An automated cross calibration fitting tool- 2

The Basics of Spectral Fitting

Elena González Alonso

(Complutense University)

Tutors: M.Stuhlinger, M.Kirsch

The project

Develop a tool which enables to **fit** the data coming from the EPIC cameras and spectrometers, on board **XMM-Newton,** in an automated way.

Let's pay attention to the **fit** concept…

…we will discover:

→ **Why** the fit concept is needed

→ **Which mathematical tool** we use to do the fitting

→ **How** to know whether we have performed a good fitting or not.

The CCDs in the EPIC cameras obtain photon counts (C) within the specific instruments channels (I). The observed spectrum, C(I), is related to the actual one, f(E), as follows:

$$C(I) = \int_0^\infty f(E) R(I, E) dE$$

R(I,E) = instrumental response (instruments on board satellites use to suffer from some technical behaviour!)

* The goal: f(E) = actual spectrum

* The data: C(I) = observed spectrum,
  R(I,E) = instrumental response

* The relation:

$$C(I) = \int_0^\infty f(E) R(I,E) dE$$

The problem! It is impossible to invert this equation as these invertions tend to be **non-unique** and **unstable** to small changes in C(I) → **the fit concept is needed.**

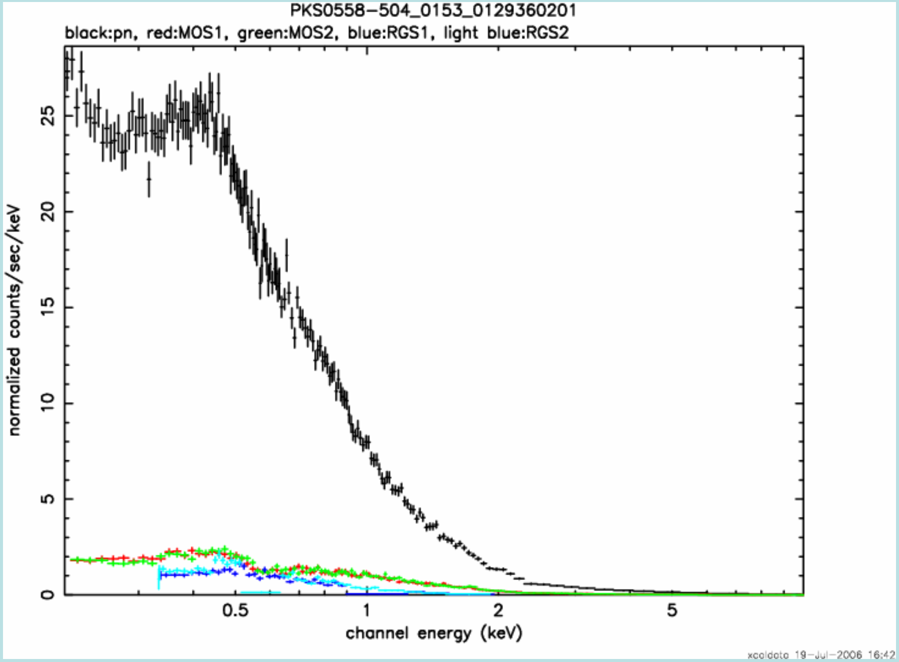
The solution

1. Choose a **model spectrum**, described in terms of n **parameters**, and calculate the **predicted count spectrum**

$$f(E, p_1, p_2, ..., p_n)$$

$$C_m(I) = \int_0^\infty f(E, p_1, p_2, ..., p_n) R(I, E) dE$$

- PKS0558-504 observed spectrum **C(I)** from EPIC cameras, RGS1 and RGS2.

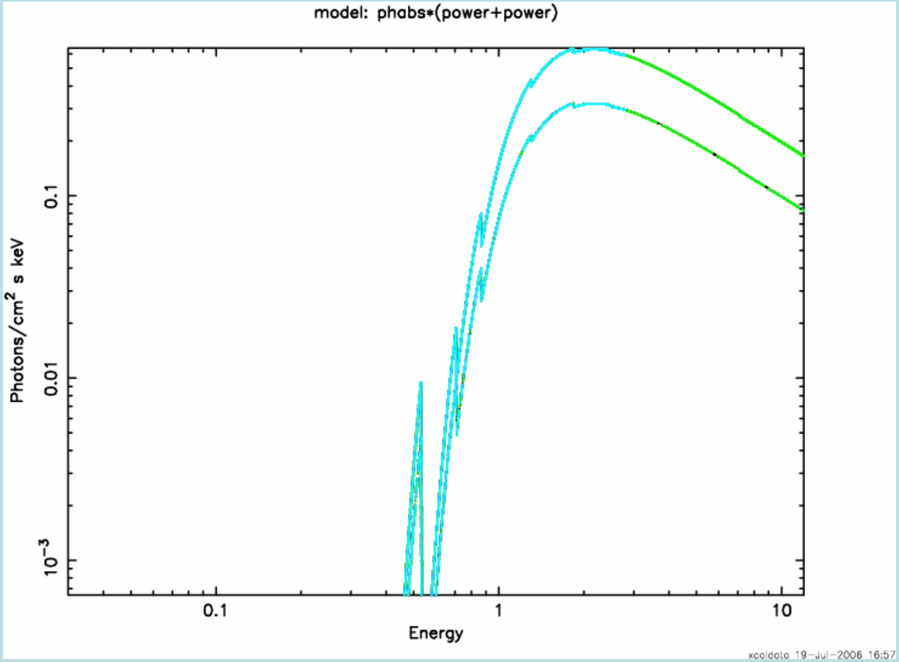


- PKS0558-504 model spectrum **C_m(I)**:

phabs * (power + power)

P_1 = nH **P_1 = phoIndex**

P_2 = norm

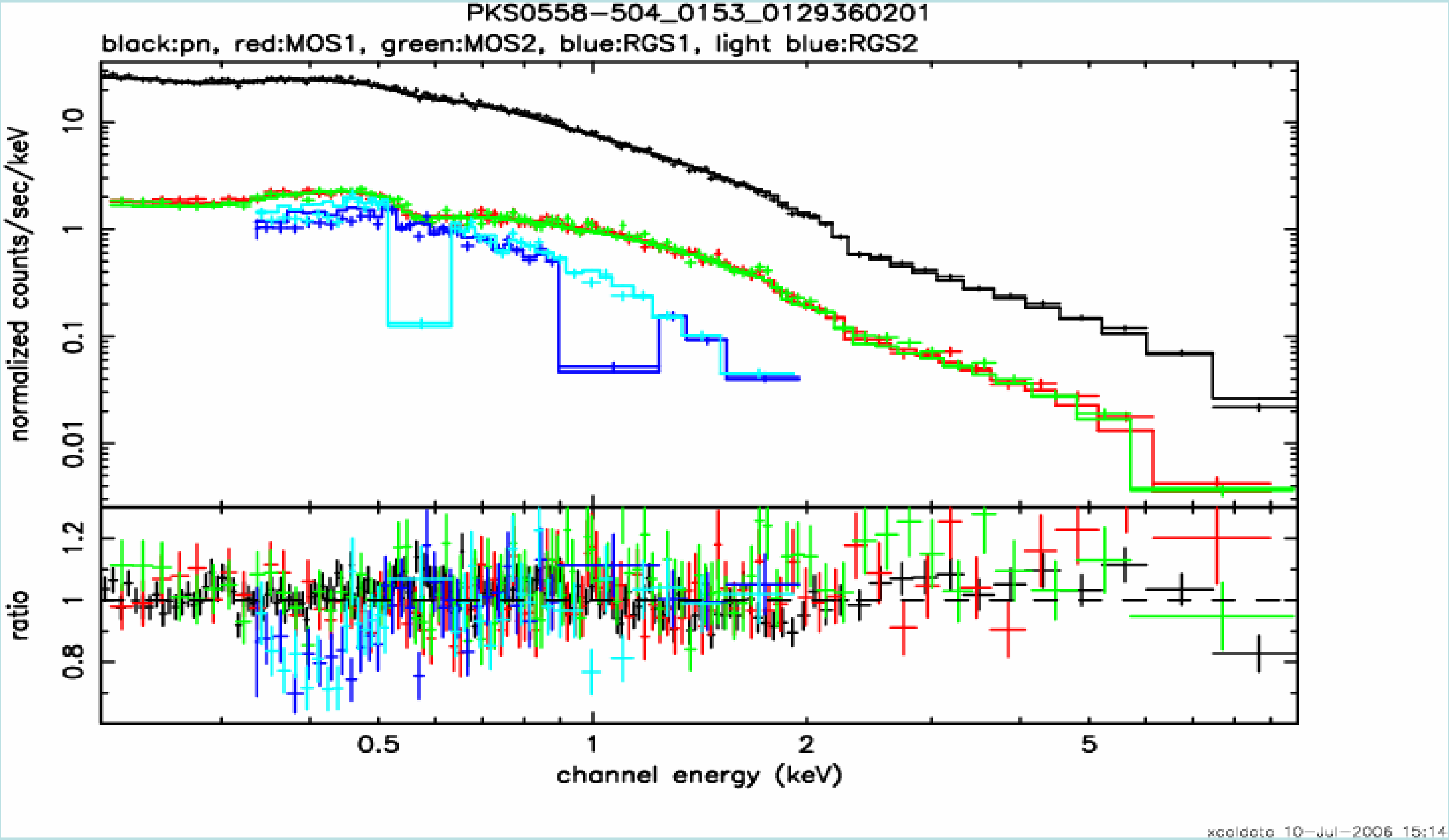


2. **Compare** the predicted count spectrum to the observed one, to decide if the model spectrum **fits** the observed data

$$\chi^2 = \sum_N \frac{(C(I) - C_m(I))^2}{\sigma(I)^2}$$

The model parameters are varied to find the **best fitting**.

PKS0558-504 simultaneous fitting



→ The **Chi-Squared** is the most common statistic in use to fit a function f to data {x, y} :

$$\chi^2 = \sum_N \frac{(y_i - f(x_i))^2}{\sigma_i^2}$$

$N =$ number of data points

$\sigma_i^2 =$ variance (related to the measurement error for y_i)

$y =$ independet variable

x = dependet variable

f = assumed relationship between x and y

$y_i =$ observed mean

$f(x_i) =$ predicted mean

Statistical concepts relevant to data fitting

- The parent distribution has a mean, variance and form that are unknown. By choosing a model we are defining these values:

parent distribution = **model distribution**

- The fitting function f describes the assumed functional relationship between x and y. So, **f predicts the mean for each data point.**

- The difference between the predicted mean and the observed mean is called **deviation**:

$$deviation = f(x_i) - y_i = \Delta y_i \ \text{(positive value)}$$

The Chi-Squared has been defined in general, so, if we suppose:

$$N = \text{ number of channels}$$

$$\sigma(I) = \text{ error for channel I}$$

$$\text{C(I) = observed spectrum (observed counts)}$$

$$\mathrm{C_m}(I) = \text{ model spectrum (model counts)}$$

The Chi-Squared turns to:

$$\chi^2 = \sum_N \frac{(C(I) - C_m(I))^2}{\sigma(I)^2}$$

The **Chi-Squared** characterizes the **dispersion of the observed frequencies from the expected frequencies**, as its components are:

$C(I)$ - $C_m(I)$ = deviation = spread of the observations

σ(I) = good measure of the expected spread.

So, for a GOOD MODEL:

$$C(I) - C_m(I) \approx \sigma(I) \quad \Leftrightarrow$$

$$\frac{(C(I) - C_m(I))^2}{\sigma(I)^2} \approx 1 \Leftrightarrow$$

$$\chi^2 = \sum_N \frac{(C(I) - C_m(I))^2}{\sigma(I)^2} \approx N$$

In fact, the **Chi-Squared** expected value is:

$$\chi^2 = \nu = N - n = \mathrm{degrees\ of\ freedom}$$

N = number of channels

n = number of model parameters

→The general rule to be sure of the **“Goodness-of-fit”** of the model is:

$$\frac{\chi^2}{\nu} = 1$$

It is called the **Reduced Chi-Squared**

Possible situations for the **Reduced Chi-Squared**:

- <u>Case 1</u>: $\dfrac{\chi^2}{\nu} > 1$

- <u>Case 2</u>: $\dfrac{\chi^2}{\nu} < 1$

Case 1: $\dfrac{\chi^2}{\nu} > 1$

$$\chi^2 > \nu = N - n \;\Rightarrow$$

$$\left(\frac{\mathrm{C(I)\text{-}C_m}(I)}{\sigma(I)}\right)^2 > 1 \;\; \forall \text{channel} \;\Leftrightarrow$$

$$\frac{\mathrm{C(I)\text{-}C_m}(I)}{\sigma(I)} > 1 \;\; \forall \text{channel}$$

There are two possibilities again:

(a) If σ(I) is "too small" $\Rightarrow$

$$\frac{\mathrm{C(I)\text{-}C_m}(I)}{\sigma(I)} \textit{ is } \text{"too large"} \Rightarrow$$

$$\left(\begin{array}{c} \text{If the model is "perfect"} \Rightarrow \mathrm{C(I)\text{-}C_m}(I) \text{ is "too small"} \Rightarrow \\ \chi^2 = \left(\frac{\mathrm{C(I)\text{-}C_m}(I)}{\sigma(I)} \right)^2 \approx 1 \; \forall \text{ channel} \Rightarrow \\ \frac{\chi^2}{\nu} \approx 1 \qquad \text{¡¡contradiction!!} \end{array} \right)$$

$\Rightarrow \sigma(I)$ UNDERESTIMATED!!!

(b) If $\text{C(I)-C}_\text{m}(I)$ is "too large" $\Rightarrow$

$$\frac{\text{C(I)-C}_\text{m}(I)}{\sigma(I)} \textit{ is } \text{"too large"} \Rightarrow$$

$$\left(\begin{array}{c} \text{If } \sigma(I) \text{ is also "too large"} \Rightarrow \\ \chi^2 = \left(\dfrac{\text{C(I)-C}_\text{m}(I)}{\sigma(I)} \right)^2 \approx 1 \ \forall \text{ channel} \Rightarrow \\ \dfrac{\chi^2}{\nu} \approx 1 \qquad \text{¡¡contradiction!!)} \end{array} \right)$$

$\text{C}_\text{m}(I)$ INCORRECT!!!

$$\underline{\text{Case 2}}: \ \frac{\chi^2}{\nu} < 1$$

$$\chi^2 < \nu = N - n \ \Rightarrow$$

$$\frac{\text{C(I)-C}_{\text{m}}(I)}{\sigma(I)} < 1 \ \ \forall \text{channel} \ \Leftrightarrow$$

$$\text{C(I)-C}_{\text{m}}(I) < \ \sigma(I) \ \forall \text{channel} \ \Rightarrow$$

$\sigma(I)$ OVERESTIMATED!!!

<u>Summary</u>: Testing the "Goodness-of-fit"

* $\dfrac{\chi^2}{\nu} \approx 1 \Rightarrow$ GOOD FITTING!!

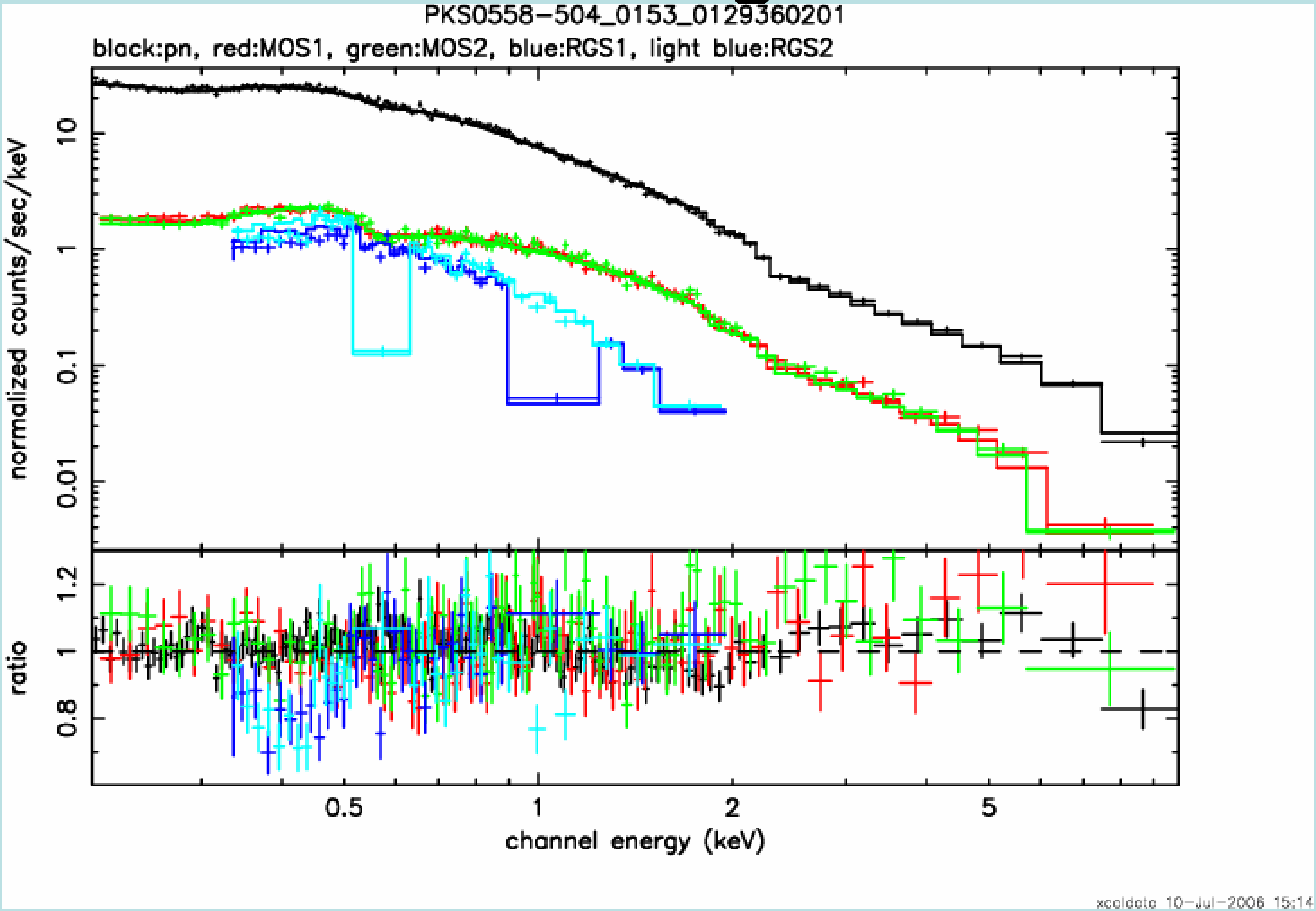
* $\dfrac{\chi^2}{\nu} > 1 \Rightarrow \{\sigma(\mathrm{I})$ UNDERESTIMATED or INCORRECT MODEL $\}$

  $\Rightarrow$ BAD FITTING!!

* $\dfrac{\chi^2}{\nu} < 1 \Rightarrow \{\sigma(\mathrm{I})$ OVERESTIMATED$\}$

  $\Rightarrow$ BAD FITTING!!

PKS0558-504 simultaneous fitting



Testing the "Goodness-of-fit" for PKS0558-304

Best model: $\text{phabs} * (\text{power} + \text{power})$

$$\frac{\chi^2}{\nu} = 1.02668 \approx 1$$

Best parameters:

$* \text{ Phabs} \rightarrow$ nH: 4.382314 atoms/cm^2

$* \text{ Powerlaw } 1 \rightarrow$ phoIndex: 2.79705 (dimensionless)

norm: $6.897345 \cdot 10^{-3}$ photons / keV / cm^2 / s at keV

$* \text{ Powerlaw } 2 \rightarrow$ phoIndex: 2.79705 (dimensionless)

norm: $6.627209 \cdot 10^{-4}$ photons / keV / cm^2 / s at keV