Standard analyses of the solar neutrino anomaly

- The **solar** anomaly relies on a combination of many ingredients.

nuclear physics ↘ ↙ statistics

Cl, Ga → $\boxed{\Delta m^2, \theta}$ ← SK, SNO

Solar models ↗ ↖ MSW (sun, earth)

1. **If** all correct: few peculiar $P_{ee}(E_\nu)$ (SMA, LMA, LOW)
   - $\Delta\chi^2$ Gaussian approximation ≈ frequentist FC fit
   - the GOF is significantly lower than usually reported.

But no d/n asymmetry, no spectrum distorsion, no seasonal variation seen so far. Significant MSW effects disfavoured.

2. Which crucial ingredients could be **slightly** wrong?
   - Homestake → $\theta \sim \pi/4$ with large $\Delta m^2_{\rm sun} \gtrsim 10^{-4}\,\mathrm{eV}^2$
   - Solar models → SMA, LMA with smaller θ (disfavored by the new SNO data: I will talk about 1.)

Why improving the statistical analysis?

Starting point of any fit: we know

$$p(\text{data}|\text{theory}) = p(R|\Delta m^2, \theta) = \frac{\exp[-\chi^2/2]}{(2\pi)^{3/2}\sqrt{\det \sigma^2}}$$

but we want $p(\text{theory}|\text{data})$. Two different methodologies:

- **Bayesian**: $p(\Delta m^2, \theta)$ updated as

$$p(\Delta m^2, \theta|R) \propto p(R|\Delta m^2, \theta) p(\Delta m^2, \theta)$$

- **Frequentist** (Neynman, 1937): for each value of $\Delta m^2, \theta$ build *a* 90% range of R. If it contains the measured data, that parameter value is accepted at 90% CL. Build the range starting from highest
  - $p(R|\theta)$ (Crow-Gardner ordering, 1959)
  - $p(R|\theta)/p(R|\theta_{\text{best}}(R))$ (Feldman-Cousins, 1998).

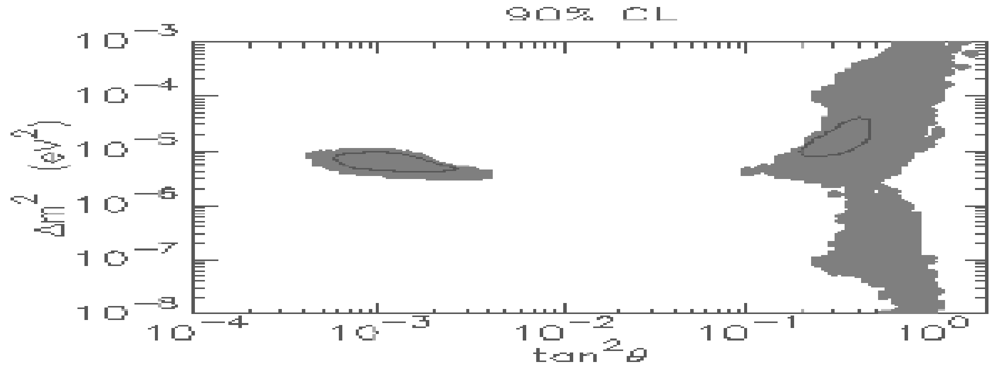
If $p(\text{data}|\text{theory})$ is Gaussian: Bayes = FC = $\Delta\chi^2$

- $p(R|\Delta m^2, \theta)$ is almost Gaussian in $R \sim \sigma\Phi$
- is highly non Gaussian in $\Delta m^2, \theta$. Many solutions LMA, SMA, LOW, VO. A gaussian has only one peak.

This is a minor effect in a Bayesian fit: $95\% \approx 99\%$ anyhow.

Total rates only

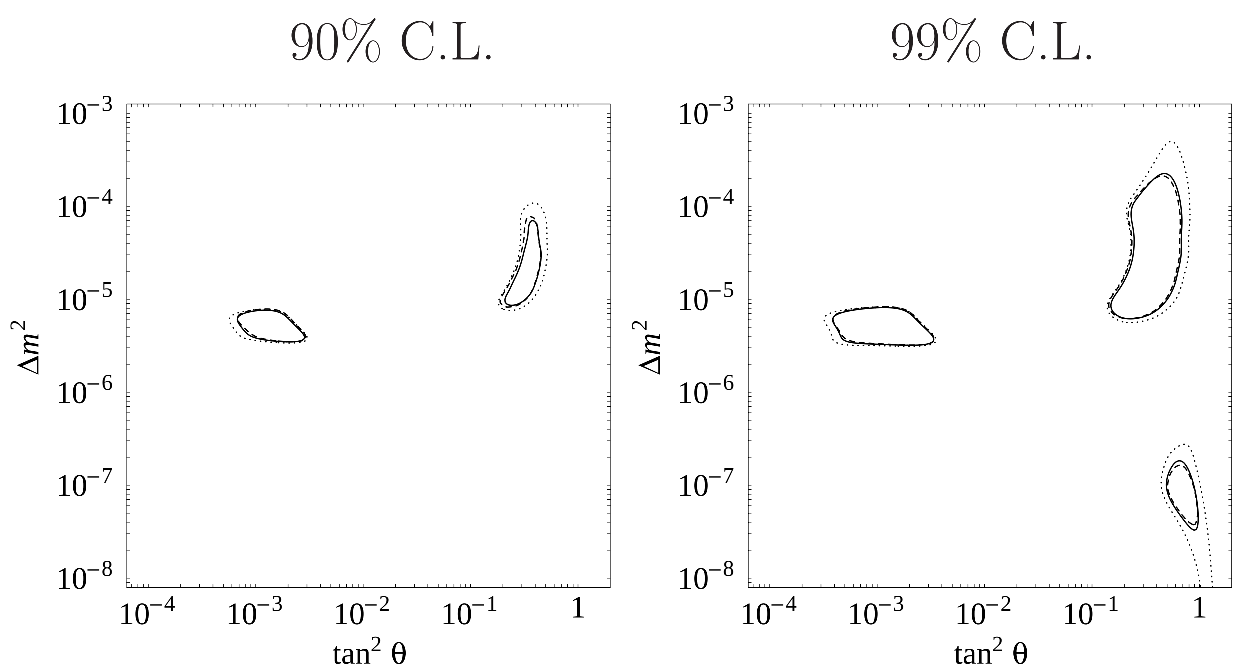
CG fit of data about $\Delta\hat{m}^2$ and $\hat{\theta}$ by Garzelli and Giunti:



FC and CG fit of the three rates. The standard procedure

1. use $\sigma^2 = \sigma^2_{\rm stat} + \sigma^2_{\rm syst} + \sigma^2_{\rm th}$
2. do not ask why

is justified, if Neyman construction is applied in a Bayesian framework.



$$\mathbf{90}\%\,\mathrm{CL} : \Delta\chi^2 = 4.6 \;\;\rightarrow\;\; \mathrm{FC} \;\;\; \Delta\chi^2 = (4.6{\div}5.5) \rightarrow \mathbf{94}\%$$

FC: a fit for each possible exp outcome $\approx 20^3$ fits $\approx$ hours

All data

SK 'gives' the energy spectrum of recoil electrons during the day and during the night (18 + 18 energy bins)

No signal, but significant shift of the allowed regions

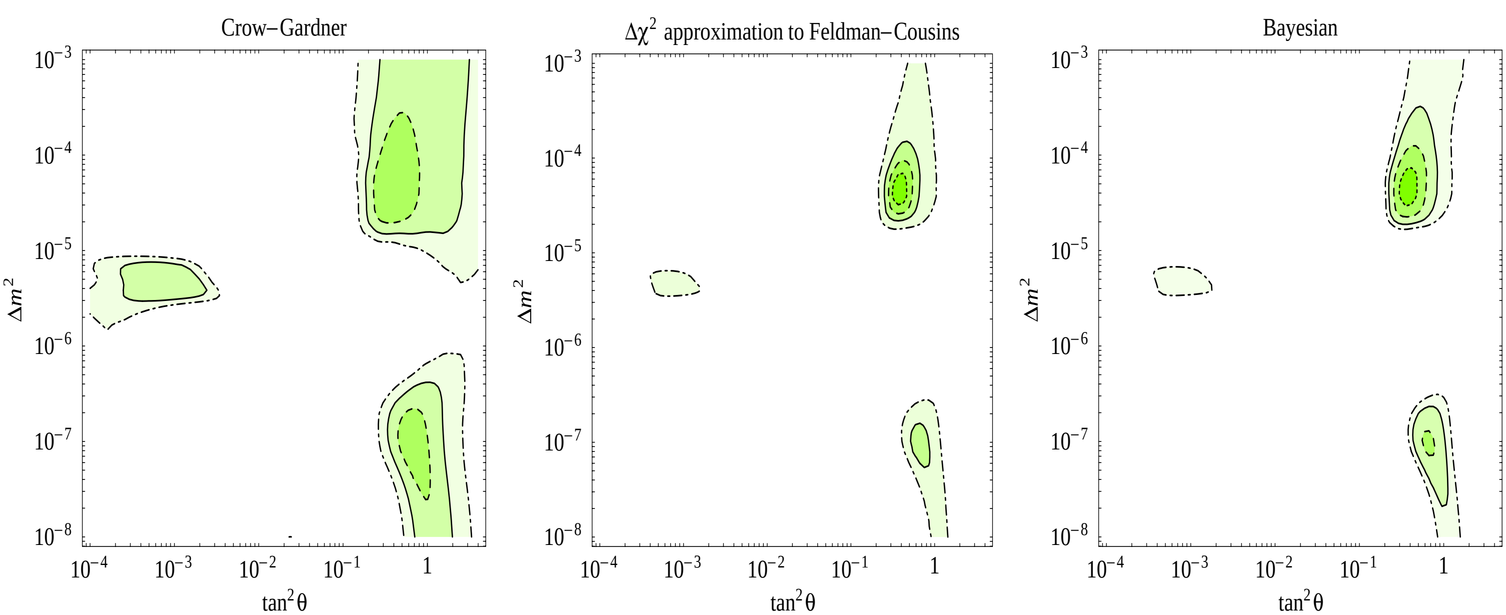
The FC ratio of probabilities

$$p(R|\Delta m^2, \theta)/p(R|\Delta m^2_{\rm best}, \theta_{\rm best}(R))$$

recognizes and eliminates the statistical fluctuations that have nothing to do with the determination of the parameters. With many data this becomes more significant, but also impossible to do numerically: $\approx 20^{38}$ fits $\gg \tau_{\rm proton}$

$$\begin{array}{ll} \text{FC} \approx \Delta\chi^2\text{–cut}: & \chi^2 - \chi^2_{\rm best} \le \chi^2_{2\,\rm dof}(\text{CL}) \\ \text{CG} = \chi^2\text{–cut}: & \chi^2 \le \chi^2_{38\,\rm dof}(\text{CL}) \end{array}$$

CG $\not\approx$ $\Delta\chi^2 \approx$ FC $\approx$ Bayes



Bayesian fit done assuming the 'prior' $dp = d\ln\Delta m^2\, d\ln\tan\theta$.

Goodness-of-fit (naïve)

Fitting only the rates: GOF(LOW) $\approx$ 0.7%.
Adding SK spectral and day/night data (35 bins without any signal) GOF(LOW) increases to 50%

How is this strange result obtained?

Based on a global Pearson χ^2 test with "too many data"

$$\hat{\chi}^2_{\rm rates} \approx 7 \qquad \gg \qquad 1 = 3 \text{ rates} - 2 \text{ parameters}$$
$$\hat{\chi}^2_{\rm global} \approx 7+35 \quad \approx \qquad 36 = 3-2+35$$

$\hat{\chi}^2_{\rm global}$ cannot recognize that there is a problem in the rates

According to $\hat{\chi}^2_{\rm global}$ good fit of sun + atm + LSND with 3ν

What does it mean "too many data"?

1. It is easy to compare two different predictions:

   $$\Delta\chi^2 \equiv \hat{\chi}^2_{\rm th1} - \hat{\chi}^2_{\rm th2}$$

   is distributed as a χ^2 with param$_1-$ param$_2$ dof.

   Irrelevant data can be added to a $\Delta\chi^2$

2. It is more difficult to judge a theory with no competitors. Compare th1 = LOW with th2 = null theory

   $$\Delta\chi^2 = \hat{\chi}^2_{\rm th1} - 0 = \hat{\chi}^2_{\rm LOW}$$

   "*Pearson's $\hat{\chi}^2$ tests the validity of a certain solution with respect to a generic alternative hypothesis, which has a sufficient number of parameters to fit all the data with infinite precision*" means that th2 is defined by the data: adding e.g. data about the ν direction would give a higher GOF for LOW, because now the comparison is done with a th2 that does not know where is the sun

3. **th2 depends on the set of data.** With 18 T_e bins, th2 is **unphsyical**: admits fuzzy energy spectra.

Goodness-of-fit

Include only the data that test the theory

(if you want a useful GOF: $\ll 1$ when there is problem)

18 energy bins give **one** significant new information:
$P_{ee}(E_\nu)$ is flat around $E_0 \sim 10\,\mathrm{MeV}$. To see this, fit

$$P_{ee}(E_\nu) = P_0 + P_0'\left(\frac{E_\nu}{E_0} - 1\right) + \frac{P_0''}{2}\left(\frac{E_\nu}{E_0} - 1\right)^2 + \cdots$$

$$\begin{aligned} P_0 &= 0.45 \pm 0.02 \pm 0.1_{\mathrm{th}} \\ P_0' &= -0.05 \pm 0.1 \\ P_0'' &= 0 \pm 1.5 \end{aligned}$$

large non linearity not present in MSW nor detectable by SK

SK tells nothing about P_0'' and higher derivatives: fit only $P_0' \sim$ rate($T_e < 9\,\mathrm{MeV}$)/rate($9\,\mathrm{MeV} < T_e < 13\,\mathrm{MeV}$).

GOF	Rates only	Naive: rates and spectra	Refined: rates and spectra
SMA	55%	30%	$\approx 2\,\%$
LMA	6%	60%	$\approx 15\,\%$
LOW	0.7%	50%	$\approx 2\,\%$
$P_{ee} = \mathrm{cte}$	0.3%	28%	$\lesssim 1\,\%$

Using another reasonable procedure, also SK finds that now SMA gives a poor fit: the SMA region favoured by total rates falls is excluded by spectral and d/n data at 97% CL.

Inclusion of SNO data

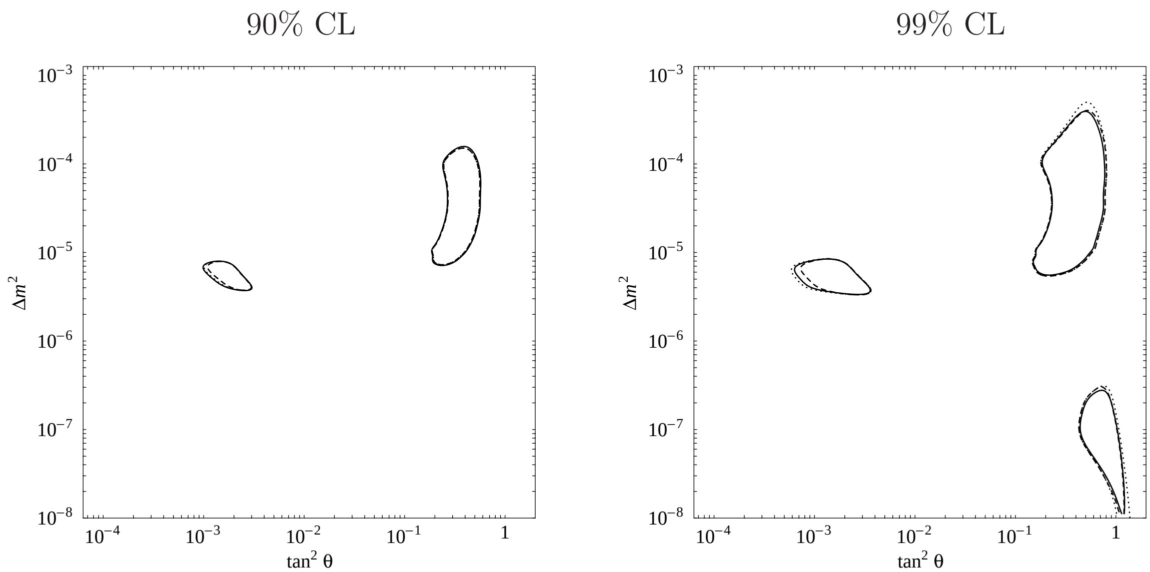


Figure 1: *The 'rates only' fit. Confidence regions at 90% (left) and 99% (right) CL obtained from the four solar rates using three different methods: the $\Delta\chi^2$ approximation (continuous line), the Feldman–Cousins procedure (dashed line) and the Crow–Gardner procedure (dotted).*

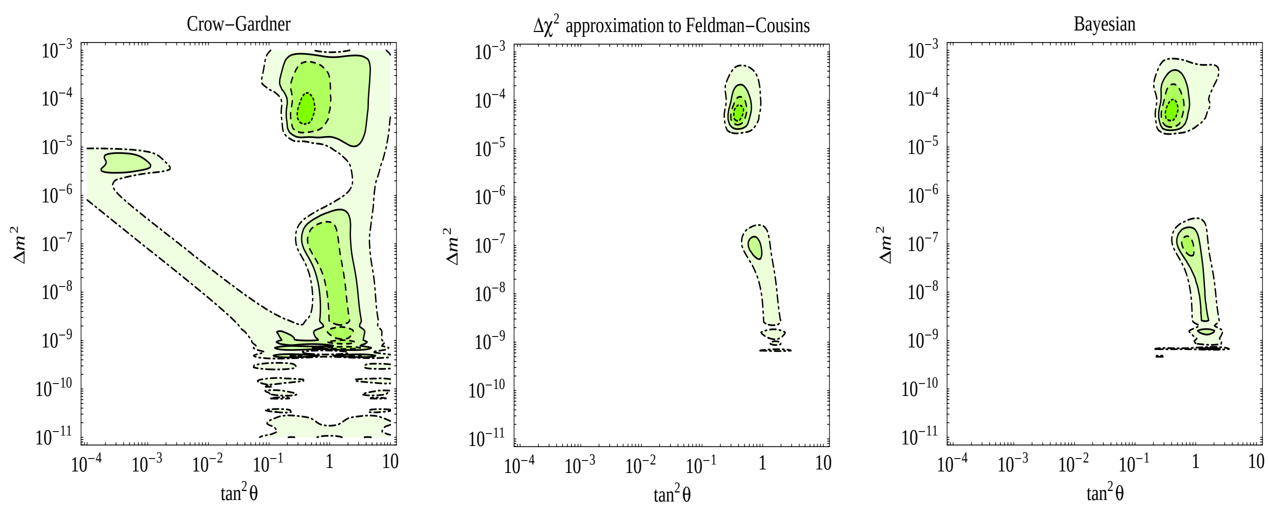


Figure 2: *The 'global' fit.*