Double Chooz Final Fit

G. Mention

CEA Saclay (DSM/Irfu/SPP)

GdR Neutrinos - June 21st 2012

Overview

***Disclaimer**: this talk is not a statistical course. The goal is to show the general idea of the statistical analysis performed in Double Chooz. Some specificities are not addressed here.*

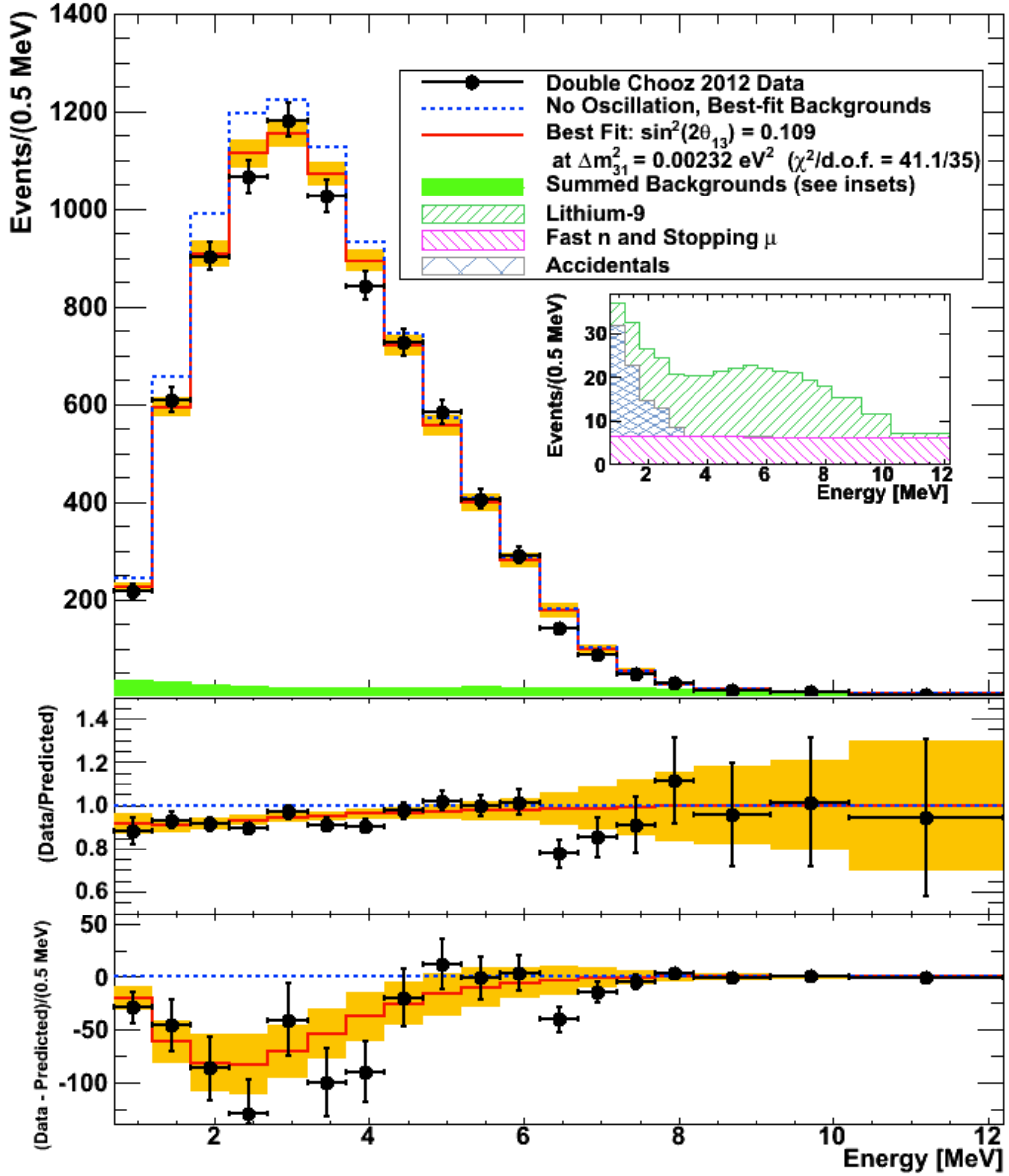
- Statistical approach
- Model and parameters
- Covariances and systematic effects
- Nuisance parameters
- Mixed approach
- Diagnostics

Statistical approach

- Double Chooz statistics is large (*i.e.* >>1 event)
- Double Chooz collaboration chose a χ^2 approach:
  - Given large statistics, χ^2 is equivalent to a binned Likelihood Ratio
  - Compared to an event by event Likelihood, need less care on the development at current stage.
- Selected events are binned in energy intervals.
- A non-constant binning is chosen to ensure sufficient statistic per bin.
- Binning was fixed prior to data release.

Model and parameters

- Physical parameters, a single one: θ_{13} (*i.e.* $\sin^2(2\theta_{13})$).
- Neutrino event rates prediction: $N_i(\theta_{13})$.
- Background events which mimic neutrino events: accidentals, fast-n and Li9.
- Systematic effects induce extra uncertainties on $N_i(\theta_{13})$ and background subtraction to data.
- Two possibilities:
    - take directly into account the systematic induced uncertainties in the χ^2 through a covariance matrix;
    - model the systematic effects as nuisance parameters.



List of systematics

Example of list of systematics and uncertainties for Double Chooz 1st publication (Phys.Rev.Lett. 108 (2012) 131801, arXiv:1112.6353 [hep-ex])

TABLE II. Contributions of the detector and reactor errors to the absolute normalization systematic uncertainty.

Detector		Reactor	
Energy response	1.7%	Bugey4 measurement	1.4%
E_{delay} Containment	0.6%	Fuel Composition	0.9%
Gd Fraction	0.6%	Thermal Power	0.5%
Δt_{e+n}	0.5%	Reference Spectra	0.5%
Spill in/out	0.4%	Energy per Fission	0.2%
Trigger Efficiency	0.4%	IBD Cross Section	0.2%
Target H	0.3%	Baseline	0.2%
Total	2.1 %	Total	1.8%

All these systematics affect the rate. There are other systematics which affect the shape: background shape uncertainties, reactor spectrum uncertainties and energy scale uncertainty.

Goal of any parameter estimation

- The goal of any parameter estimation is to transform the information we have from data into summarized information on parameters of the model

$$\chi^2 = \sum_i \left(\frac{\text{Data}_i - \text{Model}_i(\text{Params})}{\text{Uncertainty}_i} \right)^2$$

$$\chi^2 = \sum_{p,q} \left(\text{Params}_p - \text{BestEstim.}_p\right) \text{COV}^{-1}_{p,q} \left(\text{Params}_q - \text{BestEstim.}_q\right) + \chi^2_{\text{min}}$$

$$\chi^2 = (\alpha - \hat{\alpha})' \, V_\alpha^{-1} \, (\alpha - \hat{\alpha}) + \chi^2_{\text{min}}$$

In non-linear case only

χ^2 with covariance approach

Here, we take directly into account the systematic induced uncertainties in the χ^2 through a covariance matrix:

$$Y_i = N_i(\theta_{13}) + B_i + \varepsilon_i$$

$$\text{or} \quad Y_i - B_i - N_i(\theta_{13}) = \varepsilon_i$$

and $\quad \mathrm{Cov}(\varepsilon_i, \varepsilon_j) \neq \delta_{ij}\sigma_i^2$

i.e. there are correlations among the residuals due to systematic effects, background subtraction uncertainties...

The associated χ^2 (*i.e.* statistical quantity which should follow a χ^2 distribution if hypotheses are correct) is then of the form:

$$\chi^2 = \sum_{i,j} \varepsilon_i V_{i,j}^{-1} \varepsilon_j = \sum_{i,j} \left(Y_i - N_i(\theta_{13}) - B_i\right) V_{i,j}^{-1} \left(Y_j - N_j(\theta_{13}) - B_j\right)$$

The V matrix is an estimate of the covariance matrix of the errors ε_i.

This χ^2 depends on θ_{13}. Minimizing this χ^2 w.r.t. θ_{13} gives the best estimate on θ_{13} according to data Y, models N and B and covariance estimate V.

χ^2 with nuisance parameters approach

Here we model the systematic effects with nuisance parameters:

$$Y_i = N_i(\theta_{13}, \vec{\alpha}) + \varepsilon_i = N_i(\theta_{13}) + \sum_k \alpha_k S_{i,k} + \varepsilon_i$$

a priori model (known before fit)

$$N_i(\theta_{13}) = N_i(\theta_{13}, \vec{\alpha}^\star)$$

with *a priori* nuisance parameters $\vec{\alpha}^\star$ (known before fit $\pm\sigma_\alpha$)

The systematic effects S are:

$$S_{i,k} = \left.\frac{\partial N_i(\theta_{13}, \vec{\alpha})}{\partial \alpha_k}\right|_{\alpha_k = \alpha_k^\star}$$

The covariance is:

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \delta_{ij}\sigma_i^2$$

The associated χ^2 is then :

$$\chi^2 = \sum_i \left(\frac{\varepsilon_i(\vec{\alpha}^\star + \vec{\alpha}, \theta_{13})}{\sigma_i}\right)^2 + \sum_k \left(\frac{\alpha_k}{\sigma_{\alpha_k}}\right)^2$$

or

$$\chi^2 = \sum_i \left(\frac{Y_i - N_i(\vec{\alpha}^\star + \vec{\alpha}, \theta_{13})}{\sigma_i}\right)^2 + \sum_k \left(\frac{\alpha_k}{\sigma_{\alpha_k}}\right)^2$$

if correlations among α_k

$$\chi^2 = \sum_i \left(\frac{Y_i - N_i(\vec{\alpha}^\star + \vec{\alpha}, \theta_{13})}{\sigma_i}\right)^2 + \sum_{k,k'} \alpha_k W^{-1}_{k,k'} \alpha_{k'}$$

Equivalence between linear nuisance parameters and covariance matrix approaches

Nuisance parameters approach follow a class of fitting models known since a long time as Mixed Effects Models (i.e. fixed and random effects).

$$\chi^2 = \sum_i \left(\frac{Y_i - N_i(\vec{\alpha}^\star + \vec{\alpha}, \theta_{13})}{\sigma_i} \right)^2 + \sum_{k,k'} \alpha_k W^{-1}_{k,k'} \alpha_{k'}$$

minimized over all α_k is

$$\chi^2 = \sum_{i,j} \varepsilon_i V^{-1}_{i,j} \varepsilon_j = \sum_{i,j} \left(Y_i - N_i(\theta_{13}) - B_i\right) V^{-1}_{i,j} \left(Y_j - N_j(\theta_{13}) - B_j\right)$$

with V expressed as function of $S_{i,k}$ $\quad V = \Sigma + SWS'$

and $\quad V^{-1} = \Sigma^{-1} - \Sigma^{-1} S (W^{-1} + S'\Sigma^{-1} S)^{-1} S' \Sigma^{-1}$

with $\quad \Sigma_{i,j} = \delta_{i,j} \sigma_i^2$

The covariance χ^2 is the same as the nuisance parameter χ^2 already minimized on nuisance parameters.

Formalism corrrect even with non-diagonal Σ matrix, *i.e.* it is possible to mix covariance and nuisance parameters approaches!

Current choice in Double Chooz

A mixed covariance-pull approach

- pull parameters take into account
  - backgrounds
  - energy scale
  - Δm^2 uncertainty
- The rest is in covariance matrix:
  - normalization
  - reactor simulation: fuel and burn-up uncertainties, Bugey-4 anchor, power uncertainty,
  - x-section uncertainty
  - efficiencies

Some comments about pull approach

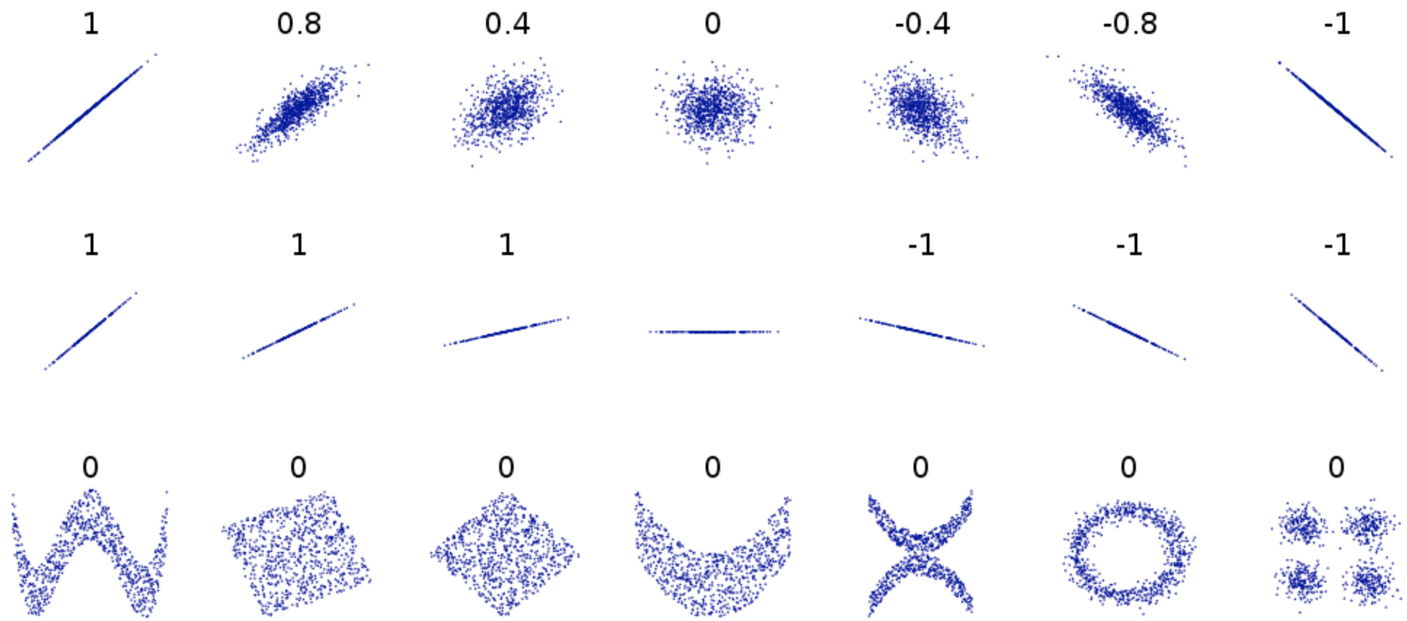
- The pull distribution should follow a normal law (if hypotheses correct)
  values of nuisance parameters at best fit ~ $\mathcal{N}(0,1)$
- QQ-plots, Normality tests, correlation tests (Durbin-Watson,...), etc.
- Can allow to point out tension in fits: statistical tests on best fit parameters possible.
- Comment on the number of degrees of freedom:
  - Each extra α_k/σ_k term can be considered as an extra bin compared to data points Y_i.

$$\chi^2 = \sum_i \left(\frac{Y_i - N_i(\vec{\alpha}^\star + \vec{\alpha}, \theta_{13})}{\sigma_i} \right)^2 + \sum_k \left(\frac{\alpha_k}{\sigma_{\alpha_k}} \right)^2$$

  - Consequently, $\chi^2_{\min} = \min_{\vec{\alpha},\theta_{13}} \chi^2(\theta_{13}, \vec{\alpha})$ should behave as a χ^2 with $N_{bins} + N_{pulls} - N_{par} = N - 1$ degrees of freedom.

# of free params here = 1

Note on covariances:
they capture only linear relationships!



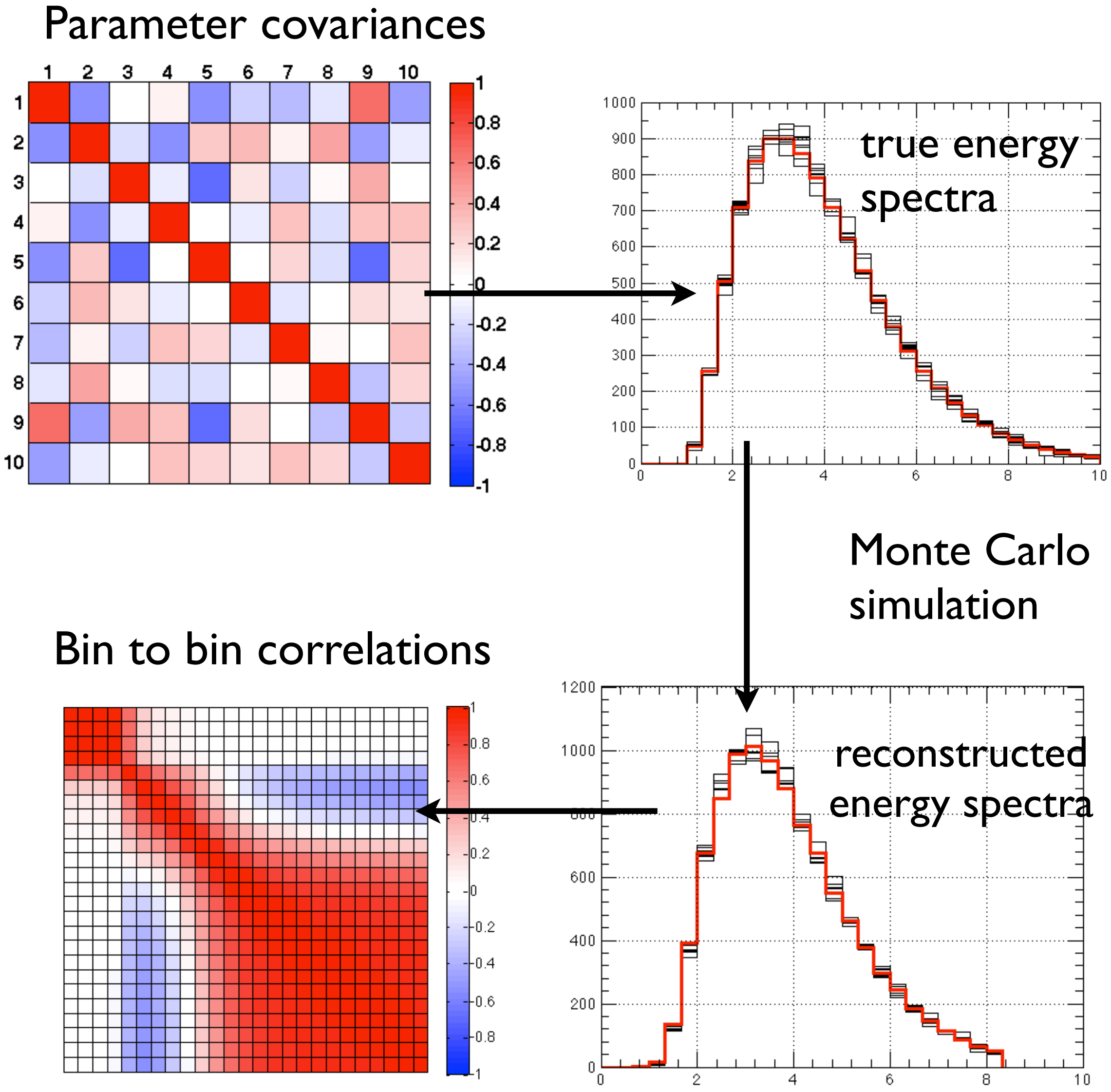
$$\mathrm{Cov}(N_i, N_j) = \langle N_i N_j - \langle N_i N_j \rangle \rangle = \frac{1}{N_{\rm sim} - 1} \sum_{k=1}^{N_{\rm sim}} \left(N_i(k) N_j(k) - \frac{1}{N_{\rm sim}} \sum_{k=1}^{N_{\rm sim}} N_i(k) N_j(k) \right)$$

Several sets of (*x*, *y*) points, with the Pearson correlation coefficient of *x* and *y* for each set. Note that the **correlation reflects the noisiness and direction of a linear relationship** (top row), but not the slope of that relationship (middle), nor many aspects of nonlinear relationships (bottom). N.B.: the figure in the center has a slope of 0 but in that case the correlation coefficient is undefined because the variance of *Y* is zero. [source: Wikipedia]

Nuisance parameters

- I advocate use of **nuisance parameters approach** instead of covariance approach:
  - Because the covariance approach is just the remaining χ^2 after the minimization over α's.
  - Information on best fit values α is lost in covariance approach and both formalisms are equivalent.
  - Using a linear minimizer can solve very quickly the χ^2 minimization w.r.t. α parameters first before eventually tackling non-linear physical parameters.
  - It means:
    - if you use Minuit as a fitter, **provide** Minuit with the gradient estimate otherwise, this nuisance parameter approach suffers from the curse of dimensionality (difficulty of convergence in high parameter dimension spaces)
    - or use a **linear fitter** (ROOT has one inside for instance TLinearFitter,...) to first solve the χ^2 w.r.t. linear parameters.
- The best fit value of nuisance parameters and deviance from a priori information give you interesting diagnostic on the model.
- Possible compatibility tests on nuisance parameters between *a priori* values and data driven values (residuals normality test, Fischer tests, t-tests, etc. + diagnostics described in next slides)

MultiSim approach



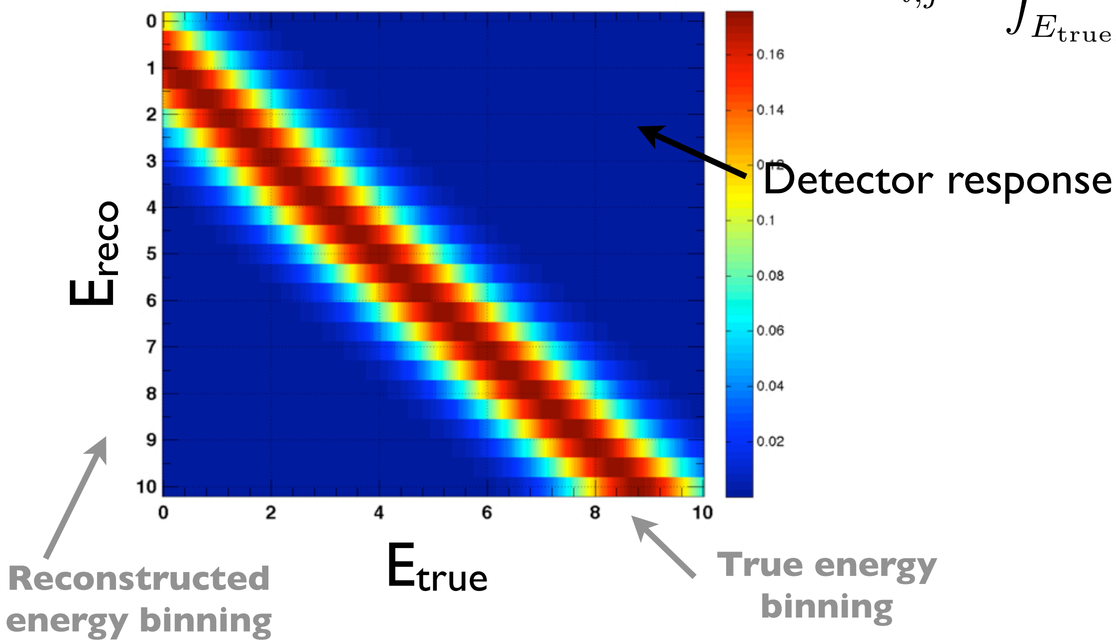
$$\mathrm{Cov}(N_i, N_j) = \langle N_i N_j - \langle N_i N_j \rangle \rangle = \frac{1}{N_{\mathrm{sim}} - 1} \sum_{k=1}^{N_{\mathrm{sim}}} \left(N_i(k) N_j(k) - \frac{1}{N_{\mathrm{sim}}} \sum_{k=1}^{N_{\mathrm{sim}}} N_i(k) N_j(k) \right)$$

- Take into account detector response through a MC simulation for each sample.
- Draw n samples
- Draw deposited energy from incident particle energy taking into account correlations (among energy bins)
- For each deposited energy computes the reconstructed energy through MC simulation.
- Computes the covariances among the reconstructed energy bins

=> Get the covariance matrix in the reconstructed energy bins.

Migration matrix

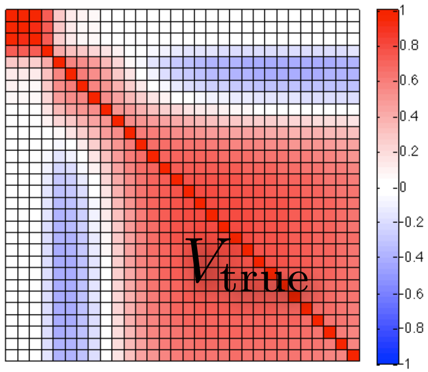
Example of migration matrix

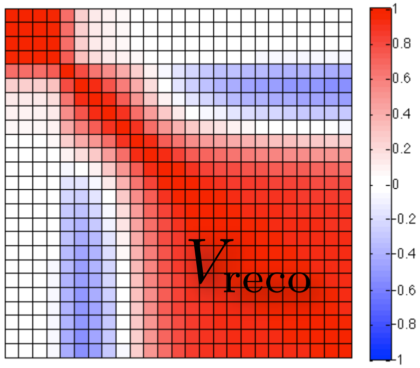


$$M_{i,j} = \int_{E_{\mathrm{true},j}}^{E_{\mathrm{true},j+1}} \int_{E_{\mathrm{reco},i}}^{E_{\mathrm{reco},i+1}} R(E_{\mathrm{T}}, E_{\mathrm{R}}) dE_{\mathrm{R}} dE_{\mathrm{T}}$$

- Models the detector response through a matrix once and for all.
- Do a MC simulation for a given input energy spectrum (no correlations among energy), for instance a flat spectrum (assures same MC statistics whatever the energy).
- For each input energy deposited in the detector the MC computes the reconstructed energy
- Performs a 2D histogram of (E_{true},E_{reco})
- => Gives the migration matrix from a bin j in true energy to a bin i in reconstructed energy $M_{i,j}$

Then, we have $V_{\mathrm{reco}} = M \times V_{\mathrm{true}} \times M'$





Rate Only / Shape Only / Rate and Shape

χ^2 definitions

Rate and Shape

$$\chi^2_{RS} = \sum_i \left(\frac{Y_i - N_i(\vec{\alpha}^\star + \vec{\alpha}, \theta_{13})}{\sigma_i} \right)^2 + \sum_k \left(\frac{\alpha_k}{\sigma_{\alpha_k}} \right)^2$$

Ndof = N-1

Rate Only

$$\chi^2_{RO} = \left(\frac{\sum_i Y_i - N_i(\vec{\alpha}^\star + \vec{\alpha}, \theta_{13})}{\sqrt{\sum_i \sigma_i^2}} \right)^2 + \sum_k \left(\frac{\alpha_k}{\sigma_{\alpha_k}} \right)^2$$

Ndof = 1-1 = 0

$\chi^2_{min} = 0$

Shape Only

$$\chi^2_{SO} = \min_{\mathcal{N}} \sum_i \left(\frac{Y_i - \mathcal{N} \times N_i(\vec{\alpha}^\star + \vec{\alpha}, \theta_{13})}{\sigma_i} \right)^2 + \sum_k \left(\frac{\alpha_k}{\sigma_{\alpha_k}} \right)^2$$

Ndof = N-2

With correlations, in RO case: $\sum_i \sigma_i^2 \longrightarrow \sum_{i,j} V_{i,j}$

Some references

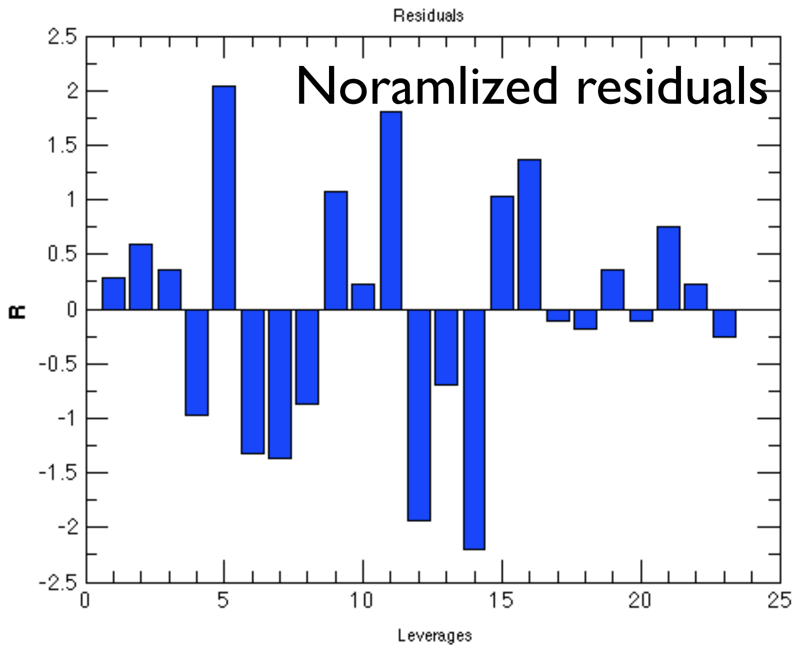
- Kendall, M., Stuart, A., Ord, J.K., Arnold, S. (1999) Kendall's Advanced Theory of Statistics, Volume 2A: Classical Inference and and the Linear Model, Oxford University Press. 6th Edition.
- James, F. (2006) Statistical methods in experimental physics, World Scientific.

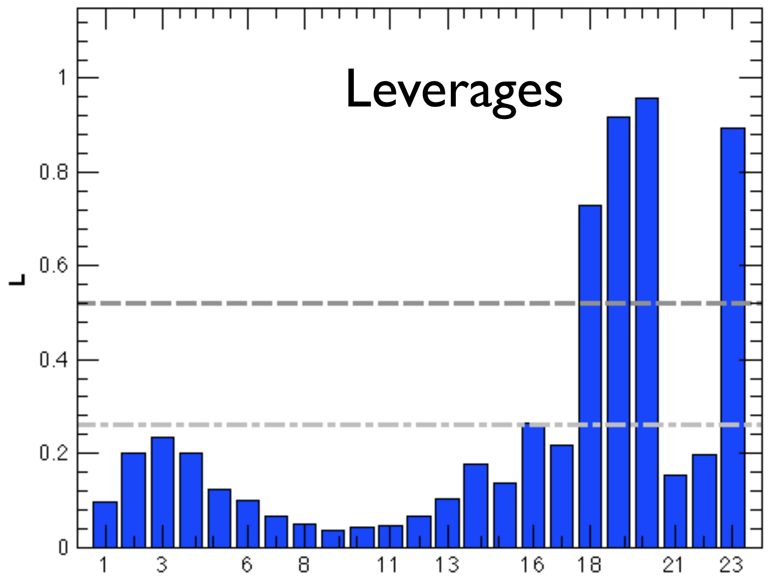
Extra information on χ^2 statistics

Going further into χ^2 analysis: not all the information have the same importance on the results

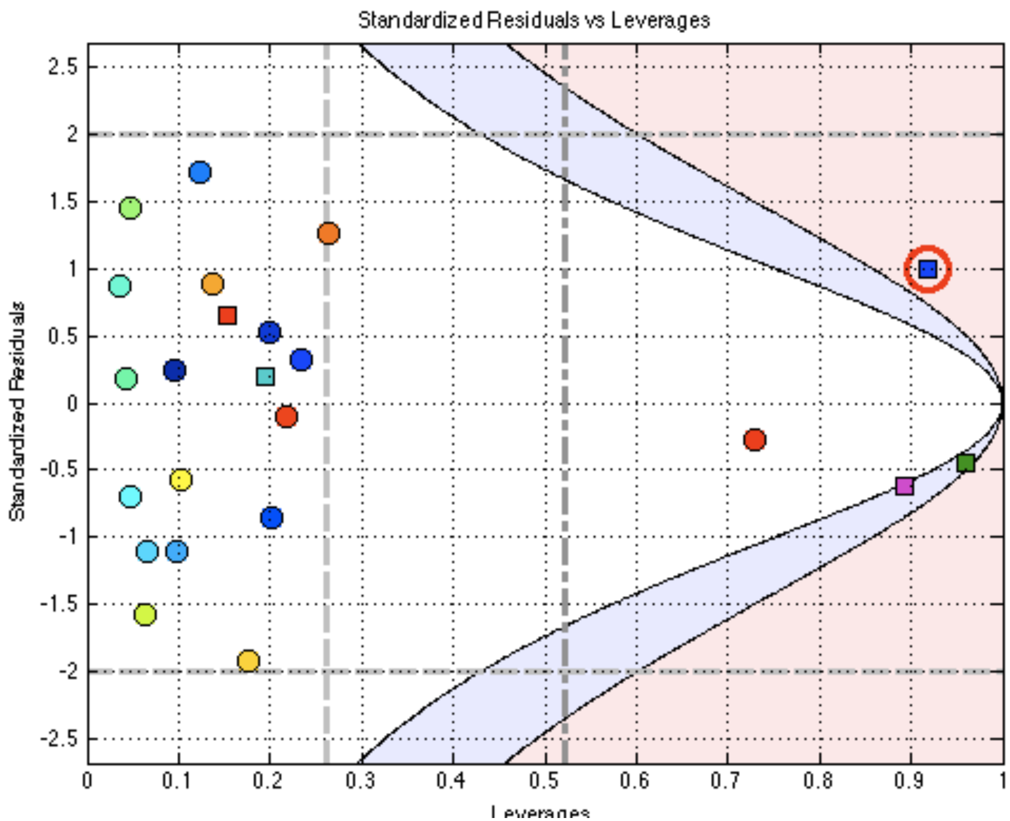
The following slides are just a very brief introduction to diagnostic quantities. References are indicated at the end of this presentation

Diagnostics



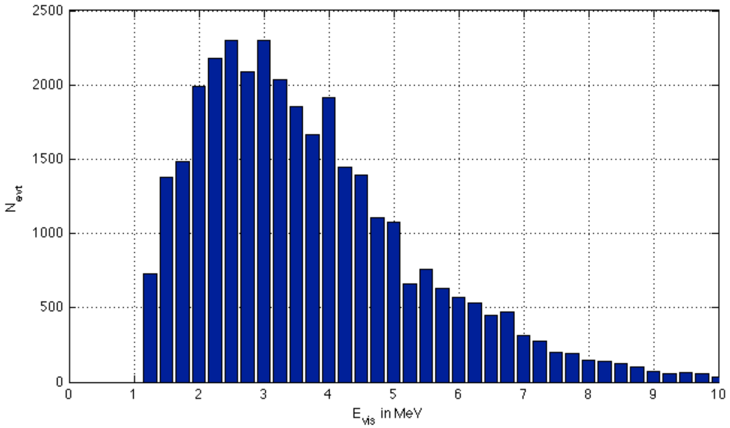


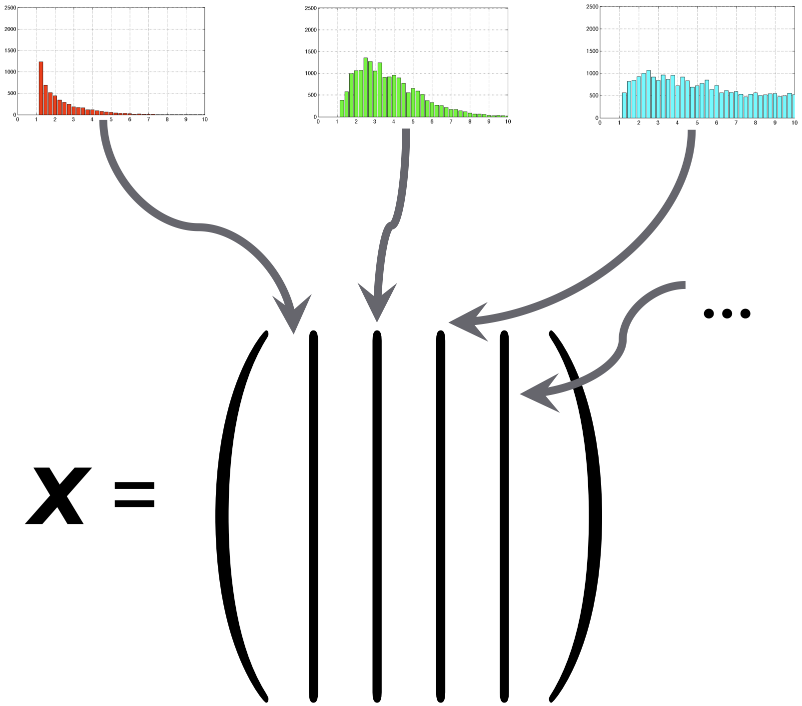
Cook distance



Basic ideas (1/)

Building the model and the χ^2





1. you have some data y_i
2. you assume there exist a model such that $\{y_i\}$ can be described by a linear equation:
$$y = Aa + e$$
where ***A*** is a matrix, ***a*** is the vector of parameters realized in nature, **e** is a vector of errors subsequent to the measurement or the process.
3. you would like to try your model:
$$y = X\alpha + \varepsilon$$
where you put the component you believe responsible of the observations $\{y_i\}$ in ***X***, columnwise (each column is a different component).
4. you estimate the α coefficients with linear regression (χ^2 fit) through minimizing:
$$\chi^2(\alpha) = \|y - f(\alpha)\|^2 = \sum_i (y_i - f_i(\alpha))^2$$
where $f(\alpha) = X\alpha$
For fixed α and ***X***, if **e** follow a normal distribution, then this function behaves like a χ^2 with all the associated properties for interval estmation, statistical tests *etc.*

Basic ideas

Introducing residuals...

So you assume $\{y_i\}$ is drawn from the equation: $y = Aa + e$

where you assume the errors are centered and follow a normal law: E(e) = 0 and V(e) =1.

Then your model is such that you would like to describe the $\{y_i\}$ sample you have: $y = X\alpha + \varepsilon$
where ε is therefore definef by: $\varepsilon = Aa - X\alpha + e$

You want to find the best α which explain the $\{y_i\}$ observations, thus you minimize $\chi^2(\alpha)$
leading to the solution: $\hat{\alpha} = (X'X)^{-1} X'y$

Then the best model is: $\hat{y} = X\hat{\alpha} = \underbrace{X(X'X)^{-1}X'}_{H} y = Hy$

The **residuals** are then defined by $\hat{\varepsilon} = \hat{y} - y = Aa - X\hat{\alpha} + e$

If your model is right, $Aa - X\hat{\alpha}$ should be close to 0 (compared to e). Then in this case $\hat{\varepsilon}$ is an estimate of the initial errors. Thus you should expect to have $E(\hat{\varepsilon}) = 0$ and $V(\hat{\varepsilon}) = 1$ as for **e**.
Of course there are several ways to get this wrong: $E(e) \neq 0$ and/or $V(e) \neq 1$, or non constant V(e), or the ***X*** model is "wrong" such that $Aa - X\hat{\alpha} \nsim 0$

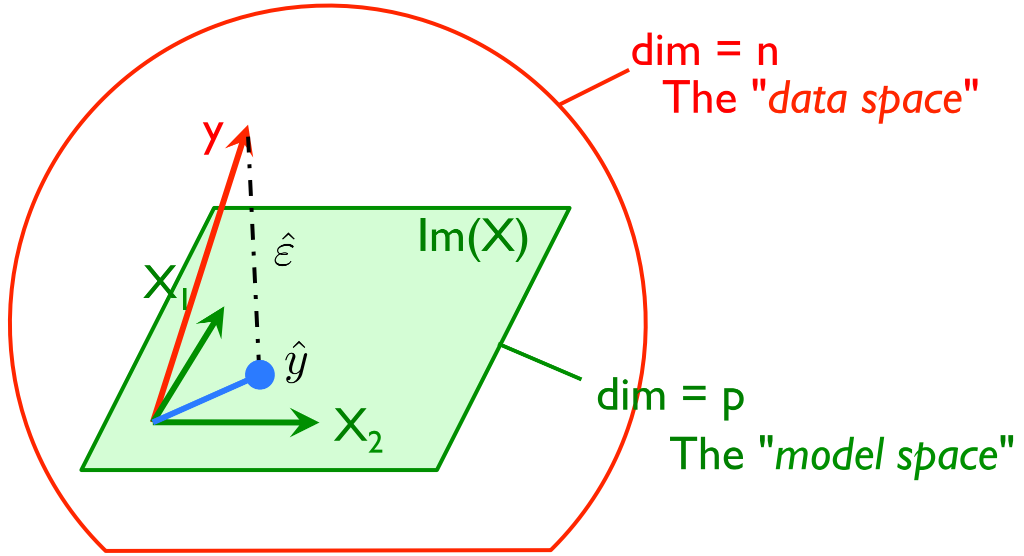
=> ***always inspect the residuals*** to check the validity of the model and the fit!

Basic ideas

...and leverages

When we got the best model: $\hat{y}$ we introduce the ***H*** matrix which is called the "hat matrix" because "it puts a hat" on the y vector: $\hat{y} = X\hat{\alpha} = \underbrace{X\left(X'X\right)^{-1}X'}_{H}y = Hy$

This hat matrix is very important.
It's a projection matrix
(you can check that $H^2 = H$).



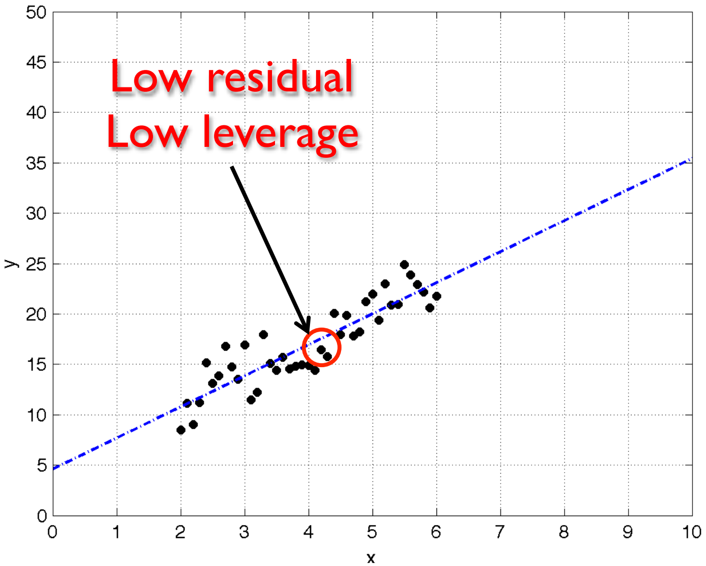
The diagonal coefficient are of upmost importance since they indicate the contribution of the y_i in the best fit model value $\hat{y}_i$

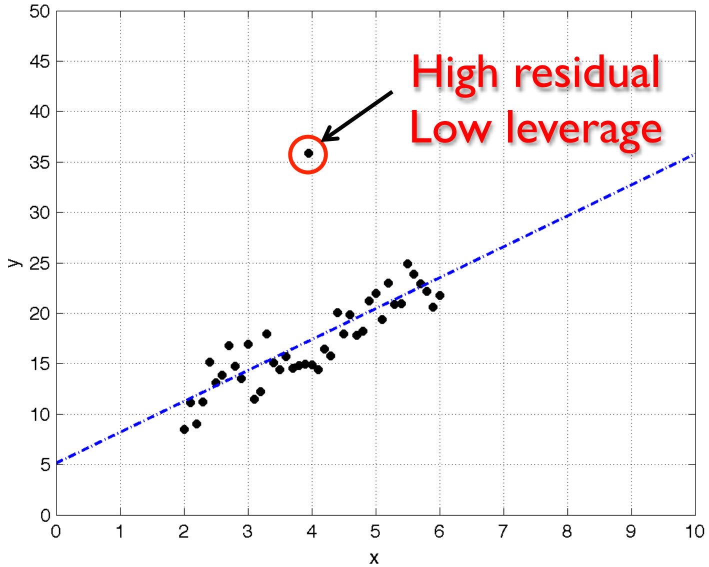
These diagonal coefficients have a particular name: they are called the **leverages**: $h_i = H_{ii}$

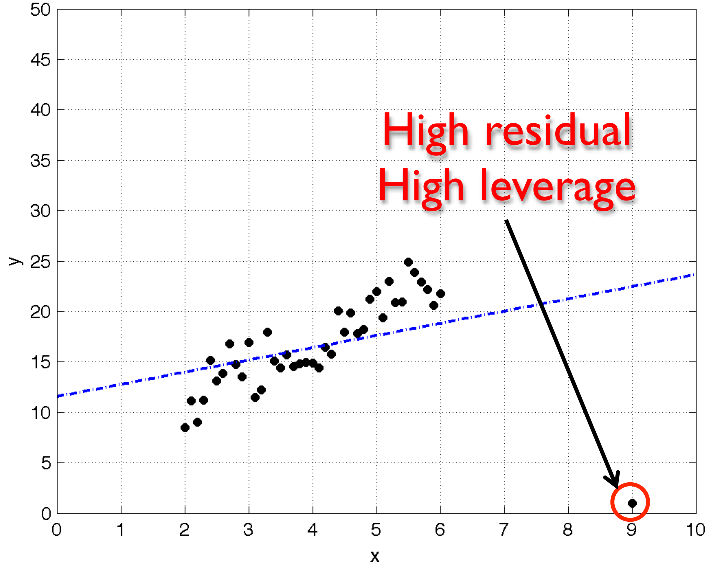
Leverages are linked but complementary to residuals as we will see ($\hat{\varepsilon} = y - \hat{y} = (Id - H)y$)

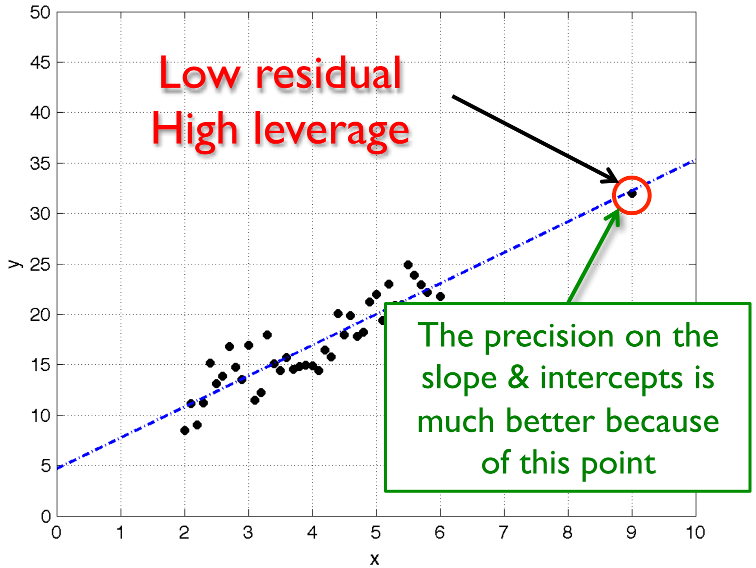
Examining some typical cases in a very simple situation: the straight line fit

○ is a non-influencial data point







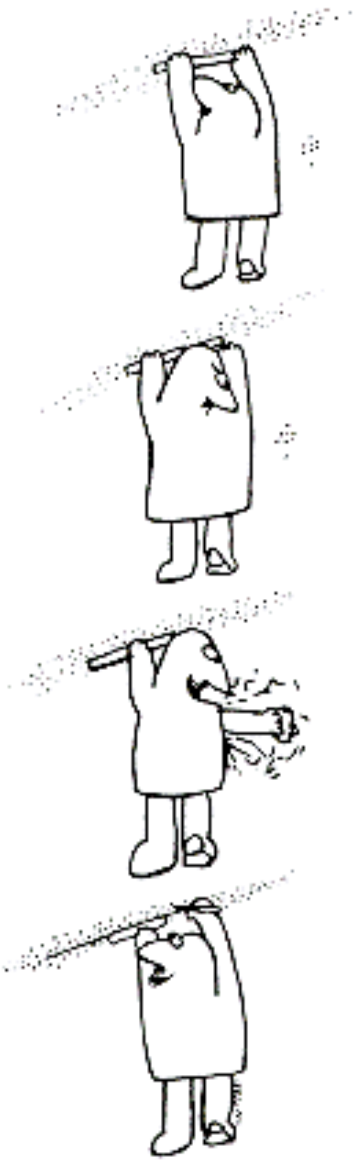


○ is an influencial data point

Diagnostic quantities

- There are a lot of interesting diagnostic quantities on χ^2 regression:
  - Residuals: check normality distribution of residuals
  - Leverages: check influence of each info on fit
  - Cook Distance: a kind of distance with respect to the cloud of data points taking into account all the available info provided
- A lot of deletion diagnostics: deleting one of the data point, what is the influence on the fit?
  - DFBETA: change on best fit param values.
  - COVRATIO: change in uncertainties on best fit params.
  - DFIT: change in model values at best fit
  - Cook Distance can be viewed also as a deletion diagnostic.
- A diagnostic to check the collinearity aspect of the model: if there are many nuisance parameters, it could be that predictive effects are not each other independent or could be closely related. Collinearity among nuisance effects causes instability in the uncertainties and best fit values. The associated diagnostic quantity is called Variance Inflation Factor (VIF).

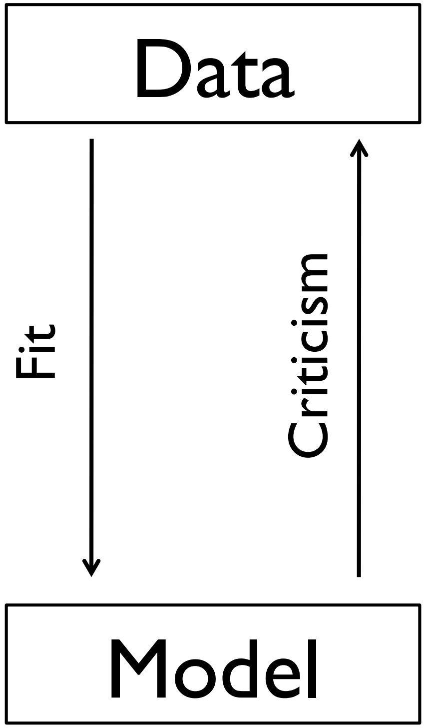
What to do with the diagnostics?



The idea behind is not to remove data points...

The idea is to go deeper in the understanding of the inputs and their impacts on the results: inputs can be data points or systematic bias and uncertainties guess. If results are really depending on a few pieces of information...
=> need to check the inputs again carefully.

(robust studies, robust results)



Some references on diagnostics

- Belsley, D. A., Kuh, E. and Welsch, R. E. (1980) Regression Diagnostics. New York: Wiley.
- Cook, R. D. and Weisberg, S. (1982) Residuals and Influence in Regression. London: Chapman and Hall.
- Williams, D. A. (1987) Generalized linear model diagnostics using the deviance and single case deletions. Applied Statistics 36, 181–191.