Received: *March 5, 2001*, Accepted: *May 25, 2001*
Revised: *April 12, 2001*

Frequentist analyses of solar neutrino data

Paolo Creminelli and Giovanni Signorelli

Scuola Normale Superiore and
INFN Sezione di Pisa, Italy
E-mail: creminelli@cibs.sns.it, Giovanni.Signorelli@sns.it

Alessandro Strumia

CERN, Geneva, Switzerland and
Dipartimento di Fisica dell'Università di Pisa and INFN
E-mail: Alessandro.Strumia@cern.ch

Abstract: The solar neutrino data are analysed in a frequentist framework, using the Crow-Gardner and Feldman-Cousins prescriptions for the construction of confidence regions. Including in the fit only the total rates measured by the various experiments, both methods give results similar to the commonly used $\Delta\chi^2$-cut approximation. When fitting the full data set, the $\Delta\chi^2$-cut still gives a good approximation of the Feldman-Cousins regions. However, a careful statistical analysis significantly reduces the goodness-of-fit of the SMA and LOW solutions.

Keywords: Standard Model, Beyond Standard Model, Neutrino Physics.

JHEP05(2001)052

Contents

1. Introduction **1**

2. Different frequentist analyses: rates only **3**
- 2.1 Feldman-Cousins fit 5
- 2.2 Crow-Gardner fit 7

3. The inclusion of the whole data set **9**

4. The goodness of fit **11**

5. Conclusions **13**

A. Details of the computation **14**

1. Introduction

The solar neutrino anomaly is an old but still controversial problem, in which many experimental data [1]–[4] and theoretical ingredients [5]–[9] have to be merged to give predictions for the oscillation parameters and to rule out other non-standard explanation of the anomaly. A correct statistical treatment is a necessary step of the analysis. The starting point for interpreting the results of an experiment is the fact that one knows the probability distribution $p(\text{data}|\text{theory})$ for obtaining a set of data under the assumption that a given theory is true. In the case of the solar neutrino anomaly (at least in its simplest version), we know $p(R_i|\Delta m^2, \theta)$, where R_i are the three neutrino rates measured in Chlorine, Gallium and SK experiments, which should be used to infer the values of the theoretical parameters Δm^2 and θ.

This can be done according to two conceptually very different approaches [10], each one with unsatisfactory aspects.

- **The bayesian approach** employs a probability distribution $p(\Delta m^2, \theta)$ to summarize our knowledge of the parameters of the theory. According to elementary properties of probability, this probability gets updated by the inclusion of the results of a new experiment as $p(\Delta m^2, \theta|R) \propto p(R|\Delta m^2, \theta)p(\Delta m^2, \theta)$. The drawback is that one needs to choose some "prior" $p(\Delta m^2, \theta)$ to start with, and the final result depends on this choice until experiments are sufficiently precise. At the moment, solar neutrino fits give multiple distinct solutions so that $p(\Delta m^2, \theta|R)$ still contains arbitrary order 1 factors. The advantage is its extreme simplicity: the laws of probability dictate what to do in any situation.

- **The frequentist approach** refuses the concept of probability of theoretical parameters. The Neyman construction [11] allows us to build range of parameters for any possible outcome of an experiment with the property that 90% (or whatever) of such ranges contain the true value. However this procedure is not univocal and the resulting regions can be quite different. For example the Crow-Gardner [12] procedure gives smaller regions in presence of unlikely statistical fluctuations in the measured outcome of the experiment, while the Feldman-Cousins [13] procedure gives ranges of roughly the same size for all possible outcomes.

In simple cases when $p(\text{data}|\text{theory})$ is a gaussian function of all its arguments (data and parameters, with no physical constraints on them), the bayesian approach (using a flat prior p) and the frequentist approach (using the Feldman-Cousins method) are numerically equivalent to the commonly employed $\Delta\chi^2$-cut approximation.

When fitting solar neutrino data one has to be careful because:

(1) $p(R_i|\Delta m^2, \theta)$ is a highly non-gaussian function of $\Delta m^2, \theta$: in fact one finds a few separate best-fit solutions (usually named "LMA", "SMA", "LOW", "VO") while a gaussian would have only one peak. This is the problem that we will address in this paper.

(2) $p(R_i|\Delta m^2, \theta)$ is not perfectly gaussian as a function of R_i. Assuming a gaussian uncertainty on the detection cross sections σ and on the solar fluxes Φ, one does not obtain a gaussian uncertainty on the rates $R \sim \sigma \cdot \Phi$. In principle this is true; in practice the errors on σ and Φ are sufficiently small that their product is also almost gaussian, up to very good accuracy.

Such issues have been studied in [14], finding that (1) apparently has a dramatic effect (see [14, figure 3]: LMA and LOW merge in a single region), while (2) has a negligible effect (see [14, table II]). We will ignore (2) and we therefore write the probability density function (pdf) for all the n solar neutrino data x_i as

$$p(\boldsymbol{x}|\Delta m^2, \theta) = \frac{\exp[-\chi^2/2]}{(2\pi)^{n/2}\sqrt{\det \sigma^2}}, \qquad \chi^2 \equiv \sum_{i,j=1}^{n} (x_i^{\rm exp} - x_i^{\rm th}) \frac{1}{\sigma_{ij}^2} (x_j^{\rm exp} - x_j^{\rm th}) \,. \quad (1.1)$$

The predicted values $x^{\rm th}$ and the covariance matrix σ^2 depend on Δm^2 and θ. The covariance matrix contains both theoretical and experimental errors, statistical and systematic, added in quadrature. This is the standard procedure, which can be justified by applying the Neyman construction in a bayesian framework (i.e. by describing

JHEP05(2001)052

theoretical and systematic uncertainties using a probability distribution). A strict frequentist framework employs a definition of probability that makes it unclear how to deal with systematic and theoretical uncertainties.

Using the analytical properties of gaussians enormously simplifies the computation: we will not need lengthy and obscure computer calculations. The probability p is computed as described in appendix A. We will study oscillations among the three active neutrinos in a two flavour setup. We could study much more general cases, but experiments indicate that this seems to be the relevant case.[1] $\Delta m^2 \equiv m_2^2 - m_1^2 > 0$ is the squared mass difference relevant to solar neutrinos, and $0 \leq \theta \leq \pi/2$ is the corresponding mixing angle.

In section 2 we fit the data about the total rates using the Crow-Gardner and Feldman-Cousins constructions, which are compared with the commonly used $\Delta\chi^2$-cut approximation. We do not find dramatic differences (see figure 1). A fit based on the $\Delta\chi^2$ approximation does not miss any relevant physical issue. In section 3 we include in the fit the SK spectral and day/night data. We now find more marked differences between the various methods for building Neyman's confidence regions (see figure 4). In section 4 we show that the well-known statement that LMA, LOW and SMA presently give a good fit is based on an inappropriate statistical test, and we recompute the goodness-of-fit of the various solutions (see table 1). Our conclusions are drawn in section 5.

2. Different frequentist analyses: rates only

We want to compare exact and approximate methods to compute confidence regions. To begin with, we consider only the total rates measured at Homestake, SuperKamiokande and the weighted sum of the two Gallium experiments: GALLEX-GNO and SAGE. All fits done so far (except [14]) use the approximated method based on the $\Delta\chi^2$-cut; this approximation will be compared with two frequentist constructions: the Crow-Gardner [12] and Feldman-Cousins [13] methods.

The use of the $\Delta\chi^2$-cut is based on the well-known *likelihood ratio theorem* [20], which states: given a conditional pdf $p(\boldsymbol{x}|\boldsymbol{m})$ ($\boldsymbol{x}$ is the data vector and $\boldsymbol{m}$ are the parameters we want to estimate) with a range for $\boldsymbol{x}$ independent from the value of $\boldsymbol{m}$, the quantity

$$\lambda(\boldsymbol{x};\boldsymbol{m}) = 2\log\left(\frac{\max_{\boldsymbol{m}} p(\boldsymbol{x}|\boldsymbol{m})}{p(\boldsymbol{x}|\boldsymbol{m})}\right) \tag{2.1}$$

is distributed as a χ^2 with $\dim(\boldsymbol{m})$ degrees of freedom (dof), independently from the

[1] In a 3 ν framework, the ν_e can also oscillate at the atmospheric Δm^2. The CHOOZ bound [15] implies that the relative mixing angle is so small that it can only have a minor effect on solar neutrinos. The LSND anomaly [16] motivates models with a fourth sterile neutrino. However LSND is significantly constrained directly by Bugey [17] and Karmen [18], and indirectly by SK [2, 19] that disfavours a significant sterile contribution to both the atmospheric and solar anomalies. These indirect bounds can be evaded in models with many sterile neutrinos.

JHEP05(2001)052

value of $\boldsymbol{m}$, in the limit $\dim(\boldsymbol{x}) \to \infty$. With the pdf of eq. (1.1) this leads to

$$\lambda(\boldsymbol{x}; \Delta m^2, \theta) = \chi^2 - \chi^2_{\text{best}} + \log\det\sigma^2 - \log\det\sigma^2_{\text{best}}\,, \tag{2.2}$$

where χ^2 is the usual sum in the exponent of eq. (1.1), σ^2 is the covariance matrix and the subscript "best" indicates that the corresponding quantity must be evaluated at the value of $\Delta m^2, \theta$ that maximizes the probability for the given measured $\boldsymbol{x}$. In the limit of infinite data, λ is distributed as a χ^2 with two degrees of freedom (the two parameters we are studying: Δm^2 and θ. If $10^{-3}\,\text{eV}^2 \lesssim \Delta m^2 \lesssim 10^{-4}\,\text{eV}^2$ one can obtain poor fits with energy independent survival probability. In this case the experimental results only depend on the single parameter θ).

The simplest way to construct confidence regions using this asymptotic property is to include all values of $(\Delta m^2, \theta)$ for which λ is less than a critical value, which can be obtained from the χ^2 distribution tables. Neglecting the $\ln\det\sigma^2$ term — which is not a constant — gives the well-known approximate rule

$$\Delta\chi^2 < \beta\,, \tag{2.3}$$

where β depends on the confidence level (CL) we want to quote. This is the method most analyses use to obtain the confidence regions. It is twofold approximate: it neglects the $\log\det\sigma^2$ dependence and, since the number of solar data is finite, does not ensure the correct CL.

The correct frequentist construction of the confidence regions is a well-known procedure: for any point $\boldsymbol{m}$ in the parameter space (Δm^2 and θ, in our case) one has to *arbitrarily* choose a region $\mathcal{A}(\boldsymbol{m})$ in the space $\{\boldsymbol{x}\}$ of the data (the three rates, in our case) which contains the CL% of the probability. The knowledge of the pdf (1.1) allows to do this. The confidence region $\mathcal{B}(\boldsymbol{x})$ is given by all the points in the parameter space that contains the measured value of the experimental data in their acceptance region:

$$\mathcal{B}(\boldsymbol{x}) = \{\boldsymbol{m}|\boldsymbol{x} \in \mathcal{A}(\boldsymbol{m})\}\,. \tag{2.4}$$

It is easy to realize that, taken any *true* value for the parameters, the quoted region contains this true value in the CL% of the cases.

The arbitrariness in the choice of the acceptance region $\mathcal{A}(\boldsymbol{m})$ can be fixed by choosing a particular ordering in the data space: the construction of $\mathcal{A}(\boldsymbol{m})$ is made adding $\boldsymbol{x}$-cells in that order until the requested probability (coverage) is reached. Crow and Gardner (CG) proposed [12] an ordering based on $p(\boldsymbol{x}|\boldsymbol{m})$, while Feldman and Cousins (FC) proposed [13][2] an ordering based on the likelihood ratio

$$\frac{p(\boldsymbol{x}|\boldsymbol{m})}{p(\boldsymbol{x}|\boldsymbol{m}_{\text{best}}(\boldsymbol{x}))}\,. \tag{2.5}$$

[2]The Feldman-Cousins method supplements Neyman's prescription in a way that solves some of the problems that characterize other methods. Some unwanted and strange properties are however still present, and the proposed way out [21] goes beyond Neyman's construction, leading, always in a completely frequentist approach, to the definition of a new quantity called Strong Confidence Level.

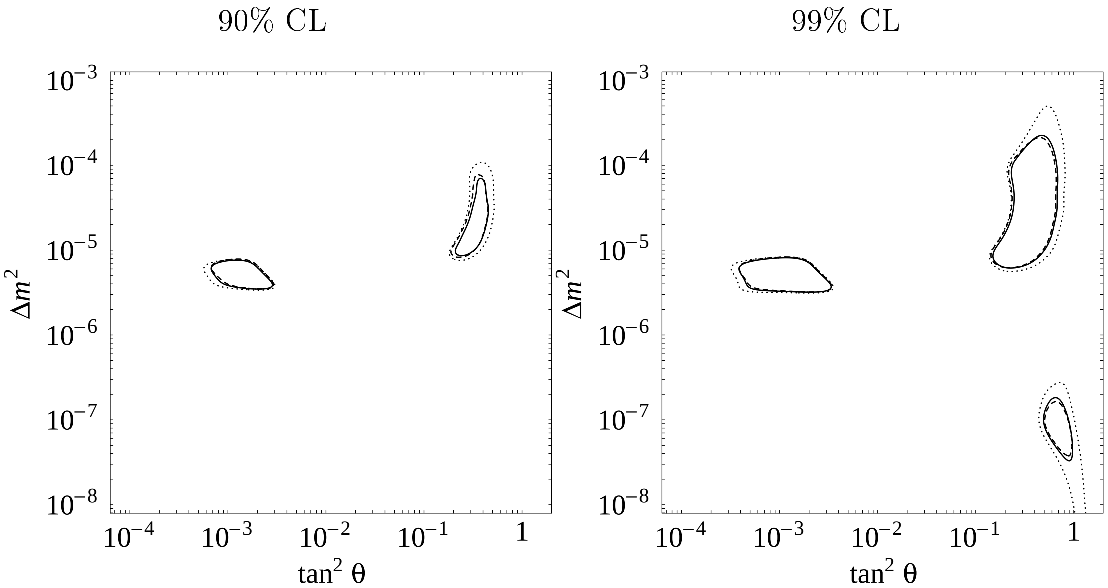

Figure 1: Confidence regions at 90% (left) and 99% (right) CL obtained from the measured solar rates using three different methods. The smallest regions (continuous line) are obtained with the $\Delta\chi^2$ approximation; they are surrounded by the Feldman-Cousins regions (dashed line). The largest regions (dotted line) are obtained with the Crow-Gardner procedure.

This means that the "priority" of a point $\boldsymbol{x}$ for fixed $\boldsymbol{m}$ is given by its probability relative to the probability obtained with the parameters set at the best-fit value $\boldsymbol{m}_{\text{best}}(\boldsymbol{x})$ corresponding to $\boldsymbol{x}$.

2.1 Feldman-Cousins fit

The FC ordering requires cumbersome numerical computations, but guarantees that the FC acceptance regions share the nice properties of the approximate $\Delta\chi^2$-cut method. The FC ordering disregards the statistical fluctuations with no information on the parameters. If the measured rates are unlikely for any value of the parameters, the FC procedure "renormalizes" the probability when determining the confidence regions. It is easy to see that the FC acceptance regions are never empty for any choice of the confidence level: every point in the data space belongs at least to the FC acceptance region of the parameter $\boldsymbol{m}_{\text{best}}$ that maximizes its probability.[3]

The FC procedure has many points in common with the approximate $\Delta\chi^2$-cut method: looking at eqs. (2.1)–(2.3) at fixed $\boldsymbol{m}$ we see that the inequality $\lambda < \beta$ chooses $\mathcal{A}(\boldsymbol{m})$ with the same ordering as the FC method. The only difference is that

[3]The Feldman-Cousins procedure gives no empty confidence regions if all the points with the same likelihood ratio are included in the acceptance region for given parameters, even after the given probability is reached: this can give a certain *overcovering*, but is essential to get this good property. Some pathological situations are still possible if the best-fit point does not exist [21].

the limit β is chosen, using the asymptotic distribution of λ, independent of $\boldsymbol{m}$, while the exact method gives a limit $\beta(\boldsymbol{m})$ that depends on the oscillation parameters $\boldsymbol{m}$.

It is easy to check that the $\Delta\chi^2$-cut is exactly equivalent to the FC construction if the pdf is gaussian with constant covariance matrix and with theoretical rates that depend linearly on the parameters (by "theoretical rates" we indicate the most likely value of the rates, for given values of Δm^2 and θ). In the linear approximation the theoretical rates, obtained varying the two parameters Δm^2 and θ, form a plane in the three-dimensional space of the rates.[4] The comparison with this linear approximation helps us to understand whether the $\Delta\chi^2$-cut is a good approximation. Two different behaviors are possible for a given $\boldsymbol{m}$:

- The value of $\beta(\boldsymbol{m})$ given by the FC procedure is smaller than the approximated one derived by the $\Delta\chi^2$-cut. This happens, for example, if we measure values of the parameters near the edges of the parameter space. The $\Delta\chi^2$ approximation assumes an infinite hyperplane of theoretical rates: the points of the data space that have maximal probability for "non-physical" values of the parameters will be included in the acceptance regions of the points near the edge, reducing their limit β to have a correct coverage.

- The value of $\beta(\boldsymbol{m})$ is larger than the approximated one. This happens when different regions of the parameter space give similar predictions for the data. A data point that is included in the acceptance region of a given $\boldsymbol{m}$ in the linear approximation may have a bigger $p(\boldsymbol{x}|\boldsymbol{m}_{\rm best})$ because of the folding of the hypersurface, which lowers its likelihood ratio. To reach the requested probability a bigger value for β is needed.

Within the LMA, LOW and SMA regions the linear approximation is pretty good, as the curvature of the "theoretical surface" is small with respect to the typical errors on the rates. The effects due to the edges of the theoretical surface can also be neglected. *The main deviation from the $\Delta\chi^2$ approximation is due to the fact that SMA, LMA, LOW points give similar predictions*: the surface of theoretical rates is folded. Constructing the acceptance region for an oscillation parameter $\boldsymbol{m}$ in the SMA region, we find points with best-fit parameters in the LMA region, so that the $\Delta\chi^2$ approximation is expected to give some undercoverage. The FC acceptance regions at 90% and 99% CL are plotted in figure 1: we see that the regions obtained from the $\Delta\chi^2$-cut are smaller than the exact FC regions. For example, the approximated $\Delta\chi^2$ cut at 90% CL is $\Delta\chi^2 < 4.6$. The value of $\beta(\boldsymbol{m})$ at 90% CL obtained with the FC construction is $\approx (4.6 \div 5.5)$ for $\boldsymbol{m}$ in the SMA region and $\approx (4.8 \div 5.5)$ for $\boldsymbol{m}$ in the LMA region. Furthermore, owing to the variation of

[4] It is interesting to note that even if a fundamental property of frequentist inference is its independence from the metric and topology of the parameter space, the validity of the $\Delta\chi^2$-cut approximation strongly depends on them.

JHEP05(2001)052

$\log\det\sigma^2$ (neglected by the $\Delta\chi^2$ approximation), in the SMA and LOW regions the FC boundary intersects the $\Delta\chi^2$ boundary, instead of surrounding it. The difference between the approximate and rigorous methods is however small enough to justify the $\Delta\chi^2$ approximation.

2.2 Crow-Gardner fit

A second way to construct confidence regions is based on the CG ordering: the acceptance regions are built beginning from the points of highest probability. Such ordering is not invariant under a reparametrization of the experimental data (e.g. a CG fit of the rates is different from a CG fit of the squared rates). Such acceptance regions are the smallest with the given coverage. The difference between the FC and the CG procedures is essential for those points that are unlikely for any value of the parameters, i.e. all the points far from the surface of theoretical rates. We have seen that, with the FC ordering, every data point is included in the acceptance region of at least one point in the parameter space, but this is obviously not true for the CG method. Consider for example the linear approximation in which the theoretical rates describe a plane in the rate space. For a given $\boldsymbol{m}$, in view of (1.1), the CG acceptance region will be an ellipsoid centered in the most likely value for the rates, which becomes larger and larger with growing CL. The FC acceptance regions will be very different and stretched to infinity in a cylindrical shape perpendicularly to the plane (see figure 2). This is clear if we consider that in this approximation the maximum likelihood point $\boldsymbol{m}_{\rm best}$ is obtained by projecting a data point on the "theoretical plane" (in the base where the covariance matrix is proportional to the identity): all the points lying on a line perpendicular to this plane and intersecting

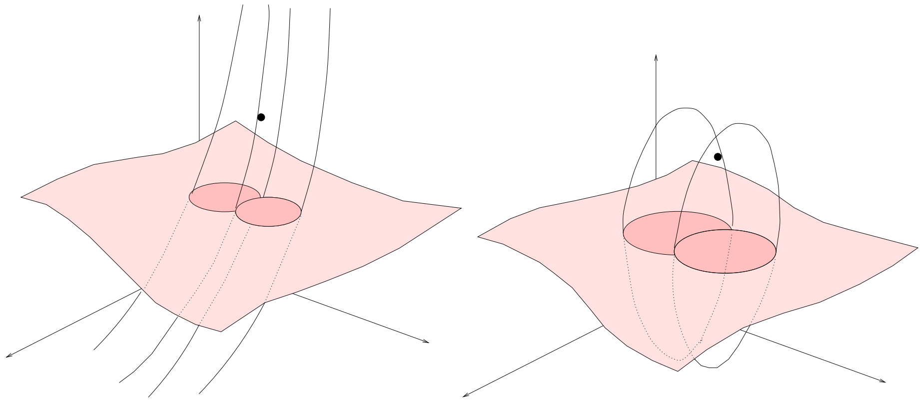

Figure 2: Approximate shapes of the acceptance regions for two near parameter points in the FC case (left) and in the CG case (right). In the first case the regions are stretched perpendicularly to the theoretical surface, while in the second they are ellipsoidal. A parameter point is accepted if the measured experimental point (black dot in the figures) lies inside its acceptance region.

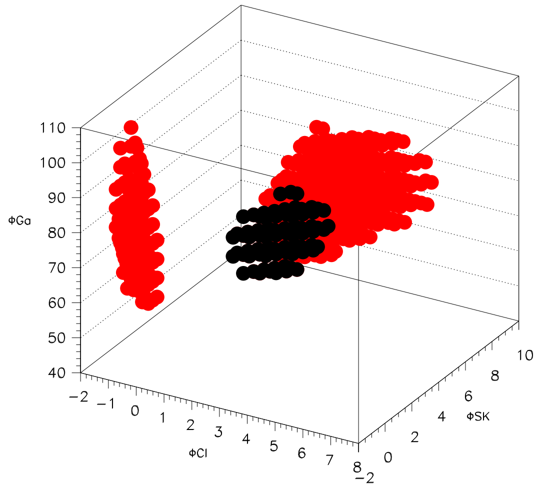

Figure 3: Comparison between the CG acceptance region (black) and the FC one (red) at 90% CL for a parameter point in the SMA region. Chlorine and Gallium rates are in SNU, while the SK rate is in $10^6\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The presence of LMA points with comparable predictions makes the FC region asymmetric and disconnected.

it in the point described by $\boldsymbol{m}$ have likelihood ratio equal to 1 and are included in the acceptance region.

In figure 3 we compare the FC and CG acceptance regions for one given SMA oscillation. We see that the CG region is ellipsoidal as expected, while the FC one is stretched, but in one direction only. This is due to the strong non linearity of the "theoretical surface": LMA and SMA have similar rate predictions even if they have very different parameters. The surface is folded and this causes the asymmetry in the FC region: the acceptance region is deformed and disconnected to get far from the LMA predictions.

The CG acceptance regions at 90% and 99% CL are plotted in figure 1. The differences between the FC and CG confidence regions are readily understood. If we fix the experimental data and begin from a very low CL, we expect an empty CG region (there is no ellipsoid that contains the data) while the FC region is small but non empty. As the CL increases, a CG region appears (roughly at 4% CL in our case) and all the regions grow. With a large CL we expect the CG regions to be larger than the FC ones as the ellipsoids have a larger projection on the "theoretical surface" than the stretched FC acceptance regions (see figure 2). All these features can be checked explicitly in figure 1.

In conclusion two points must be stressed: first of all we have checked that a correct frequentist approach gives results only slightly different from the naive analysis based on the $\Delta\chi^2$-cut. This is apparently in contrast to what is obtained in [14]. The main difference between that analysis and ours is that we use all three rates to construct confidence regions, while [14] finds, with a Monte Carlo simulation, the distribution of the maximum likelihood estimators $\Delta\hat{m}^2$ and $\hat{\theta}$ and construct from

JHEP05(2001)052

this the confidence regions, using the Crow-Gardner ordering. Since the two $\Delta\hat{m}^2$, $\hat{\theta}$ are not a *sufficient statistics* for the three rates, this procedure implies a certain loss of information, which leads to larger confidence regions.

A second point is the comparison between the two methods, CG and FC: the results are pretty similar. As we will see in the next section, this is no longer the case when the SK data on the angular and energy distribution are included in the fit. We have not shown "vacuum oscillation" fits of the solar rates because they are strongly disfavoured by the SK data.

3. The inclusion of the whole data set

The SuperKamiokande collaboration has also measured the energy spectrum of the recoil electrons as a function of the zenith-angle position of the sun. The full data set usually employed in solar neutrino fits contains 38 independent dof (see appendix A). With such a number of data it is practically impossible to perform a complete numerical construction of the acceptance regions, without any approximation. For example, even if we divide every dimension of the data space in only 20 cells, we arrive to a 38-dimensional space divided into $\sim 10^{50}$ cells. For this reason we cannot construct the FC confidence regions with all the data set. However, for the same reasons described in the previous section, the approximated $\Delta\chi^2$-cut method is expected to be a reasonable approximation of the FC construction and to give confidence regions slightly smaller than the FC ones.

For the CG ordering the situation is better. Since we have approximated the pdf (1.1) as a gaussian function of the data, the CG construction is equivalent to a cut on the χ^2 with 38 dof (rather than on the $\Delta\chi^2$). Note that in this case the procedure is exact even if $\log\det\sigma^2$ is not constant. For any $\boldsymbol{m}$ the χ^2-cut defines an ellipsoidal acceptance region $\mathcal{A}(\boldsymbol{m})$, and the confidence region is given by the set of parameters with χ^2 smaller than a given value. The comparison between the CG ordering and the approximated $\Delta\chi^2$-cut can be done analytically: for any given value of CL

$$\begin{aligned} \text{FC} \approx \Delta\chi^2\text{−cut}: &\qquad \chi^2(\Delta m^2;\theta) - \chi^2_{\text{best}} \leq \text{Quantile}(\chi^2_{2\,\text{dof}}, \text{CL})\,, \\ \text{CG} = \chi^2\text{−cut}: &\qquad \chi^2(\Delta m^2;\theta) \leq \text{Quantile}(\chi^2_{38\,\text{dof}}, \text{CL})\,. \end{aligned} \tag{3.1}$$

The comparison between the CG method and the $\Delta\chi^2$-cut, which we can consider an approximation of the FC ordering, is shown in figure 4 and presents all the features described in the previous section. The differences are now rather evident. The CG regions are empty until the $\approx 40\%$ CL, while the $\Delta\chi^2$ regions are never empty (as the FC ones). The two methods give equal regions for $\sim 45\%$ CL. With a larger CL the CG regions are bigger than the $\Delta\chi^2$ regions. Figure 4 shows that the arbitrariness in constructing frequentist acceptance regions can be quite significant.

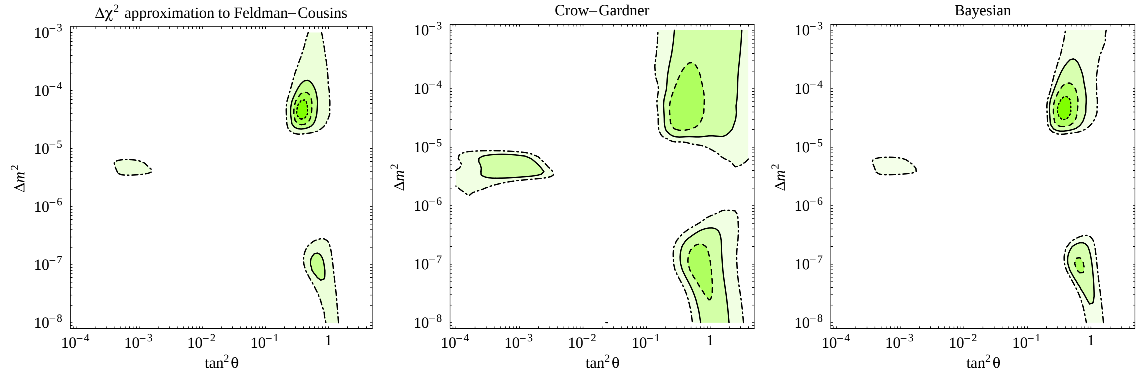

Figure 4: Comparison between the $\Delta\chi^2$-cut (left), the CG method (middle) and a bayesian fit done assuming a flat prior in the plotted variables (right). We show the 35% (dotted lines), 68% (dashed), 90% (continuous) and 99% CL (dot-dashed) regions.

For example a CG fit accepts a large $\Delta m^2 \approx 10^{-3}\,\mathrm{eV}^2$ at $\approx 75\%$ CL, while the $\Delta\chi^2$ approximation to the FC fit rules it out at $\approx 98\%$ CL. One should keep in mind the statistical assumptions behind this limit, when using it to demonstrate the necessity of a hierarchy between the squared-mass differences characteristic of the solar and atmospheric anomalies. We now explain the reason of this difference and argue that the FC fit is the relevant one.

It could seem strange that, after including all data, the CG confidence regions with high CL are much larger than when fitting only the rates. This happens because the CG method does not use the information on the parameters contained in the data in the most efficient way. Statistical fluctuation leads to experimental results that do not lie on the theoretical surface. The CG ordering, treating in the same way the directions in the data space "perpendicular" and "parallel" to the surface, does not use the information contained in the "distance" from the theoretical surface to obtain further information on the parameters. This is why the CG ordering leads to strange results, though correct from the coverage point of view. The FC method uses the "distance" from the theoretical surface to recognize and eliminate the statistical fluctuations that have nothing to do with the determination of the parameters. This difference between the two procedures is much more significant when fitting the full data set than when fitting the rates only. There are now many more data than unknown parameters: the theoretical surface is two-dimensional in a 38-dimensional data space.

In figure 4 we also show a bayesian fit, done assuming a flat prior $dp(\Delta m^2, \theta) = d\ln\Delta m^2\, d\ln\tan\theta$. Unlike the CG fit, this bayesian fit should coincide with the $\Delta\chi^2$ approximation to the FC fit if the pdf were a gaussian function of $\ln\Delta m^2$ and $\ln\tan\theta$. By comparing the two fits we can again see visible but not crucial corrections due to the non-gaussianity in the theoretical parameters $\Delta m, \theta$. The arbitrariness in the prior distribution function $p(\Delta m^2, \theta)$ gives an uncertainty comparable to the effect due to the non-gaussianity.

JHEP05(2001)052

4. The goodness of fit

A goodness-of-fit (GOF) test studies the probability of the experimental data $p(\text{data}|\text{theory})$ under given theoretical hypotheses (a model of the sun, the assumption that neutrinos are oscillating rather than decaying…) leading to statements of the form: if all the hypotheses are true, the probability that the discrepancy between predictions and data is due to statistical fluctuations is less than a certain amount. The purpose is to understand if the theoretical hypotheses used to explain the data are plausible or not.

When analyzing the most recent data, one encounters the following paradoxical situation: the LOW solution gives a *poor* fit of the solar rates only (e.g. [23] finds a GOF of 0.5%). After including the full data set, LOW gives a *good* fit (e.g. [23] finds a GOF of 40%). The paradox is that we have added 35 zenith and energy bins, in which there is no signal for neutrino oscillations. It is clearly necessary to understand better the meaning of the GOF test before we can decide if LOW gives a decent fit or not. This important question also applies to the other alternative solutions. Such tests are based on Pearson's χ^2: the quantity [20]

$$\sum_{ij}(x_i^{\rm exp} - \hat{x}_i^{\rm th})\hat{\sigma}_{ij}^{-2}(x_j^{\rm exp} - \hat{x}_j^{\rm th}) \tag{4.1}$$

is asymptotically distributed as a χ^2 with $N_{\rm data} - N_{\rm param}$ dof.[5] The hats indicate that the corresponding quantity must be evaluated at the maximum likelihood parameter point. The CG ordering is deeply linked to this GOF test: the absence of a confidence region until a given CL value is correlated to the goodness of the fit.

The paradoxical increase of the GOF of LOW is clearly due to the fact that the Pearson test does not recognize that there is a problem concentrated in the three solar rates that contain all the evidence for neutrino oscillations. It only sees that the total χ^2 is roughly equal to the large total number of dof (38), so that the fit seems good. We now explain why the Pearson's test is not adequate for such a situation. Pearson's test does a precise thing: it 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. Therefore the inclusion of more data changes also the set of alternative hypotheses which we compare with. *Describing the recoil electron spectrum in terms of 18 energy bins implies that we admit alternative theories with fuzzy energy spectra.* No physical mechanism could generate an irregular spectrum, so that we do not want to test this aspect. The measured spectrum is of course regular, and Pearson test rewards the LOW solution for this reason. To better understand

[5] The result is exact if the variation of $\log\det\sigma^2$ can be neglected. Furthermore, with a finite number of data, this test is exact only with theoretical rates depending linearly on the parameters. The deviation from this approximation leads to a small overestimate of the GOF: for example, fitting only the rates, the GOF of SMA gets corrected from 51.8% to 48.4% [14].

Goodness of fit	(*a*) Rates only	(*b*) Rates and spectra: naive result	(*c*) Rates and spectra: refined result
SMA	55% (58% [22], 50% [23],51.8% [14])	30% (34% [23])	$\lesssim 2\,\%$
LMA	6% (10% [22], 8% [23], 6.3% [14])	60% (59% [23])	$\lesssim 15\,\%$
LOW	0.7% (1.6% [22], 0.5% [23], 1.1% [14])	50% (40% [23])	$\lesssim 2\,\%$
$P_{ee} = \mathrm{cte}$	0.3%	28%	$\lesssim 1\,\%$

Table 1: GOF of the SMA, LMA, LOW and energy-independent solutions obtained by fitting (a) only the rates, (b) including all the data, (c) including only the "most significant" data. The energy-independent solution contains only one free parameter, while SMA, LMA, LOW have two free parameters. Our results are compared with the corresponding ones in [14, 22, 23]. The symbol $\lesssim$ recalls that the GOF values could be slightly lower.

this point, suppose that we add as new data the direction of arrival of the interacting neutrinos. All solutions (including the no-oscillation hypothesis) would have a higher GOF. It is obvious that these solutions are much better than a generic one, because they at least "know" where the Sun is. A meaningful χ^2 test should include only those data that really test the hypothesis under consideration. On the contrary, the inclusion of irrelevant data does not affect the confidence regions built with the FC ordering, so that a naive application of the $\Delta\chi^2$-cut correctly approximates the best-fit regions.

We therefore conclude that testing the goodness of the fit using a lot of energy bins gives a formally correct answer to an irrelevant question. If a smaller set of data were used to describe the spectral and angular information, the set of alternative hypotheses would be more reasonable and one would conclude that there is a goodness-of-fit problem. Most of the information on the energy and zenith-angle spectra can be condensed into observables such as the mean recoil electron energy and the day/night asymmetry, as shown in fits presented by the SK collaboration [4].

Within our assumption of 2-neutrino oscillations, the main new information encoded in SK spectral and day/night data is that the survival probability $P_{ee}(E_\nu)$ can only have a mild energy dependence around $E_\nu \sim 10\,\mathrm{MeV}$. This can be seen in a simple way by parameterizing $P_{ee}(E_\nu)$ as

$$P_{ee}(E_\nu) = P_0 + P_0' \cdot \left(\frac{E_\nu}{10\,\mathrm{MeV}} - 1\right) + \frac{P_0''}{2}\left(\frac{E_\nu}{10\,\mathrm{MeV}} - 1\right)^2 + \cdots$$

The SK spectral data measure $P_0' = -0.05 \pm 0.1$ and disfavour the SMA solution because it prefers a larger slope P_0'. Significant non-linearities in $P_{ee}(E_\nu)$ are not predicted in SMA, LMA, LOW oscillations, nor could be recognized by SK (the present error on P_0'' is ~ 1). In conclusion most of the information contained in the SK spectral data can be conveniently condensed into a single observable f, that measures the slope of $P_{ee}(E_\nu)$. One possible choice is the ratio between the rate of "low energy" (i.e. $T_e < 9\,\mathrm{MeV}$) and "high energy" (i.e. $9\,\mathrm{MeV} < T_e < 13\,\mathrm{MeV}$)

recoil electrons, as measured by SK. The upper bound on the recoil electron energy T_e has been chosen in order to avoid potential problems due to an enhanced flux of hep neutrinos. The measured value and the uncertainty on $f(\rho_i)$ can be easily deduced from the SK data on the full energy spectrum: $\sigma_f^2 = f_i \sigma_{ij}^2 f_j$ (in gaussian approximation), where $f_i \equiv \partial f / \partial \rho_i$, $\{\rho_i\}$ is the full set of SK bins and σ_{ij}^2 is the full error matrix.

By supplementing the fit of the total neutrino rates with a single observable f we find the GOF values shown in table 1. The symbol $\lesssim$ recalls that, especially in the SMA case, it could be possible to obtain slightly lower GOF values by identifying another observable more sensitive to the energy dependence of the neutrino survival probability. However, a variation of the GOF between, say, 1% and 4%, is within the uncertainty due to arbitrariness inherent in any statistical analysis. The important point is that the GOF values are significantly lower than the values based on a naive χ^2 test, and motivate a non-standard analysis of the solar neutrino anomaly [24]. The SK collaboration [2] finds that SMA now gives a poor fit using another reasonable procedure: at 97% CL, the region favoured by total rates falls inside the region disfavoured by spectral and day/night data.

5. Conclusions

Fits of solar neutrino data are usually done using the $\Delta\chi^2$-cut valid in the gaussian approximation and find few distinct best-fit solutions (LMA, SMA, LOW, ...). Since a gaussian would have only one peak, it is useful to check the validity of the $\Delta\chi^2$-cut approximation by comparing its results with the exact confidence regions built using the Neyman construction. This has been done using two different ordering prescriptions, proposed by Crow-Gardner (CG) and by Feldman-Cousins (FC). *We find that the $\Delta\chi^2$ cut provides a good approximation to the Feldman-Cousins confidence regions.*

When the full data set is used, there is some significant difference between the CG and FC regions. Even if the two methods give regions with the same statistical meaning, their conceptual significance is different. The FC regions are not influenced by statistical fluctuations with no information on the oscillation parameters, while the CG regions are composed by all the oscillation parameters that provide an acceptable fit of the data. The meaning of the results is deeply influenced by the assumptions involved in the statistical analysis. We finally show that a correct understanding of the meaning of Pearson "goodness-of-fit" test invalidates the statement that all solutions (LMA, LOW and SMA) presently give a good fit.

We hope that our refined statistical analysis has been useful for clarifying some aspects of solar neutrino fits. Only refined experimental data will allow to identify the solution of the solar neutrino problem.

Acknowledgments

We thank Andrea Gambassi, Alberto Nicolis and Donato Nicolò for useful discussions. We thank Y. Suzuki for the unpublished SuperKamiokande data.

A. Details of the computation

The energy spectra for the independent components of the solar neutrino flux have been obtained from [8]. The neutrino production has been averaged for each flux component over the position in the sun as predicted in [5, 8]. This averaging does not give significant corrections. MSW oscillations inside the sun have been taken into account in the following way. The 3×3 density matrix ρ_S for neutrinos exiting from the sun is computed using the Landau-Zener approximation with the level-crossing probability appropriate for an exponential density profile [6, 7]. The density profile has been taken from [8] and is quasi-exponential: small corrections to ρ_S have been approximately included. Oscillation effects outside the sun are described by the evolution matrix U, so that at the detection point $\rho_E = U \rho_S U^\dagger$. In particular, earth regeneration effects have been computed numerically using the mantle-core approximation for the earth density profile. We have used the mean mantle density appropriate for each trajectory as predicted by the preliminary Earth model [25]. The detection cross sections in Gallium and Chlorine experiments have been taken from [8], performing appropriate interpolations. We have used the tree-level Standard Model expression for the neutrino/electron cross section at SK.

The total neutrino rates measured with the three kinds of experimental techniques are [1]–[4]

$$\begin{aligned} R_{\rm Cl}|_{\rm exp} &= (2.56 \pm 0.22) \mathrm{SNU}\,, \\ R_{\rm Ga}|_{\rm exp} &= (74.7 \pm 5) \mathrm{SNU}\,, \\ R_{\rm SK}|_{\rm exp} &= (2.40 \pm 0.08) \cdot 10^6\,\mathrm{cm}^{-2}\mathrm{s}^{-1} \end{aligned} \tag{A.1}$$

where SNU $\equiv 10^{-36}$ interactions per target atom and per second. We have combined systematic errors in quadrature with statistical errors. The probability distribution function $p(\text{data}|\text{theory})$ is computed using the covariance matrix described in [9, 22]. Around the best-fits, it agrees well with the simpler pdf used in [24]. The experimental energy resolution at SK has been taken into account as suggested in [26, 27].

The solar-model-independent SK data included in the fit are the energy spectrum of the recoil electrons measured separately at SK during the day and during the night. Each energy spectrum [28] is composed of 17 energy bins between 5.5 and 14 MeV, plus one bin between 14 and 20 MeV. For these data we have used the pdf suggested by the SK collaboration [28], and described in [27] for a slightly different set of data.

References

[1] The results of the Homestake experiment are reported in K. Lande, T. Daily, R. Davis, J.R. Distel, B.T. Clevland, C.K. Lee, P.S. Wildenhain and J. Ullman, *Measurement of the solar electron neutrino flux with the Homestake Chlorine detector*, *Astrophys. J.* **496** (1998) 505.

[2] SuperKamiokande collaboration, *Measurement of the solar neutrino energy spectrum using neutrino electron scattering*, *Phys. Rev. Lett.* **82** (1999) 2430 [hep-ex/9812011]; *Solar neutrino results from Super-Kamiokande*, *Nucl. Phys.* **77** *(Proc. Suppl.)* (1999) 35; *Neutrino oscillations at Super-Kamiokande*, *Nucl. Phys.* **81** *(Proc. Suppl.)* (2000) 133; New data recently presented in Super-Kamiokande collaboration, *Constraints on neutrino oscillations using 1258 days of Super-Kamiokande solar neutrino data*, hep-ex/0103033 (after 1258 days of data taking) do not present significant differences with respect to the ones employed in our fit (1117 days)

[3] Gallex collaboration, *Gallex solar neutrino observations: results for Gallex IV*, *Phys. Lett.* **B 447** (1999) 127;
SAGE collaboration, *Measurement of the solar neutrino capture rate with gallium metal*, *Phys. Rev.* **C 60** (1999) 055801 [astro-ph/9907113].

[4] Talks by E. Bellotti, V. Gavrin and Y. Suzuki at the conference *Neutrino 2000*, Sudbury, Canada, June 2000.

[5] J.N. Bahcall, S. Basu and M.H. Pinsonneault, *How uncertain are solar neutrino predictions?*, *Phys. Lett.* **B 433** (1998) 1 [astro-ph/9805135]; *Solar models: current epoch and time dependences, neutrinos, and helioseismological properties*, astro-ph/0010346 and references therein.

[6] L. Wolfenstein, *Neutrino oscillations in matter*, *Phys. Rev.* **D 17** (1978) 2369;
S.P. Mikheyev and A.Yu. Smirnov, *Resonance amplification of oscillations in matter and spectroscopy of solar neutrinos*, *Sov. J. Nucl. Phys.* **42** (1986) 913.

[7] S. Parke, *Nonadiabatic level crossing in resonant neutrino oscillations*, *Phys. Rev. Lett.* **57** (1986) 1275;
P. Pizzochero, *Nonadiabatic level crossing in neutrino oscillations for an exponential solar density profile*, *Phys. Rev.* **D 36** (1987) 2293;
S.T. Petcov, *Exact analytic description of two neutrino oscillations in matter with exponentially varying density*, *Phys. Lett.* **B 200** (1998) 373;
for a review see T.K. Kuo, J. Pantaleone, *Neutrino oscillations in matter*, *Rev. Mod. Phys.* **61** (1989) 937.

[8] J.N. Bahcall, http://www.sns.ias.edu/~jnb.

[9] G.L. Fogli and E. Lisi, *Standard solar model uncertainties and their correlations in the analysis of the solar neutrino problem*, *Astropart. Phys.* **3** (1995) 185.

[10] There is an enormous amount of literature, much of it written by people who are fervent advocates of one or the other approach. For a reasonably balanced discussion, see e.g. G. D'Agostini, CERN 'yellow' report 99-03, available at `www.cern.ch` (for a presentation of the bayesian approach) and recent editions of the Particle Data Group, available at `pdg.lbl.gov` (for a presentation of the frequentist approach).

[11] J. Neyman, *Outline of a theory of statistical estimation based on the classical theory of probability*, *Philos. Trans. R. Soc. London* **A236** (1937) 333.

[12] E.L. Crow, R.S. Gardner, *Biometrika* **46** (1959) 441.

[13] G.J. Feldman and R.D. Cousins, *A unified approach to the classical statistical analysis of small signals*, *Phys. Rev.* **D 57** (3873) 1998 [`physics/9711021`], the Feldman-Cousins ordering was implicit in A. Stuart and J.K. Ord, Kendall's Advanced Theory of Statistics, Vol. 2, *Classical inference and relationship* 5th ed., Oxford University Press, New York, 1991;
see also earlier editions by Kendall and Stuart, sec. 23.1.

[14] M.V. Garzelli and C. Giunti, *A frequentist analysis of solar neutrino data*, `hep-ph/0007155`.

[15] CHOOZ collaboration, *Limits on neutrino oscillations from the CHOOZ experiment*, *Phys. Lett.* **B 466** (1999) 415 [`hep-ex/9907037`];
see also F. Boehm et al., *Search for neutrino oscillations at the Palo Verde nuclear reactors*, *Phys. Rev. Lett.* **84** (3764) 2000 [`hep-ex/9912050`].

[16] The LSND collaboration, *Evidence for neutrino oscillations from muon decay at rest*, *Phys. Rev.* **C 54** (1996) 2685 [`nucl-ex/9605001`];
The LSND collaboration, *Evidence for $\nu_\mu \to \nu_e$ oscillations from pion decay in flight neutrinos*, *Phys. Rev.* **C 58** (1998) 2489 [`nucl-ex/9706006`].

[17] Bugey collaboration, *Search for neutrino oscillations at 15-meters, 40-meters, and 95-meters from a nuclear power reactor at Bugey*, *Nucl. Phys.* **B 434** (1995) 503.

[18] KARMEN collaboration, *Latest results of the Karmen2 experiment*, *Nucl. Phys.* **91** *(Proc. Suppl.)* (2000) 191 [`hep-ex/0008002`].

[19] Super-Kamiokande collaboration, *Tau neutrinos favored over sterile neutrinos in atmospheric muon neutrino oscillations*, *Phys. Rev. Lett.* **85** (3999) 2000 [`hep-ex/0009001`].

[20] W.T. Eadie et al., *Statistical methods in experimental physics*, Amsterdam, 1971.

[21] G. Punzi, *A stronger classical definition of confidence limits*, `hep-ex/9912048`.

[22] G.L. Fogli, E. Lisi, D. Montanino and A. Palazzo, *Day-night asymmetry of high and low energy solar neutrino events in Super-Kamiokande and in the sudbury neutrino observatory*, *Phys. Rev.* **D 62** (2000) 113003 [`hep-ph/0008012`].

[23] M.C. Gonzalez-Garcia and C. Peña-Garay, *Global and unified analysis of solar neutrino data*, *Nucl. Phys.* **91** *(Proc. Suppl.)* (2000) 80 [hep-ph/0009041].

[24] R. Barbieri and A. Strumia, *Non standard analysis of the solar neutrino anomaly*, *J. High Energy Phys.* **12** (2000) 016 [hep-ph/0011307].

[25] A.M. Dziewonski and D.L. Anderson, *Preliminary reference earth model*, *Phys. Earth Planet. Interior* **25** (1981) 207.

[26] J.N. Bahcall and E. Lisi, *Tests of electron flavor conservation with the sudbury neutrino observatory*, *Phys. Rev.* **D 54** (1996) 5417 [hep-ph/9607433].

[27] M.C. Gonzalez-Garcia, M. Maltoni, C. Peña-Garay and J.W. F. Valle, *Global three-neutrino oscillation analysis of neutrino data*, *Phys. Rev.* **D 63** (2001) 033005 [hep-ph/0009350].

[28] Super-Kamiokande collaboration, private communication.