta_1)^2 + \theta_2^2}$$



| Fit results                | Single minimization            | Bootstrap                      |
|----------------------------|--------------------------------|--------------------------------|
| $\hat{\theta}_0$           | $3767 \pm 17$                  | $3767 \pm 17$                  |
| $\hat{\theta}_1$           | $-(2.5 \pm 5.9) \cdot 10^{-3}$ | $-(2.5 \pm 5.9) \cdot 10^{-3}$ |
| $\hat{\theta}_2$           | $0.998 \pm 0.007$              | $0.998 \pm 0.007$              |
| " $\chi^2_{MIN}$ " / $\nu$ | 0.67                           | 0.67                           |

➡ Standard minimization with gradient method (10,000 replicas)

## A further check using a toy model

### ➤ Systematic errors - I (common scale factor)

Same sample as before

«experimentally estimated» overall possible systematic error  $\sigma_{i,sys}/y_i = 5\%$

- $\theta_0$  normalisation factor
- $\theta_1$  peak position = 0
- $\theta_2$  FWHM = 1

$$T(x) = \theta_0 \cdot \frac{\theta_2}{(x - \theta_1)^2 + \theta_2^2}$$

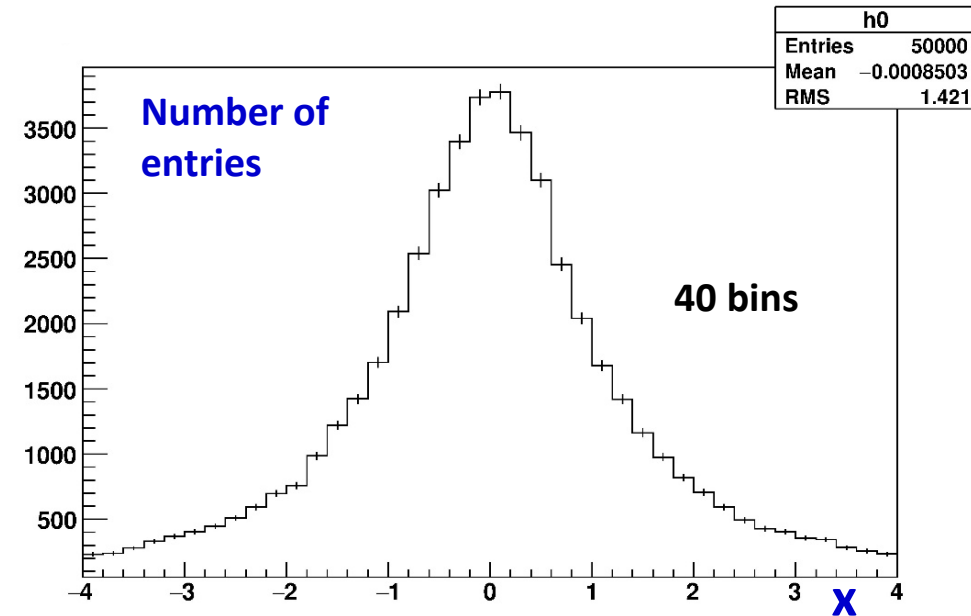
$$\chi_{\text{mod}}^2 = \sum_{I=1}^{N_{\text{subset}}} \left( \frac{\alpha \cdot y_i - t_i(\theta)}{\alpha \cdot \sigma_{i,\text{exp}}} \right)^2 + \left( \frac{\alpha - 1}{\sigma_{\text{sys}}} \right)^2$$

The only difference ➡

$$\sigma_{\hat{\theta}} / \hat{\theta}_0 = 5\%$$

different systematic error distributions  
give the same result, but ....

What is the p.d.f of  $\hat{\theta}_0$  ?



| Fit results          | Single minimization            | Bootstrap                      |
|----------------------|--------------------------------|--------------------------------|
| $\hat{\theta}_0$     | $3767 \pm 189$                 | $3764 \pm 190$                 |
| $\hat{\theta}_1$     | $-(2.5 \pm 5.9) \cdot 10^{-3}$ | $-(2.5 \pm 5.9) \cdot 10^{-3}$ |
| $\hat{\theta}_2$     | $0.998 \pm 0.007$              | $0.998 \pm 0.007$              |
| $\hat{\alpha}$       | $1 \pm 0.05$                   | 1                              |
| " $\chi_{MIN}^2$ "/v | 0.67                           | 0.67                           |

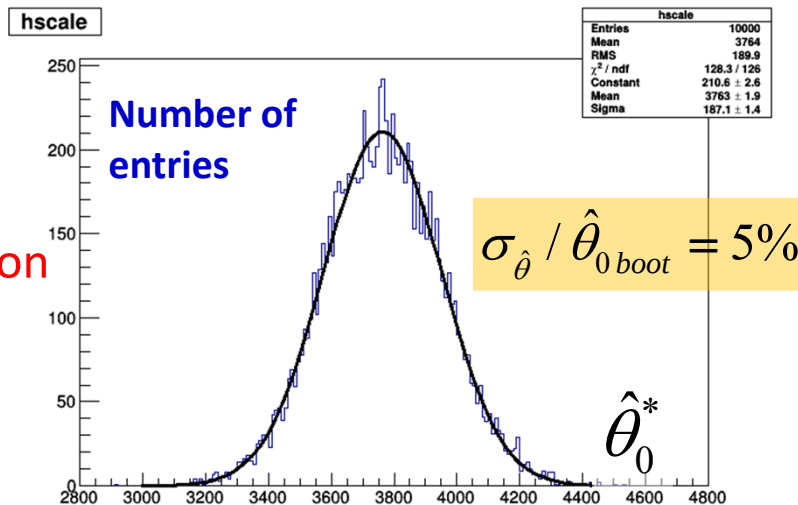
# A check using a toy model

## ➤ Fit with Bootstrap

.... the only way to be always able to answer to the previous question ....

Systematic error  
distribution

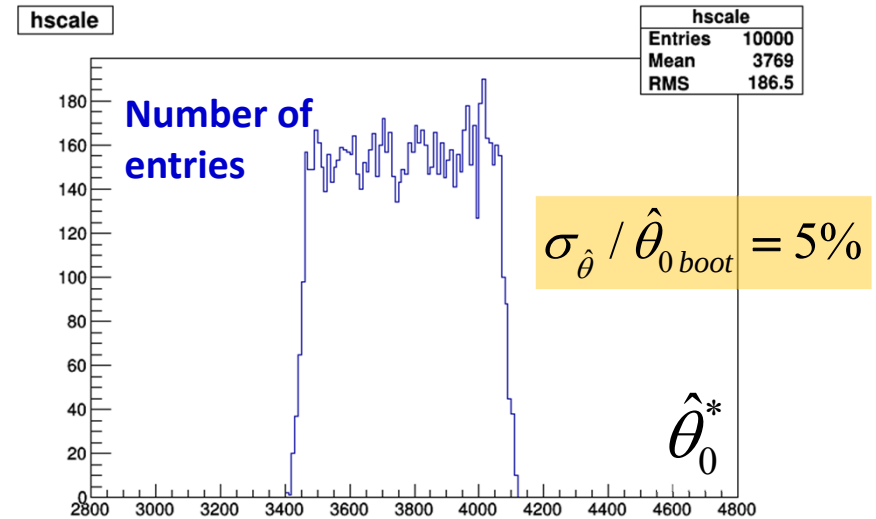
Gaussian



Gaussian

1 $\sigma$  Confidence Level = 68%

Uniform



Uniform

1 $\sigma$  Confidence Level = 58% ( $\neq$  68% !!)

In this special case, the  $\hat{\theta}_0^*$  distribution is the same as the systematic error distribution since  $\sigma_{stat} \ll \sigma_{sys}$

## Another check using a toy model

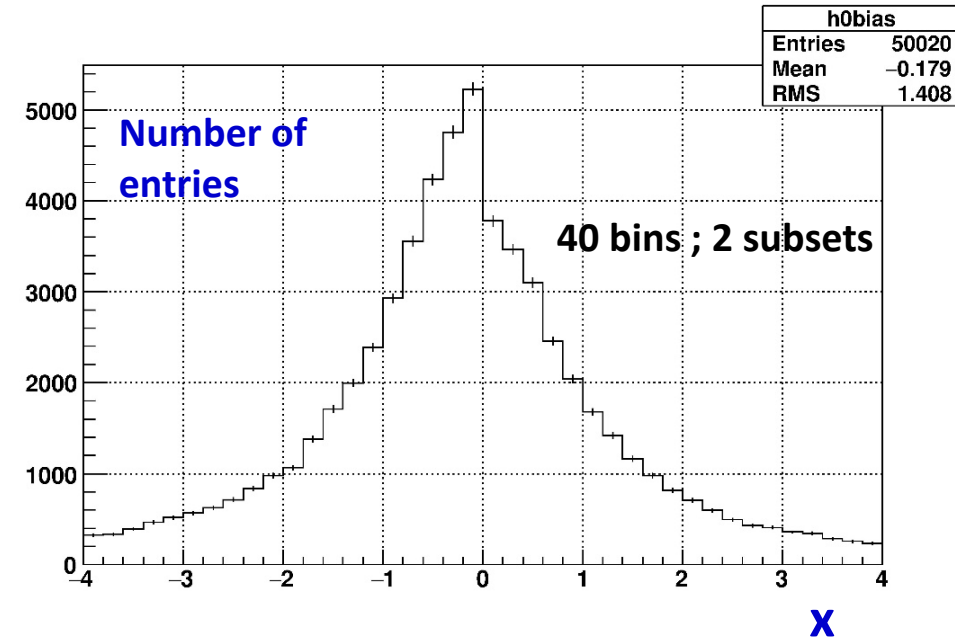
### ➤ Systematic errors -II (common scale factor)

Same sample as before

**Artificial bias** →  $y_i \cdot 1.4$  for  $x < 0$

- $\theta_0$  normalisation factor
- $\theta_1$  peak position = 0
- $\theta_2$  FWHM = 1
- $\alpha$  bias = 0.714 (=1/1.4)
- «experimentally estimated»  $\sigma_{i,sys} = 30\% \cdot y_i$   
(gaussian distribution) for  $x < 0$

$$T(x) = \theta_0 \cdot \frac{\theta_2}{(x - \theta_1)^2 + \theta_2^2}$$



$$\chi_{\text{mod}}^2 = \sum_{I=1}^{N_{\text{subset}}} \left( \frac{\alpha \cdot y_i - t_i(\theta)}{\alpha \cdot \sigma_{i,\text{exp}}} \right)^2 + \left( \frac{\alpha - 1}{\sigma_{\text{sys}}} \right)^2$$

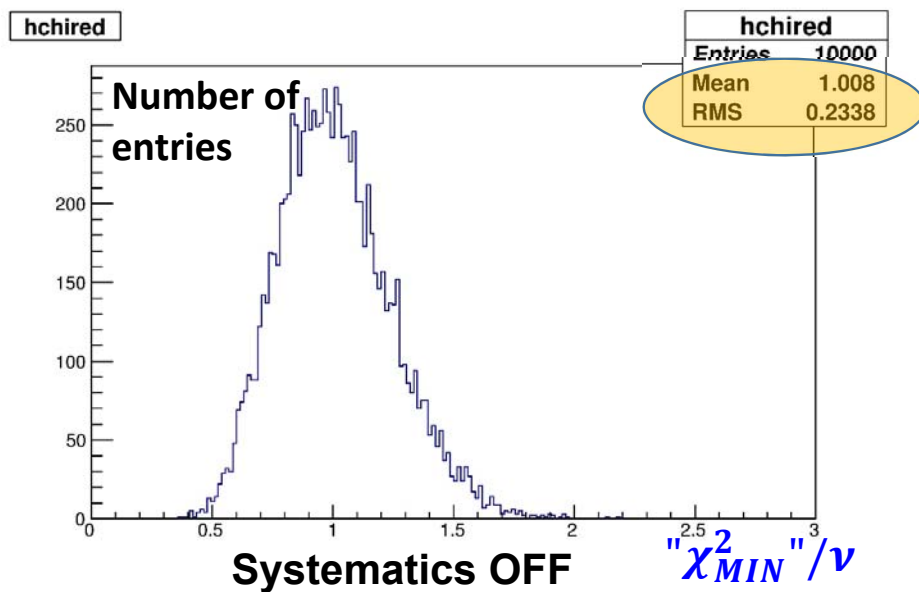
**Only for  $x < 0$**

Is this really a reduced- $\chi^2$  variable ? ➡

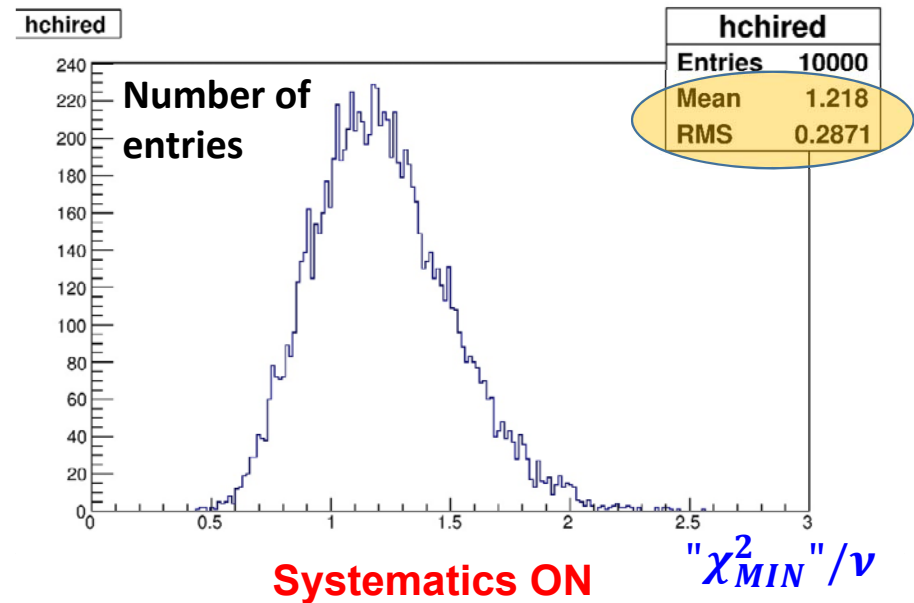
| Fit results                 | Single minimization | Bootstrap         |
|-----------------------------|---------------------|-------------------|
| $\hat{\theta}_0$            | $3831 \pm 49$       | $3780 \pm 16$     |
| $\hat{\theta}_1$            | $-0.02 \pm 0.02$    | $-0.01 \pm 0.01$  |
| $\hat{\theta}_2$            | $0.998 \pm 0.007$   | $0.998 \pm 0.007$ |
| $\hat{\alpha}$              | $0.74 \pm 0.02$     | 0.719             |
| " $\chi_{\text{MIN}}^2$ "/v | 0.79                | 0.78              |

## A “true” bootstrap model

- generate several samples from SCD (real population of the data);
- evaluate " $\chi^2_{MIN}$ "/ $\nu$  for each sample
- the final histogram of all values gives the true " $\chi^2_{MIN}$ "/ $\nu$  distribution



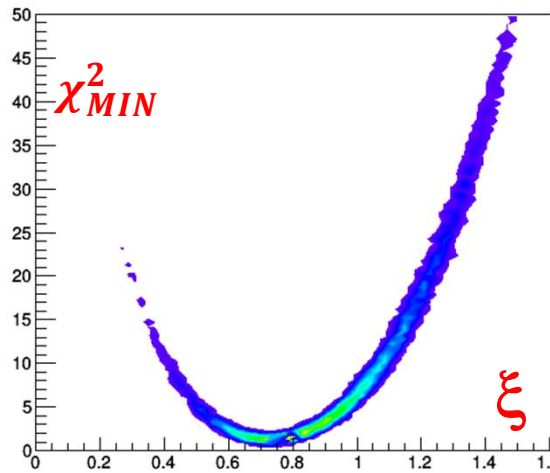
Expected  $\chi^2 / \nu$  distribution



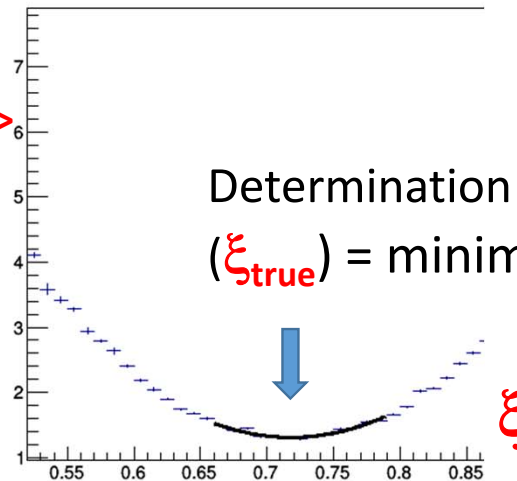
NOT a reduced  $\chi^2$  distribution

# Fit with bootstrap

➤ **First step**  $Y_{j,boot}^{SYS} = \xi_j^{Subset} \cdot Y_{j,boot}$  ;  $\xi \sim P(1, \sigma_{SYS}^2)$



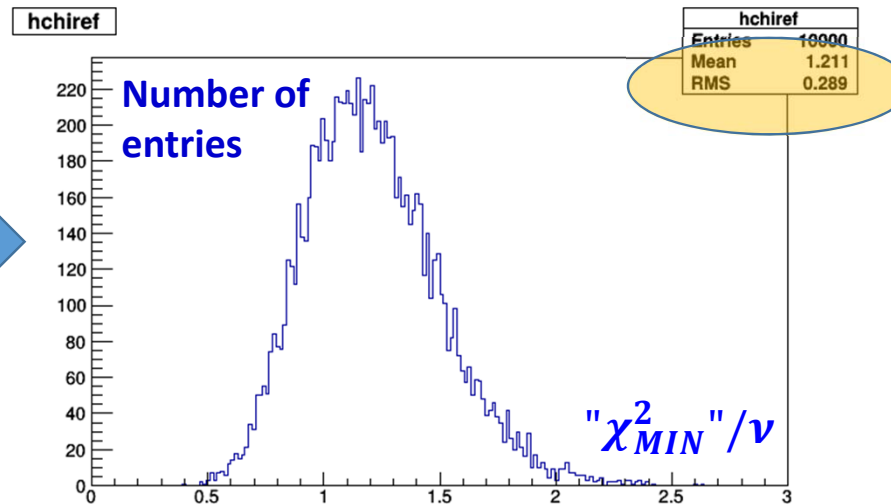
$\langle \chi^2_{MIN} \rangle$



Determination of the «true» bias value  
( $\xi_{true}$ ) = minimum of  $\langle \chi^2_{MIN} \rangle \equiv f(\xi)$

➤ **Second step**

Bootstrapped events  
with  $\xi = \xi_{true}$



Same distribution as  
the «true» one

## Conclusions

### ➤ Fit with Bootstrap

- Very useful and versatile method for data fitting when the “standard” gradient method can not be used and when systematic uncertainties have to be taken into account
- No a priori assumption needed for the experimental errors
- Effect of systematic uncertainties included in a straightforward way
- Can be used in other minimization approaches (maximum likelihood ...)
- As any other bootstrap-based techniques, the reliability of the method is higher when “large” data bases with “small” errors are analysed

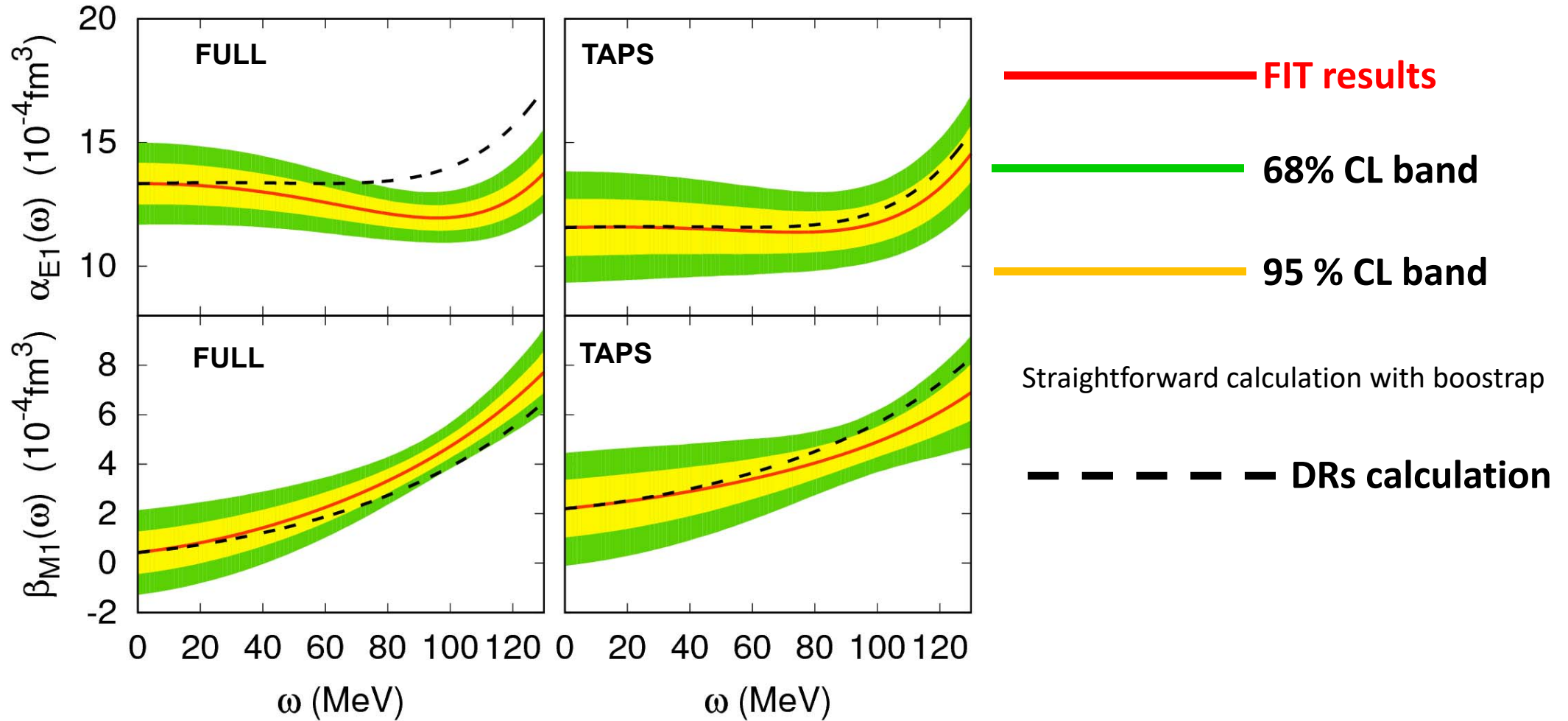


### ➤ First evaluation of the scalar proton dynamical scalar polarizabilities

An experiment ongoing at Mainz (A2@MAMI collaboration) will significantly improve the quality of the existing data base on proton Compton scattering

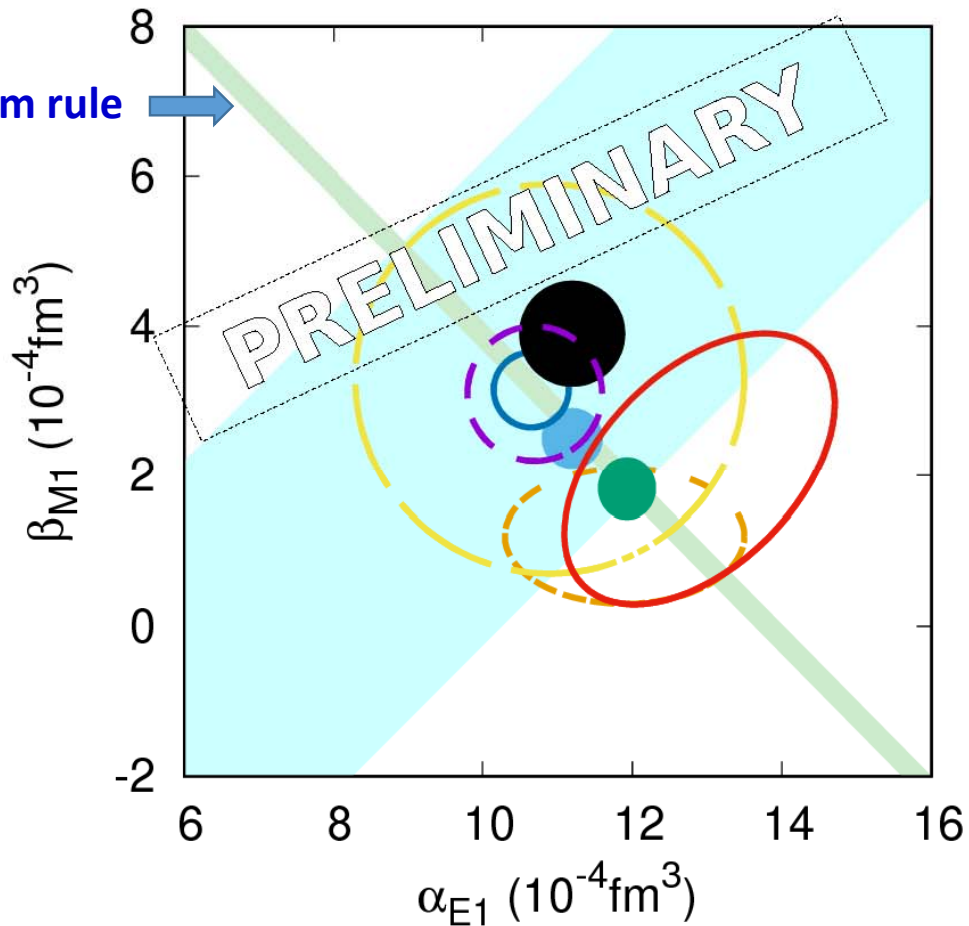
Backup

## Dynamical polarizabilities: fit results



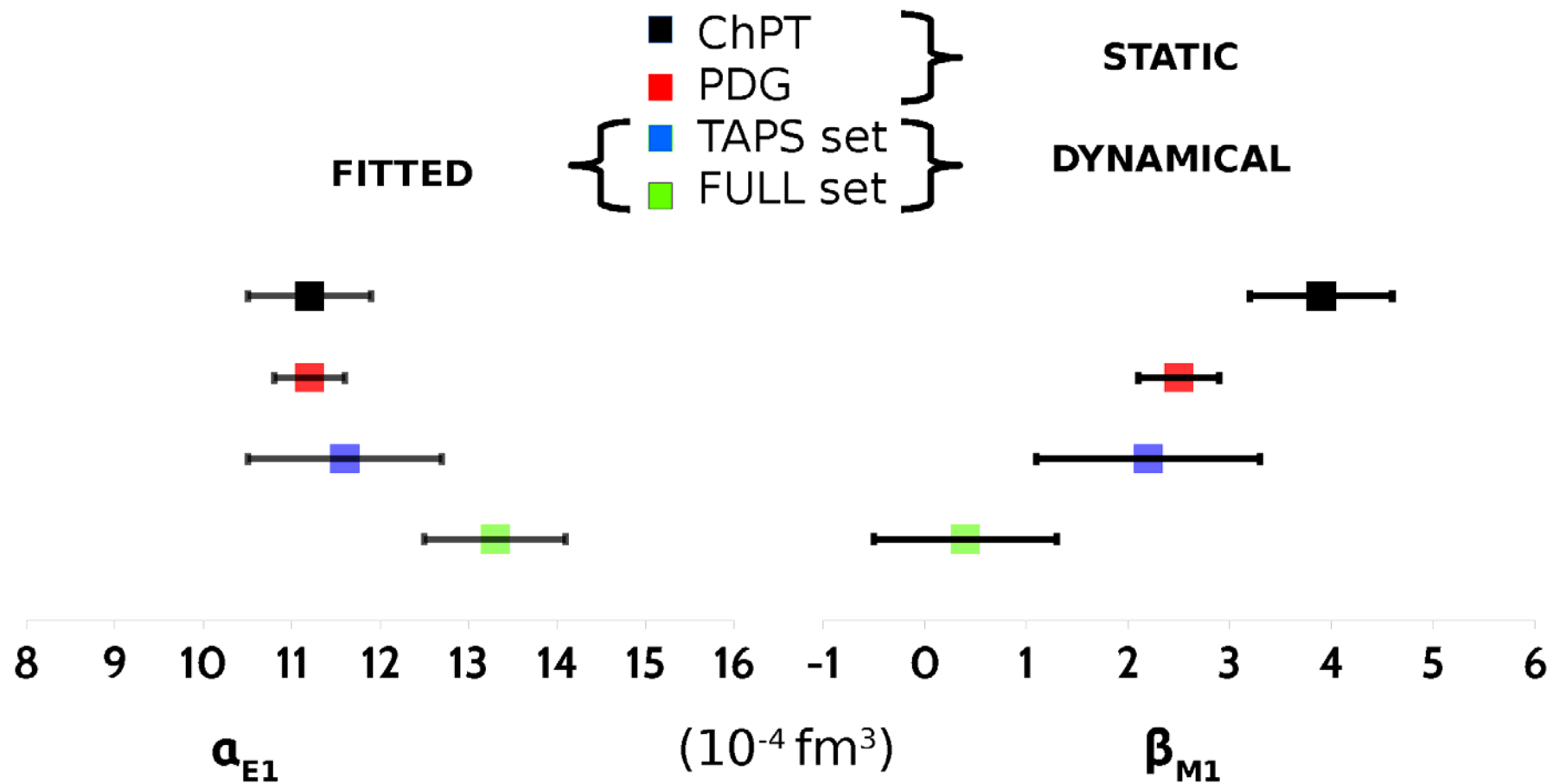
## Summary plot

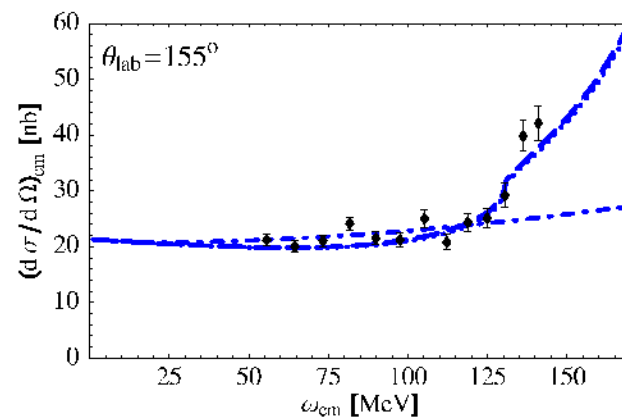
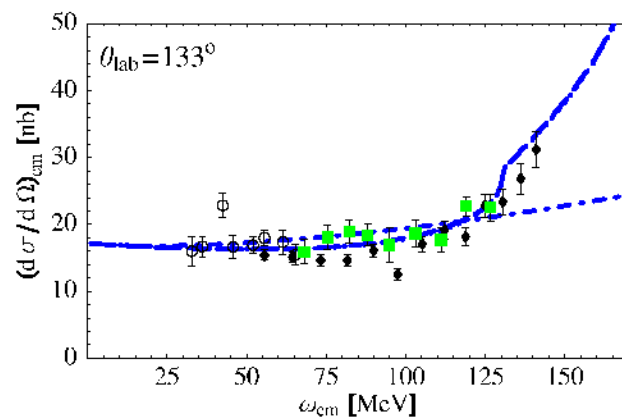
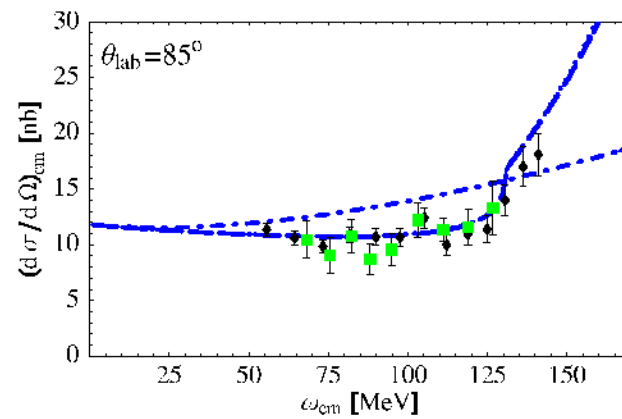
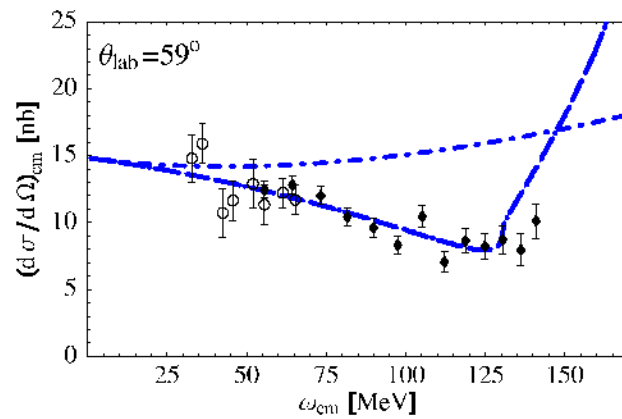
Baldin's sum rule



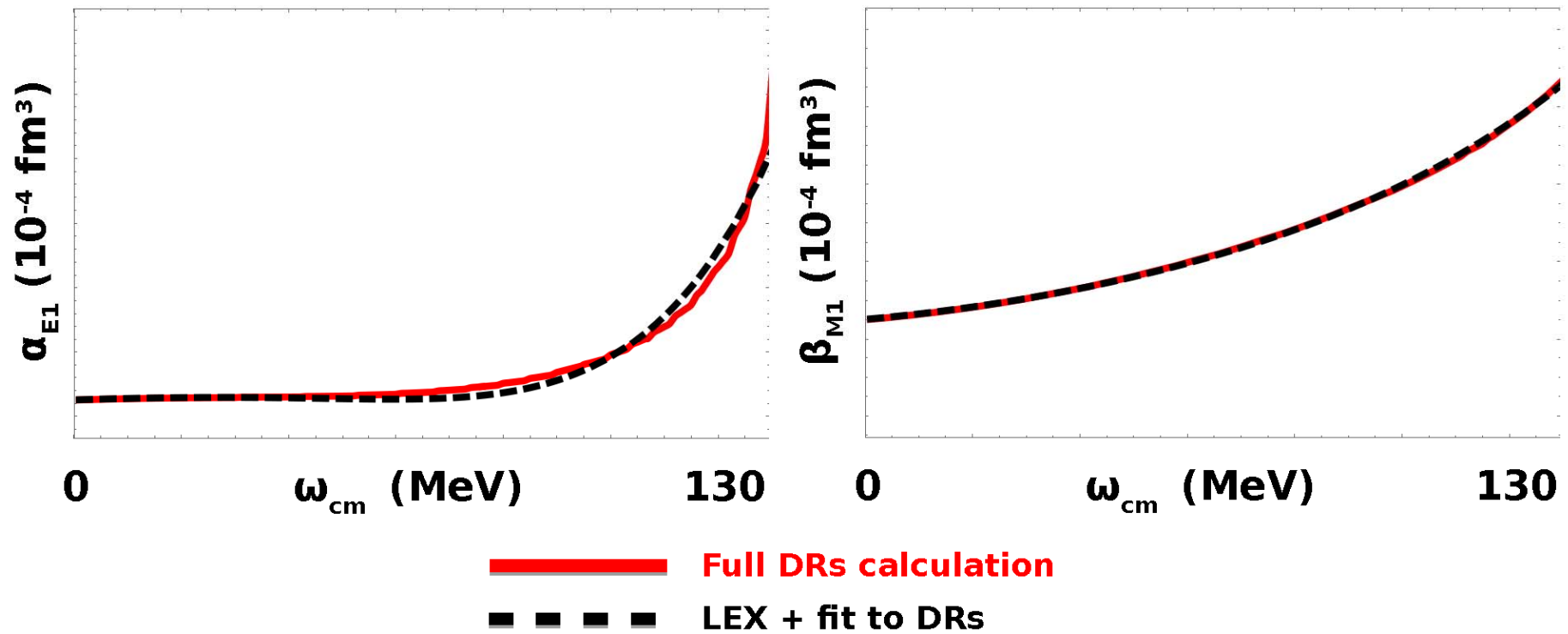
We include Baldin's  
uncertainty &  
systematic sources!

## Dynamical polarizabilities: comparisons

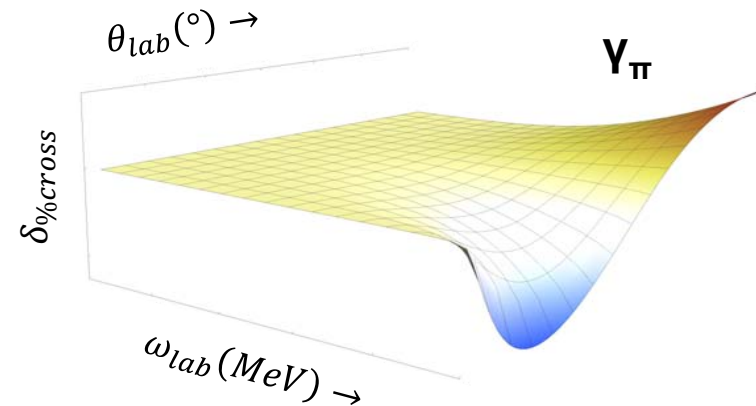
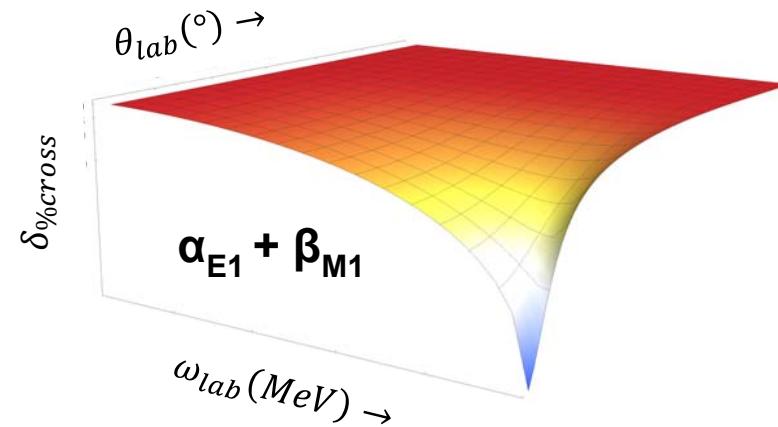
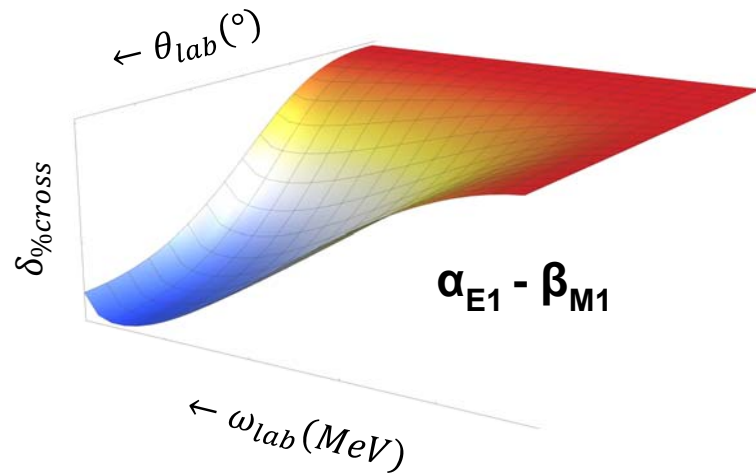




## LEX + residual functions

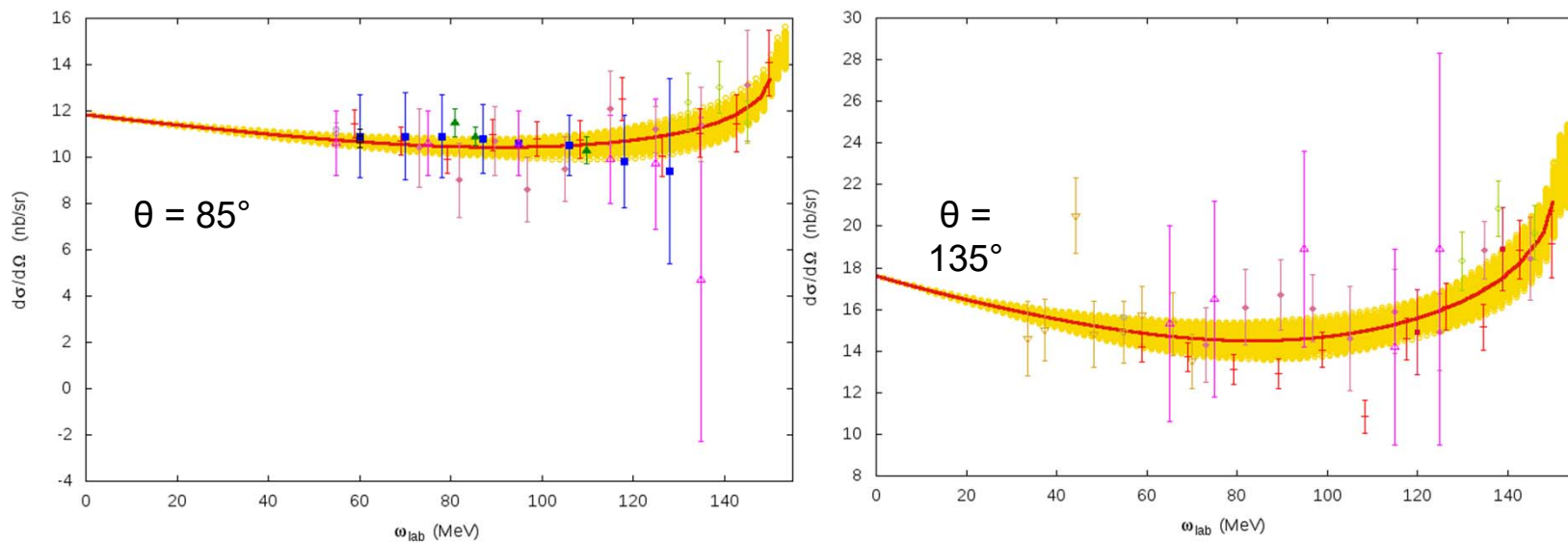


# Unpolarized differential cross section: sensitivity plots



% variation of the observable  
when the particular  
polarizability is changed by a  
factor  $\pm 10\%$

# Differential cross section



$d\sigma/d\Omega$  VS lab energy

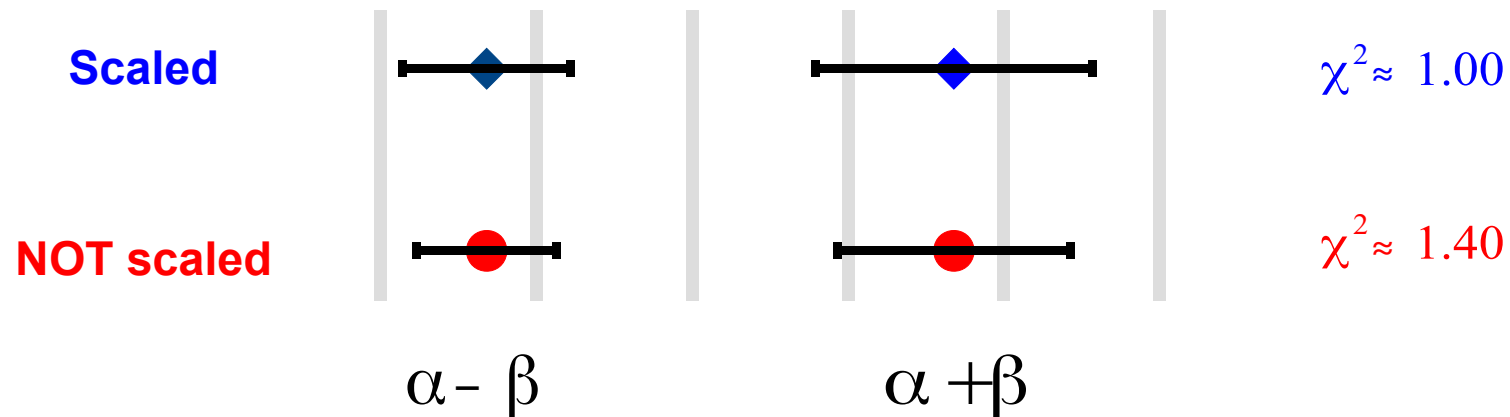
100% error band from the bootstrap fit

# Problems with the data set

Very strong dependence of polarizabilities on the specific data set!

Outliers → rescaling of all the statistic uncertainties by a factor

$$\sqrt{\chi^2}$$



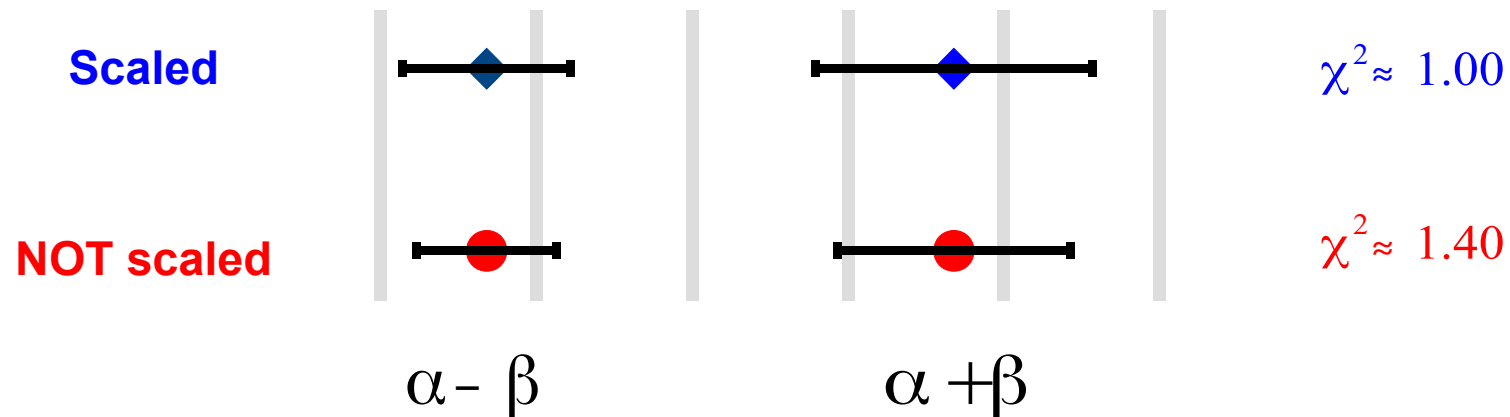
Effect: enlarging of errors on fitted parameters (~20%)

# Problems with the data set

Very strong dependence of polarizabilities on the specific data set!

Outliers → rescaling of all the statistic uncertainties by a factor

$$\sqrt{\chi^2}$$



Effect: enlarging of errors on fitted parameters (~20%)