VALOR Oscillation Fits

Lorena Escudero

for the VALOR group:

C. Andreopoulos, C. Barry, F. Bench, A. Chappell, T. Dealtry, S. Dennis, L. Escudero, R. Jones, N. Grant, M. Roda, D. Sgalaberna, R. Shah

Phystat-ν Fermilab
19th September 2016

VALOR

[1,2] **T2K ν_μ disappearance analysis** (T. Dealtry thesis and 2014 PRL)
[3,4] **T2K joint ν_μ disappearance and ν_e appearance analysis** (L. Escudero thesis and PRD 2015)
[5,6] **T2K anti-ν_μ disappearance analysis and NSI** (S. Dennis thesis and PRL 2016)
[7] **T2K joint neutrino and antineutrino oscillations** (D. Sgalaberna thesis)
[8] D. Sgalaberna, poster at Phystat-nu Japan
[9] **T2K anti-ν_e appearance,** R. Shah, poster at Neutrino 2016
[10] R. Shah, poster at Phystat-nu Japan

…and other ongoing analysis in T2K, HK and DUNE (visit https://valor.pp.rl.ac.uk)

VALOR Oscillation Fit

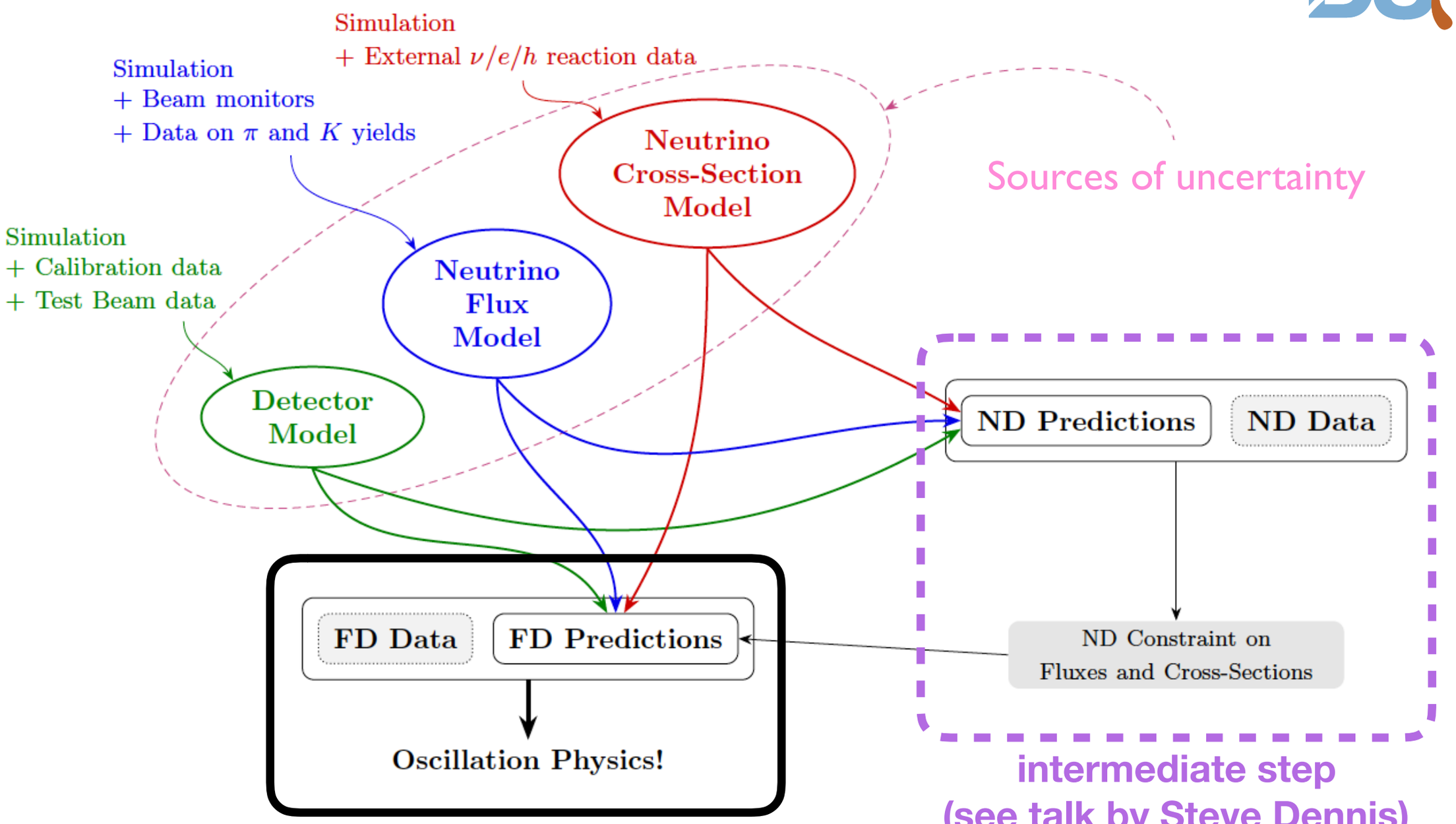
- Joint measurement of oscillation parameters
- In a 3-flavour framework (with 3+1, 3+2, 1+3+1 extensions) including matter effects
- For multiple samples (selections based on topologies), detectors and beam configurations
- Including correlated systematic uncertainties

- Historically, performs an indirect extrapolation near-to-far detector for long baseline neutrino experiments
- Extended to perform neutrino oscillation fits with multiple detectors, and multiple experiments

- Uses a binned likelihood-ratio method
- Minimisation using profiling or a mixed method combining profiling+marginalization

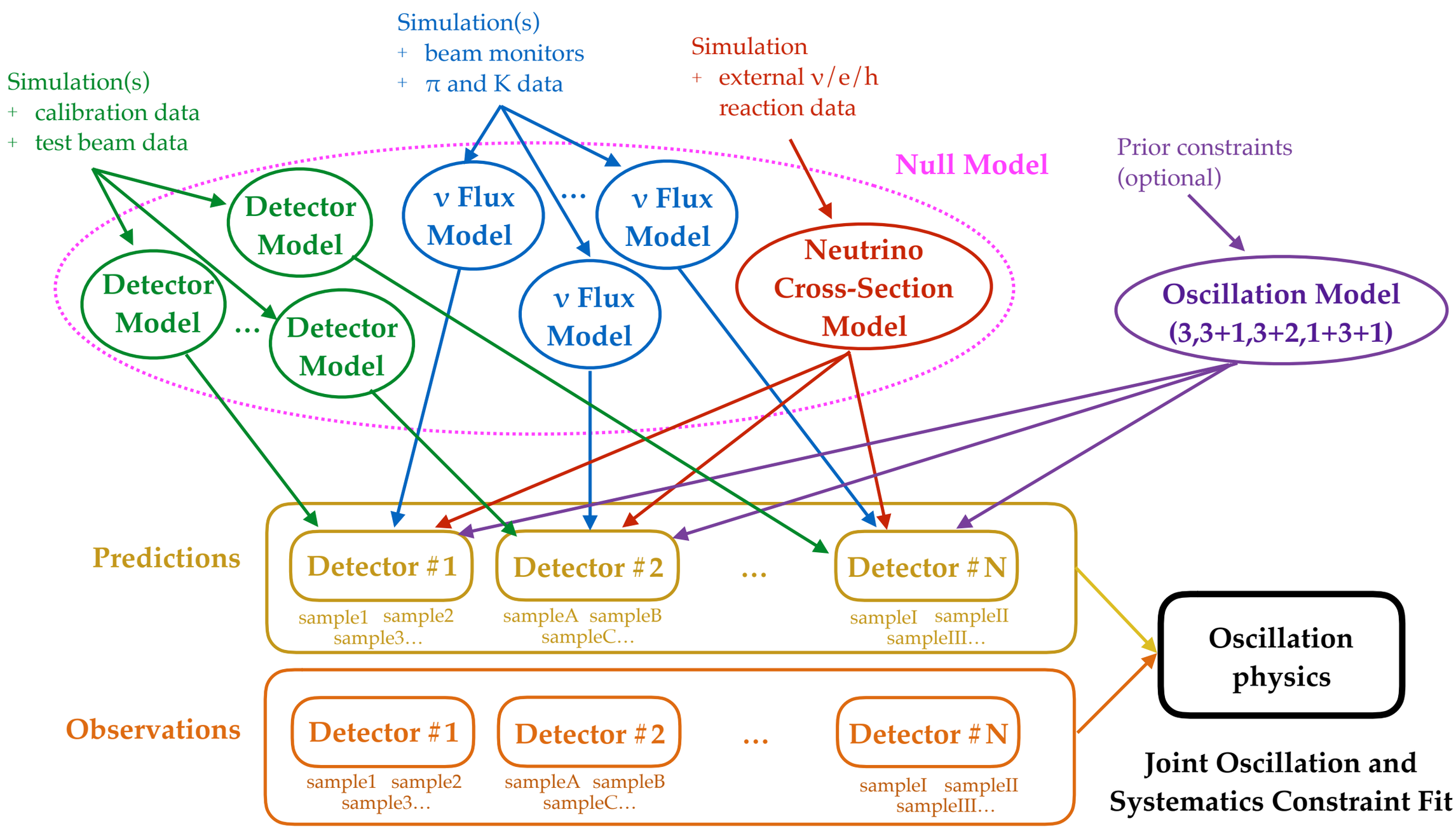
VALOR: ND+FD

Historically developed for **T2K long baseline** neutrino oscillation fits
Schematic for DUNE, including near and far detector samples



VALOR: Multiple Detectors

Adapted and extended to perform neutrino oscillation fits for **multiple detectors** and **multiple experiments**



Spectra Prediction

The VALOR framework can fit together an arbitrary number of samples to determine the parameters of a physics hypothesis in the presence of systematic uncertainties.

Each sample corresponds to a given:

- d - detector or subdetector
- b - beam configuration (FHC, RHC)
- s - selection, i.e. final state topology
- r,t - **multi dimensional** kinematical reconstructed and true phase space

systematics

physics

true kinematics

true interaction mode

$$n^{pred}_{d;b;s}(r;\vec{\theta};\vec{f}) = \sum_m \sum_t P_{d;b;m}(t;\vec{\theta}) \, R_{d;b;s;m}(r,t;\vec{f}) \cdot T_{d;b;s;m}(r,t)$$

Predicted number of events

reconstructed kinematics

Oscillation probabilities

Systematic parameters variations

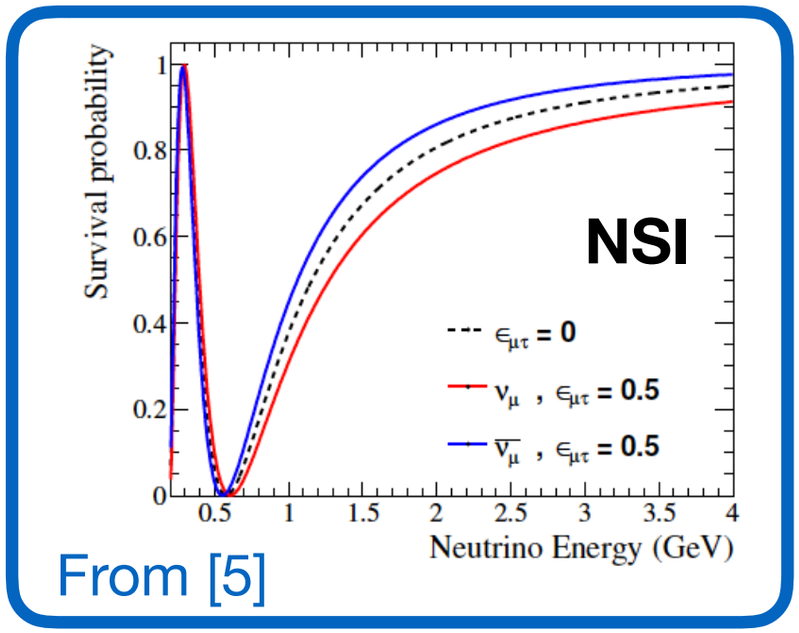
Nominal unoscillated number of events

Oscillation Probabilities

VALOR incorporates **its own library** for the calculation of oscillation probabilities:

- fast, no need to calculate and load tables from other libraries
- extensively checked against other available libraries (Prob3++,GLoBES)
- allows calculations including matter effects in a 3-flavour framework and extensions: 3+1, 3+2, 1+3+1

- addition of phenomenological-based NC NSI effects studied, approximated by the addition of one real free parameter that would affect the ν_μ survival probability

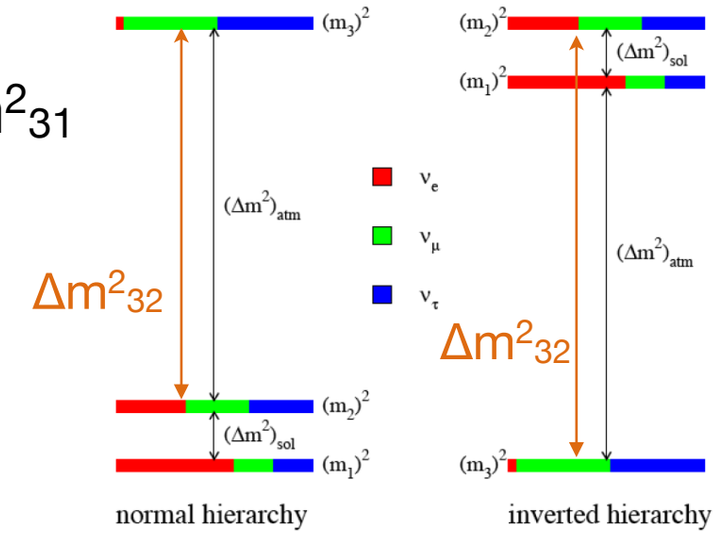


From [5]

- allows usage of different conventions:
  - δ_{CP}, $\sin(\delta_{CP})$
  - double and single mixing angles ($\sin^2\theta_{23}$ vs $\sin^2 2\theta_{23}$ for octant studies)
  - atmospheric mass splitting can be input as Δm^2_{32}, Δm^2_{31}

or as Fogli and Lisi convention: $\Delta m^2_{FL} = m^2_3 - \frac{m^2_2 + m^2_1}{2}$

useful since its absolute value is the same in both mass hierarchies, while the usual $|\Delta m^2_{32}|$ is the largest splitting in IH but the second largest in NH



Likelihood Construction

VALOR analyses obtain measurements of the parameters of interest by comparing prediction and observation typically using a **binned likelihood-ratio method:**

The log-likelihood function is constructed **for each sample s, detector d, beam configuration b**

$$ln\ \lambda_{d;b;s}(\vec{\theta};\vec{f}) = -\sum_{r}\left\{\left(n^{pred}_{d;b;s}(r;\vec{\theta};\vec{f}) - n^{obs}_{d;b;s}(r)\right) + n^{obs}_{d;b;s}(r)\cdot ln\frac{n^{obs}_{d;b;s}(r)}{n^{pred}_{d;b;s}(r;\vec{\theta};\vec{f})}\right\}$$

This statistic is summed over all datasets

$$\lambda(\vec{\theta};\vec{f}) = \left(\prod_{d}\prod_{b}\prod_{s}\lambda_{d;b;s}(\vec{\theta};\vec{f})\right)\cdot\lambda_{prior}(\vec{\theta};\vec{f})$$

adding a penalty term for those parameters with prior constraints, where C are covariance matrices

$$ln\ \lambda_{prior}(\vec{\theta};\vec{f}) = -\frac{1}{2}\left\{(\vec{\theta}-\vec{\theta}_0)^T C_\theta^{-1}(\vec{\theta}-\vec{\theta}_0) + (\vec{f}-\vec{f}_0)^T C_f^{-1}(\vec{f}-\vec{f}_0)\right\}$$

Notice that the sum over all datasets (for different detectors d, etc) can include datasets from different experiments in a multi-experiment analysis

Finally, best-fit values of the parameters of interest are obtained by maximising $\lambda(\vec{\theta};\vec{f})$ or minimising $-2ln\lambda(\vec{\theta};\vec{f})$ (usually denoted as χ^2)

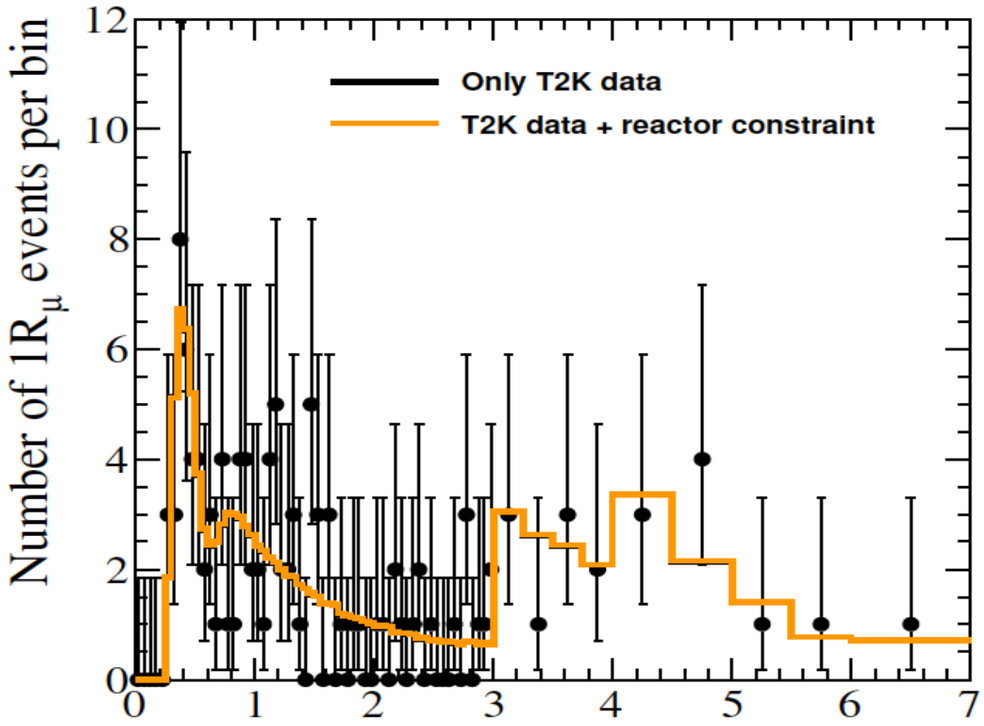
External Constraints

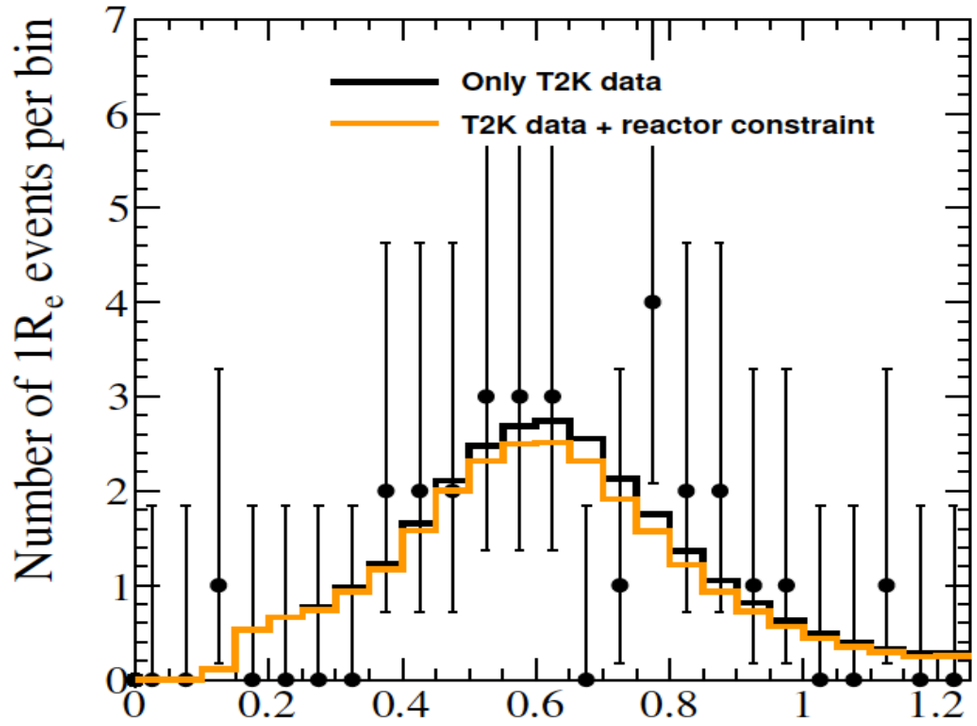
External constraints can be included in the analysis by adding the appropriate penalty term to the log-likelihood.

For example, the constrain on θ_{13} based on the reactor values is included with the penalty term:

$$\chi^2_{reactor} = \left(\frac{\sin^2\theta_{13} - (\sin^2\theta_{13})_{reactor}}{\sigma_{reactor}}\right)^2$$

T2K joint ν_μ disappearance and ν_e appearance analysis [3]





Which affects the best-fit values of the other oscillation parameters and the best-fit spectra

Reactor	MH	$\lvert\Delta m^2_{32}\rvert$ (NH) or $\lvert\Delta m^2_{13}\rvert$ (IH) $10^{-3}eV^2/c^4$	$\sin^2\theta_{23}$	$\sin^2 2\theta_{13}$	δ_{CP}	$N^{1R\mu}_{exp}$	N^{1Re}_{exp}	χ^2/ndf
NO	NH	2.512	0.524	0.162	1.909	119.915	27.999	84.5395/94
YES	NH	2.509	0.527	0.0967	-1.554	120.383	25.870	85.067/94
NO	IH	2.488	0.523	0.187	1.005	119.948	27.998	84.529/94
YES	IH	2.481	0.533	0.0984	-1.556	121.204	23.571	85.931/94

Parameter Elimination

The VALOR group use both profiling and marginalisation. Now we have a mixed scheme **Bayesian-frequentist**, obtaining the parameters of interest by maximising the likelihood while marginalising the nuisance parameters

Profiling

Profiling methods like MIGRAD algorithm in MINUIT can be used to minimise χ^2. VALOR framework is adapted to use also other minimisers from GSL libraries

Marginalisation

The marginal likelihood with prior distributions π: can be numerically approximated by generating toy MC experiments (see backup). Many toys are needed to sample the nuisance parameters phase space.

$$\lambda_{marg}(\vec{\theta'}) = \int \lambda(\vec{\theta'}; \vec{f'})\pi(\vec{f'})d\vec{f'}$$

$$\lambda_{marg}(\vec{\theta'}) \approx \frac{1}{n}\sum_{i=0}^{n-1} \lambda(\vec{\theta'}; \vec{f'}_i)$$

During the marginalisation process, random values of the nuisance parameters are drawn from the π distribution containing the prior knowledge. Thus, the penalty term for those parameters is not necessary.

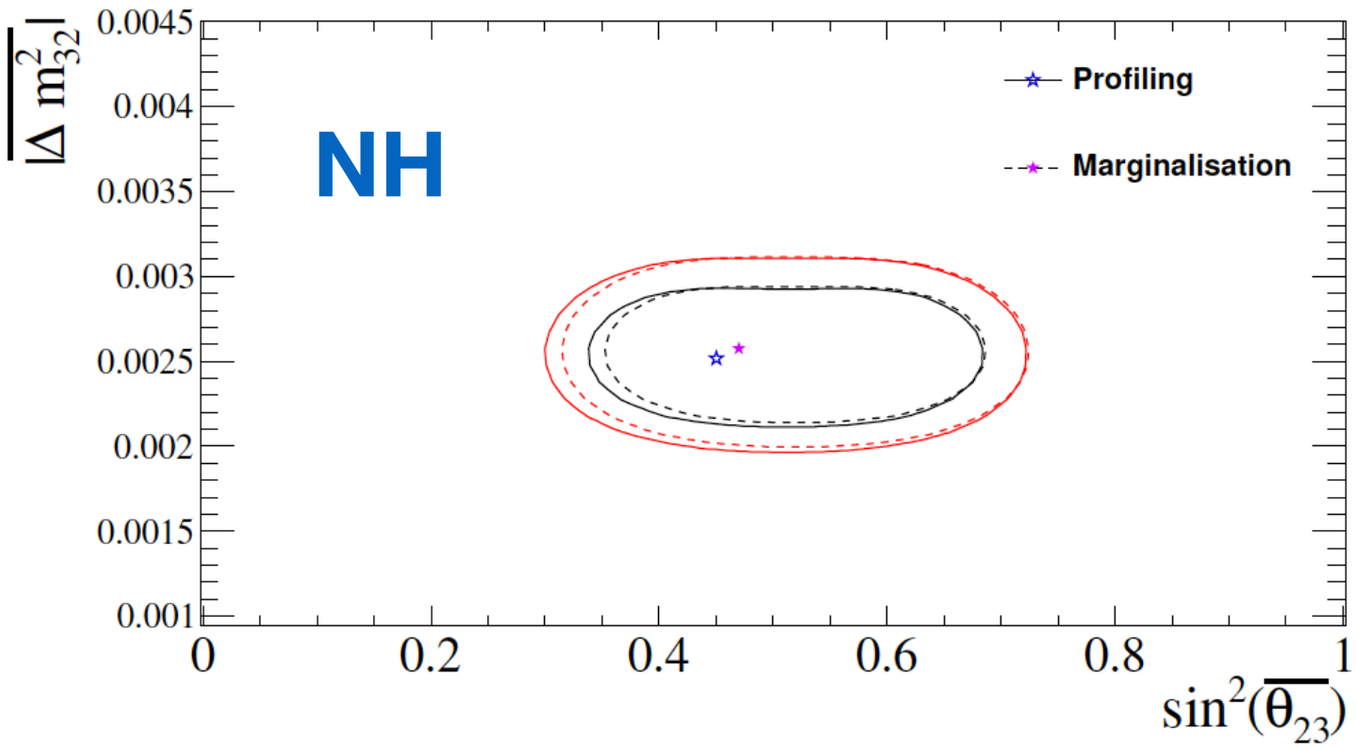
The MH and θ_{23} octant use discrete priors, minimizing for each choice independently

Parameter Elimination

Example of usage of both methods for parameter elimination

- profiling: using MIGRAD algorithm in MINUIT
- marginalisation: integrating summing over a large sample of toy experiments generated with prior π:
    - for each systematic parameter, this example uses a Gaussian prior π, with mean/rms as the nominal value/1σ error
    - Cholesky method used for correlated systematics (see backup)

Muon antineutrino disappearance analysis from [5]



In this example, only a small difference is observed in the contours and best-fit values.

How to choose?

- How well we know the priors
- Coverage studies

Fit Validation

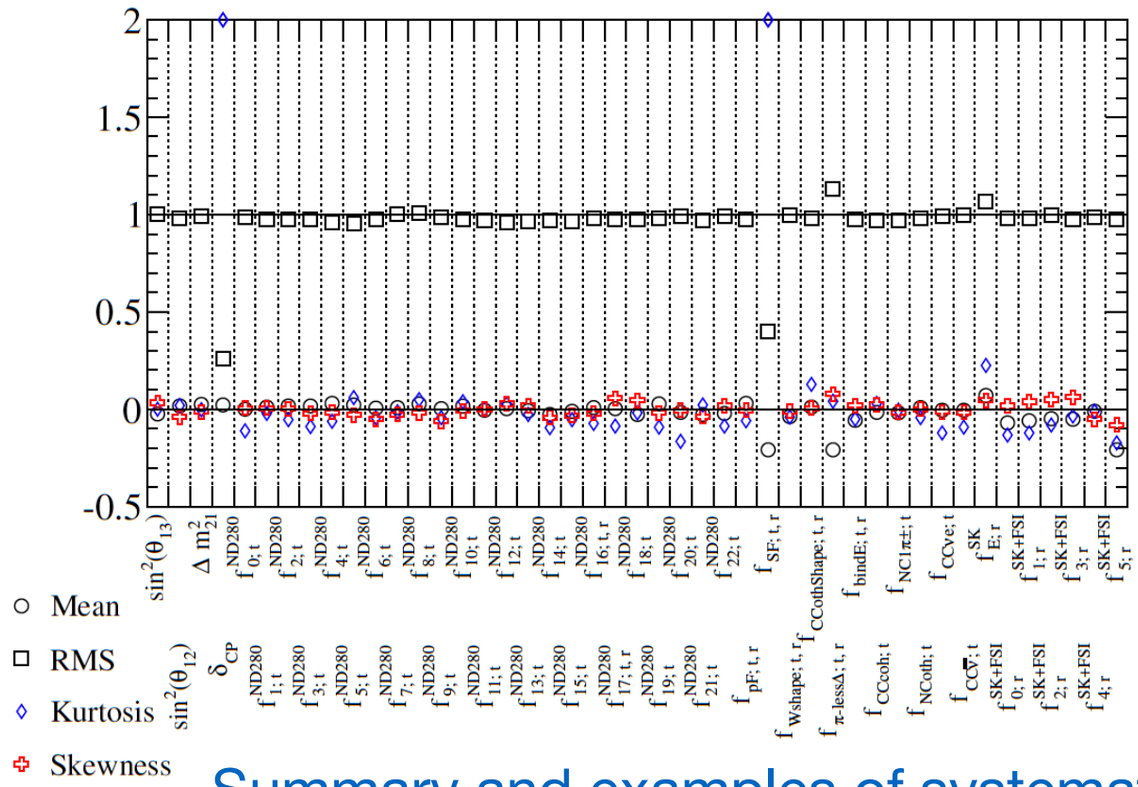
Extensive tests of the fitter performance have been performed (see [1,3,5,7]).
In addition to the ones described here, studies with fake datasets were also performed.

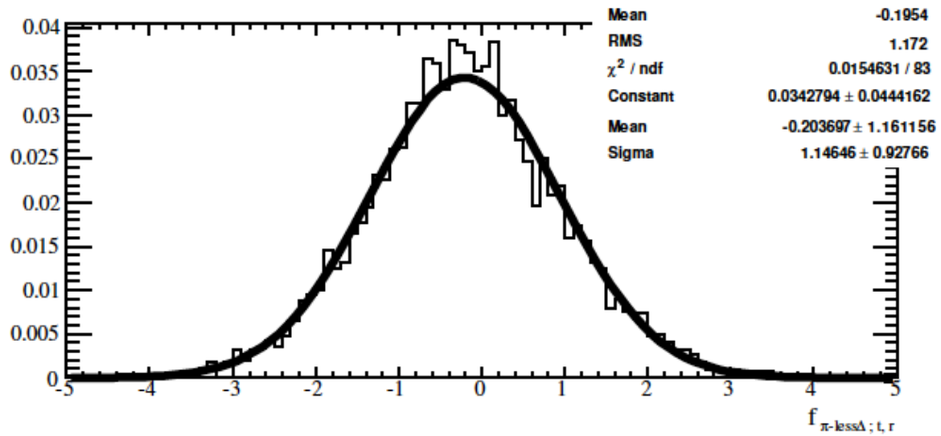
Systematic uncertainties

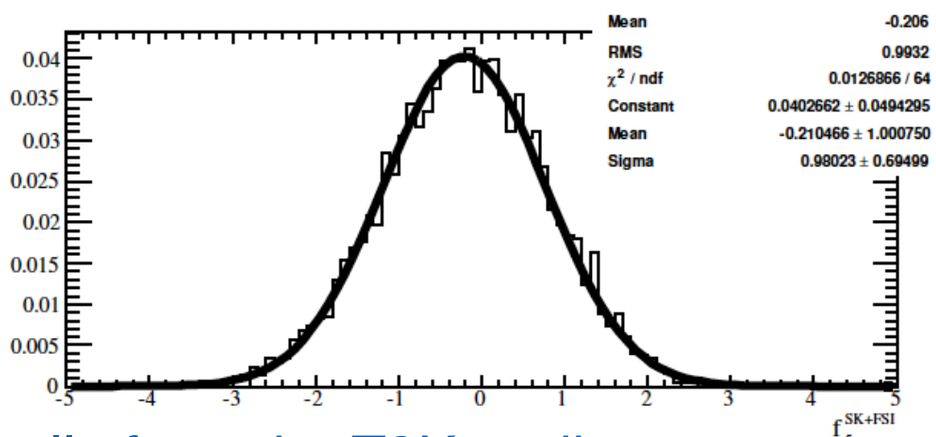
For the systematic uncertainties, the pulls are calculated as: $\frac{f_{\text{bestfit}} - f_{\text{nominal}}}{\sigma}$,

Any pathologies were studied in detail:

- bias may appear due to unphysical values limiting parameters range
- bias may appear due to low statistics
- Gaussian not expected for Energy scale







Summary and examples of systematic parameters pulls from the T2K ν_μ disappearance [1]

Fit Validation

Oscillation parameters

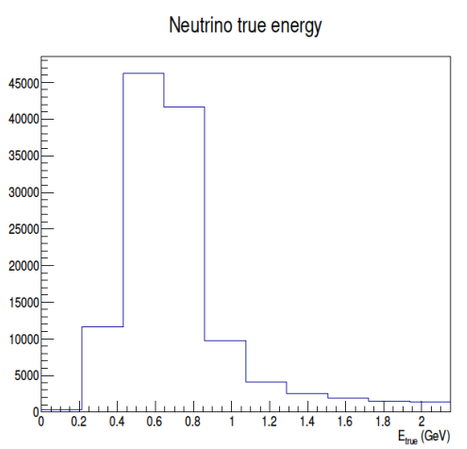
For the oscillation parameters, we studied their distribution of residuals and understood them taking into account the effects of:

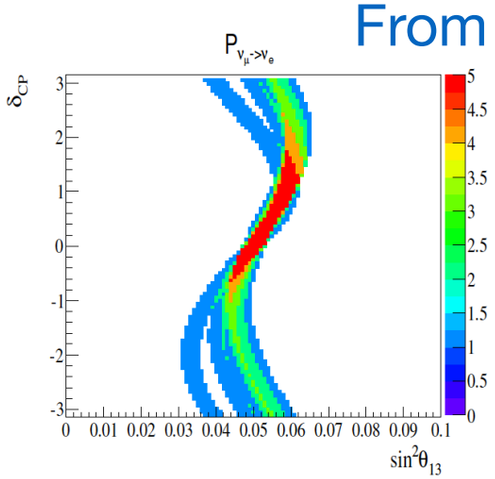
- Physical boundaries of the oscillation parameters
- Degeneracies and correlations between them
- Statistical fluctuations

Non-maximal values explained with statistical fluctuations of the number of 1Rμ events at the oscillation dip

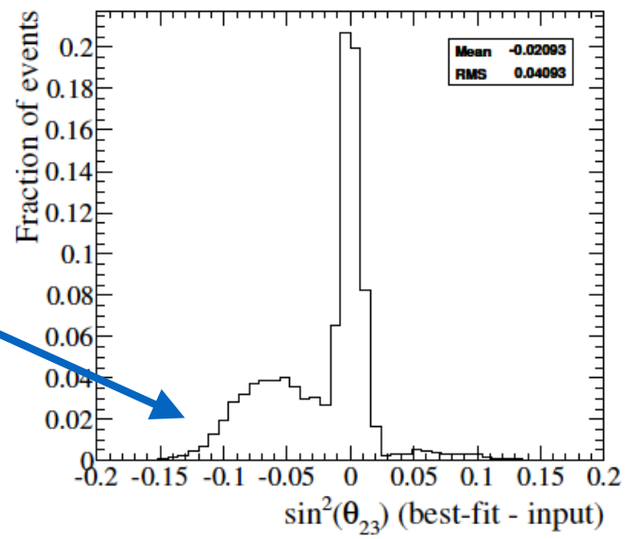
Asymmetry in $\sin^2\theta_{13}$ and δ_{CP} residuals due to the convolution of the energy spectrum and oscillation probability

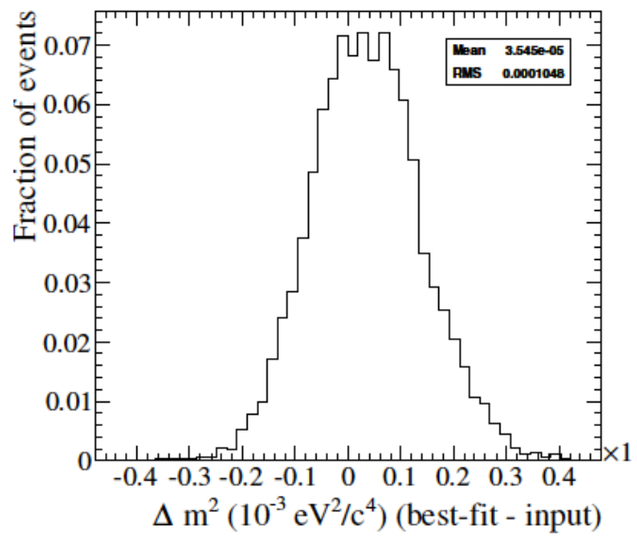
From [3]

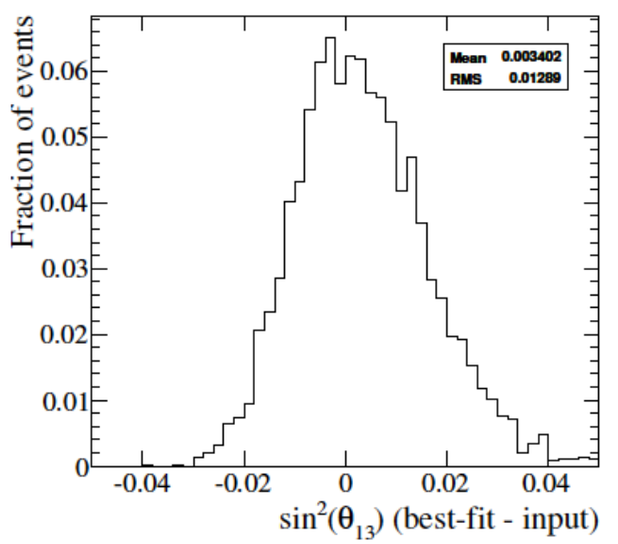


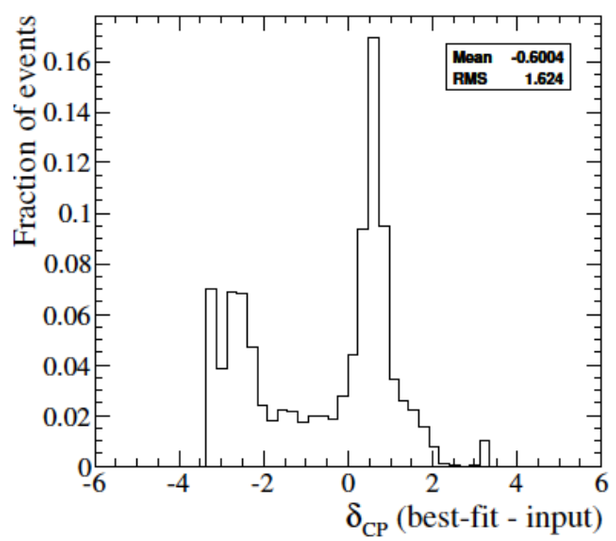


T2K ν_μ disap. + ν_e app. analysis [3]









Goodness of fit tests

The advantage of the likelihood ratio method is that, in the large sample limit, $-2ln\lambda(\vec{\theta};\vec{f})$ has a χ^2 distribution and can be used as goodness-of-fit test, which complement best-fit parameter estimation analysing the agreement between data and (best-fit) model

When bins are sparsely populated, instead of χ^2 a **p-value** is calculated to determine the goodness of fit, with a large number of toy experiments generated at the best-fit point of the fit to data, fitted in a coarser bin χ^2_{gof}

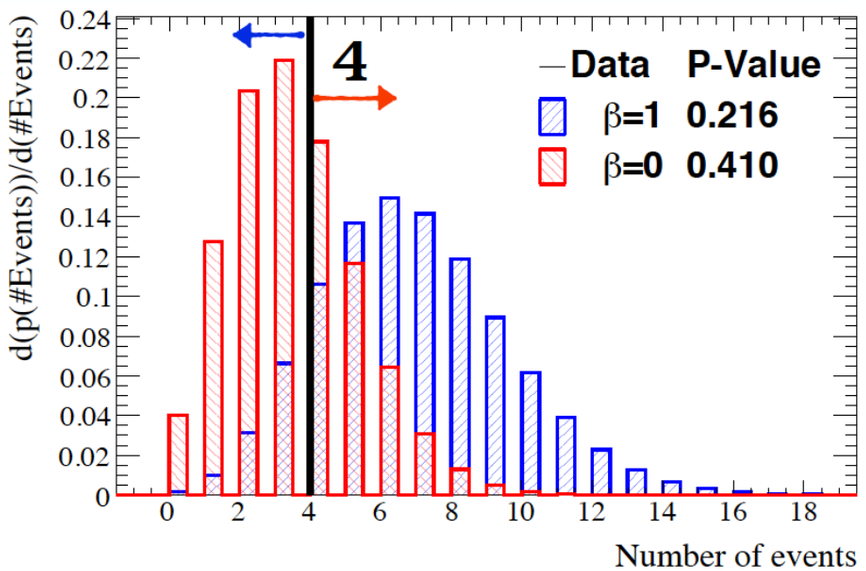
p-value = proportion of expts with $\chi^2_{gof} > (\chi^2_{gof})_{data}$

Example: T2K anti-ν_e appearance analysis [9,10]

$P_{osc}(\bar{\nu}_\mu \to \bar{\nu}_e) = \beta P_{osc}(PMNS)$

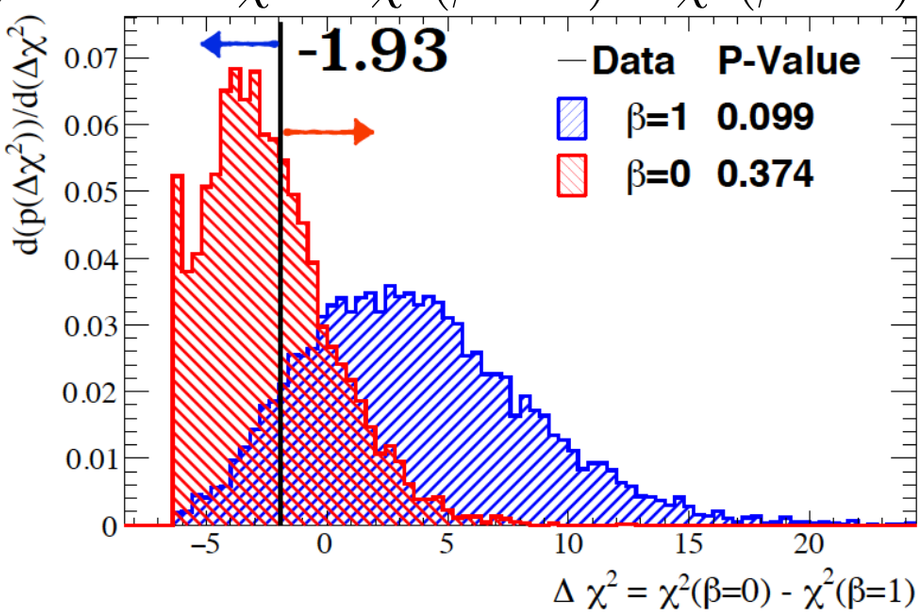
Null hypothesis: no anti-ν_e appearance, i.e. β=0
Bayes factor also calculated (backup)

Rate-only statistic: # of ν_e events



Rate+shape statistic:

$\Delta\chi^2 = \chi^2(\beta=0) - \chi^2(\beta=1)$



Goodness of fit tests

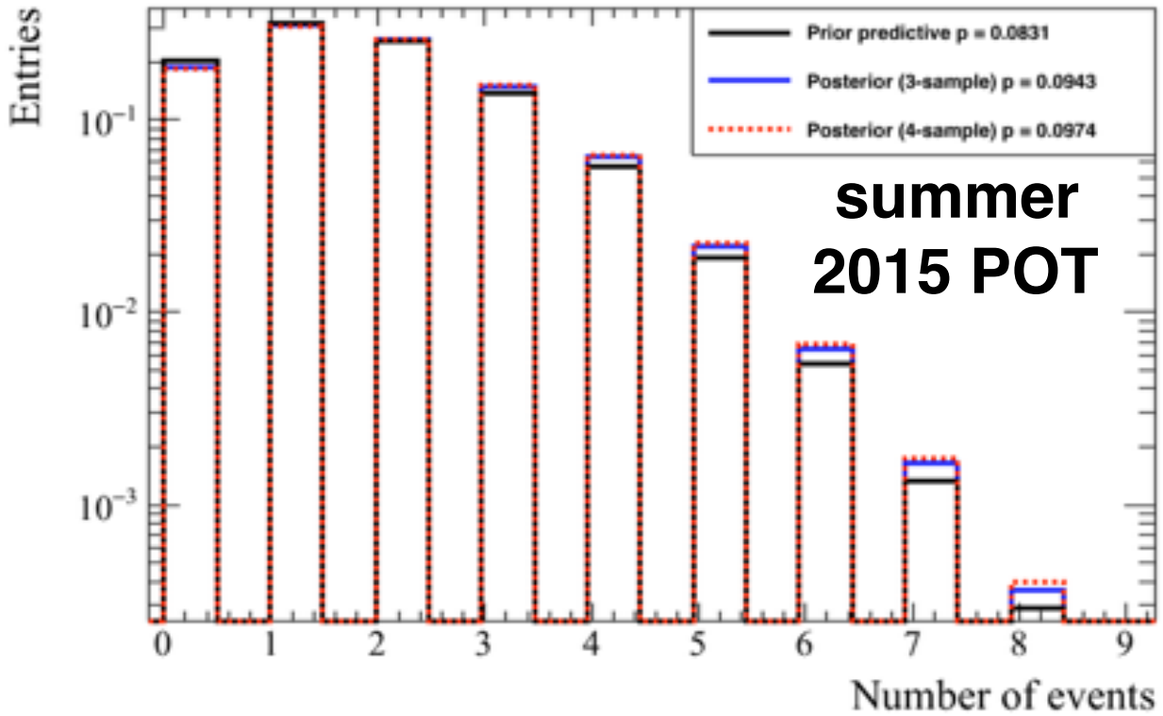
Continuation: T2K anti-ν_e appearance analysis [9,10]

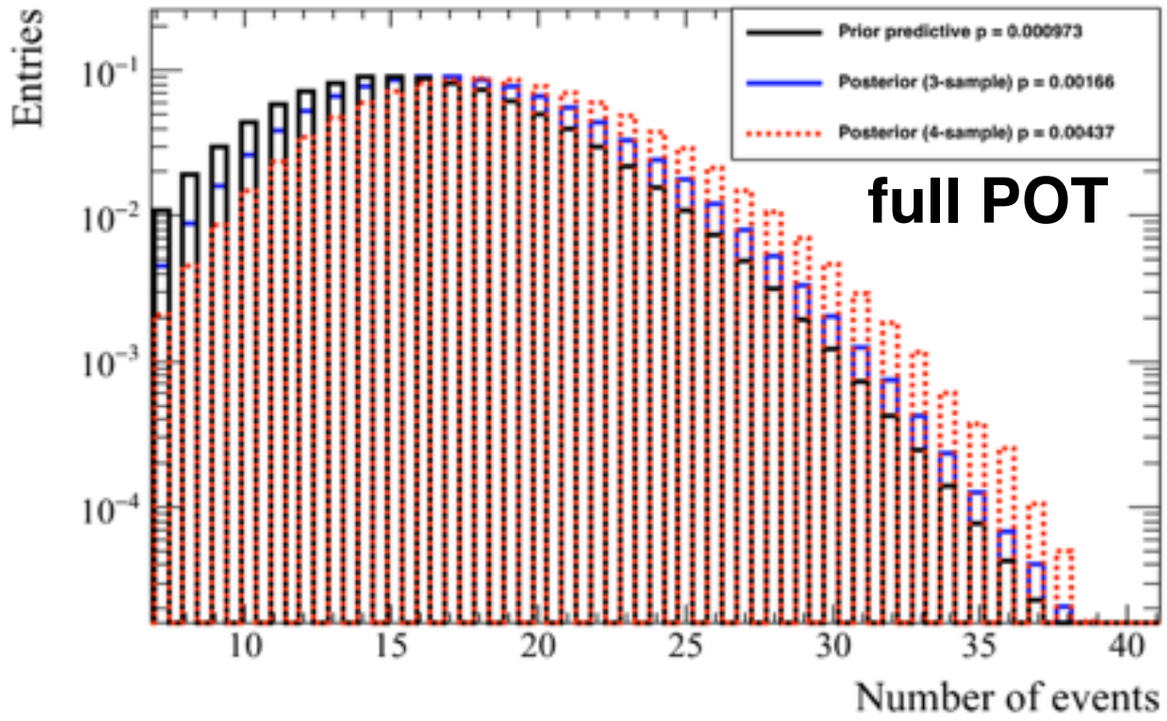
In this analysis, studies were also performed to study the **effect of sampling of the nuisance parameter space** to generate the distribution of the test statistic:

- Prior predictive: flat (uniform or gaussian) prior distributions for nuisance params
- Posterior predictive: using likelihood computed with control samples (Total 4 SK samples: μ-like, e-like in ν and anti-ν modes) to reweight the statistical throws:

$$-2\ln\lambda(N^{obs},T^{exp}) = 2\cdot\sum_{bin_i=0}^{N-1}\left(n_i^{obs}\cdot\ln(n_i^{obs}/t_i^{exp})+(t_i^{exp}-n_i^{obs})\right)$$

Rate-only statistic: # of ν_e events





Small effect with current statistics, significant with larger POT

Confidence Intervals

Confidence intervals are calculated by shifting the χ^2 distribution (grid points) wrt the best-fit value $\Delta\chi^2(\vec{\theta}) = \chi^2(\vec{\theta}) - \chi^2(\vec{\theta}_{bf})$ such that $\Delta\chi^2(\vec{\theta}) > \Delta\chi^2_{crit}$

Two methods, differing in the calculation of $\Delta\chi^2_{crit}$, are used in VALOR analyses

Constant $\Delta\chi^2$ method

Fully frequentist treatment using the Gaussian approximation, and canonical critical values (e.g. $\Delta\chi^2_{crit}$ (68% CL) = 1.00 for 1 parameter)

Feldman-Cousins method

If the gaussian regime is not satisfied, the constant $\Delta\chi^2$ method is not reliable and new critical values are calculated by generating **many toy MC experiments**:

- Produced at the oscillation hypothesis of the **grid point θ** (true osc params)
- With statistical fluctuations and systematic parameters randomised
- For each toy, χ^2 is minimised twice: fixing θ (χ^2_{true}) and fitting θ (χ^2_{bf})
- **$\Delta\chi^2 = \chi^2_{true} - \chi^2_{bf}$** is computed for the ensemble of toys and critical values are found such that $\Delta\chi^2_{crit} : \int_{-\infty}^{\Delta\chi^2_{crit}} f(\Delta\chi^2) d(\Delta\chi^2) = X\%$

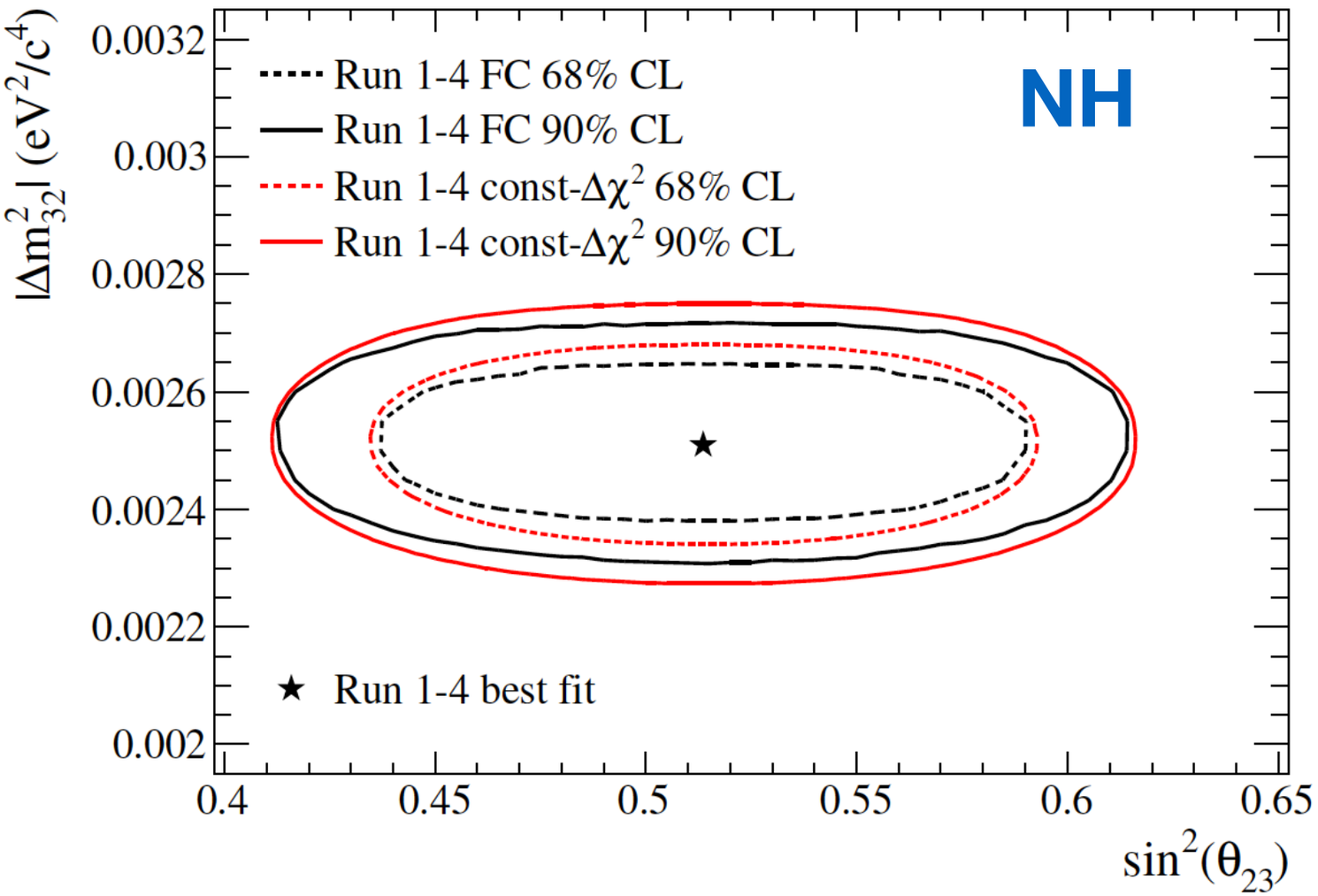
But, in the Feldman-Cousins method, there are no recommendations for including systematic uncertainties, or how to reduce the number of dimensions measured

Parameter Estimation: Confidence Intervals

2D confidence regions

Example comparing constant $\Delta\chi^2$ and Feldman-Cousins methods

T2K ν_μ disappearance analysis [1]



For the Feldman-Cousins one:

- 2 fits are done per toy expt:
  - fitting osc+systematic parameters: $\chi^2_{\text{best-fit}}$
  - fitting systematics, osc fixed to true: χ^2_{true}

i.e. χ^2 is *independently* minimised for *each* toy wrt nuisance parameters

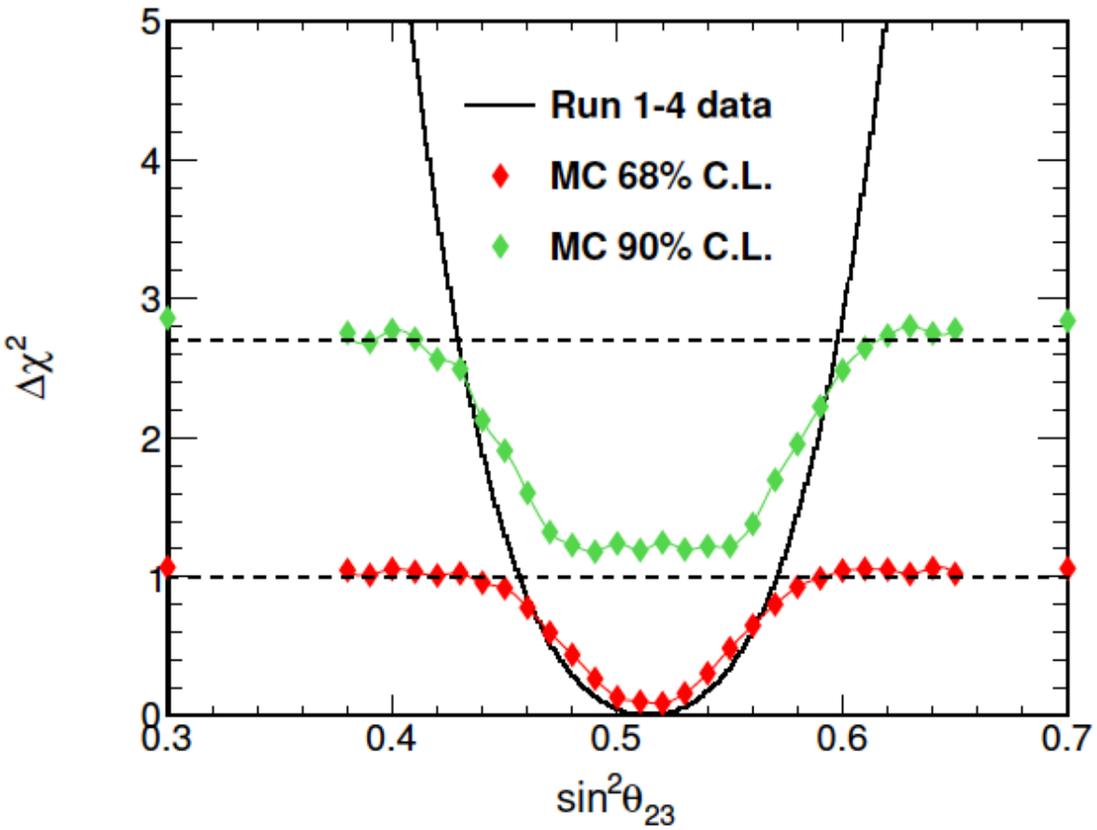
Feldman-Cousins regions are narrower at maximal disappearance, as the best-fit values of $\sin^2\theta_{23}$ pile-up at the boundary, resulting in critical $\Delta\chi2$ values smaller than the canonical ones

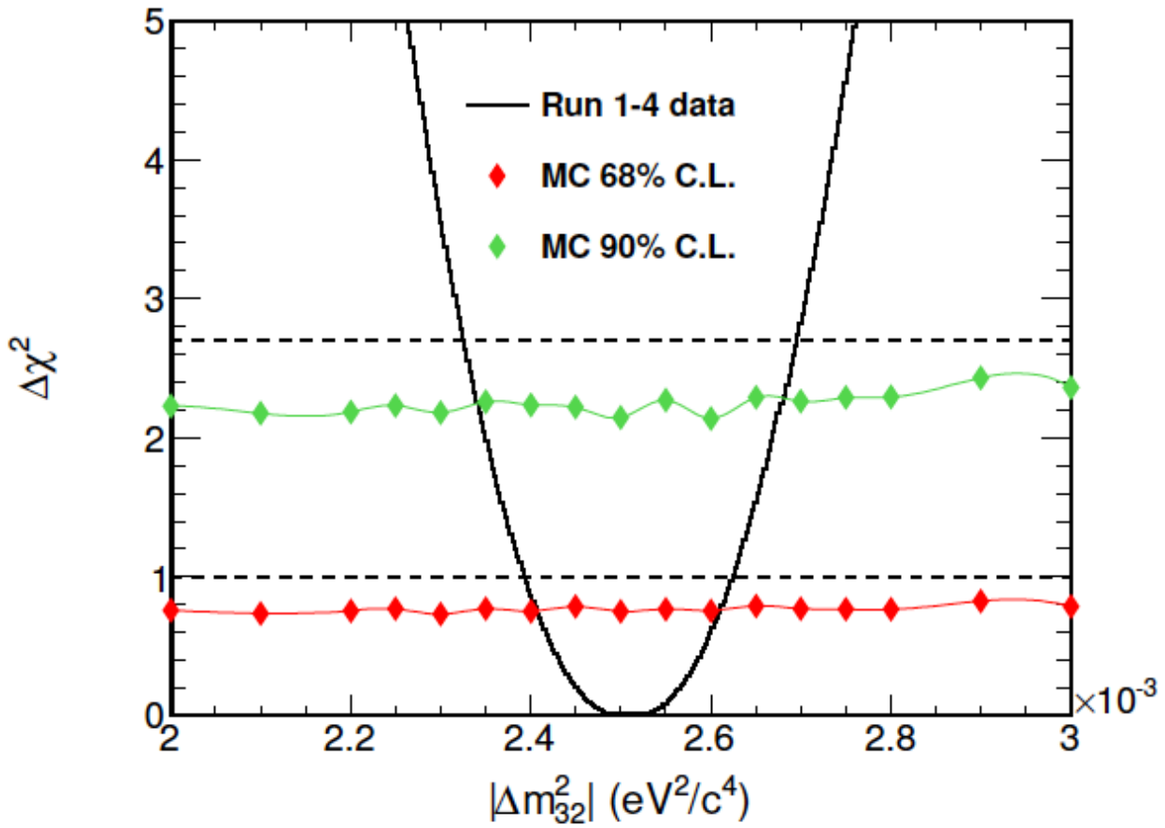
Parameter Estimation: Confidence Intervals

1D Feldman-Cousins confidence regions

Toys are generated with true input values of the nuisance oscillation parameters in proportion to: $e^{-\frac{\Delta\chi^2}{2}}$, where $\Delta\chi^2$ is the 2D $\Delta\chi^2$ surface from the data fit. In this way we take into account the uncertainty on the nuisance oscillation parameters, but also its values preferred by the data fit.

T2K ν_μ disappearance analysis [1]



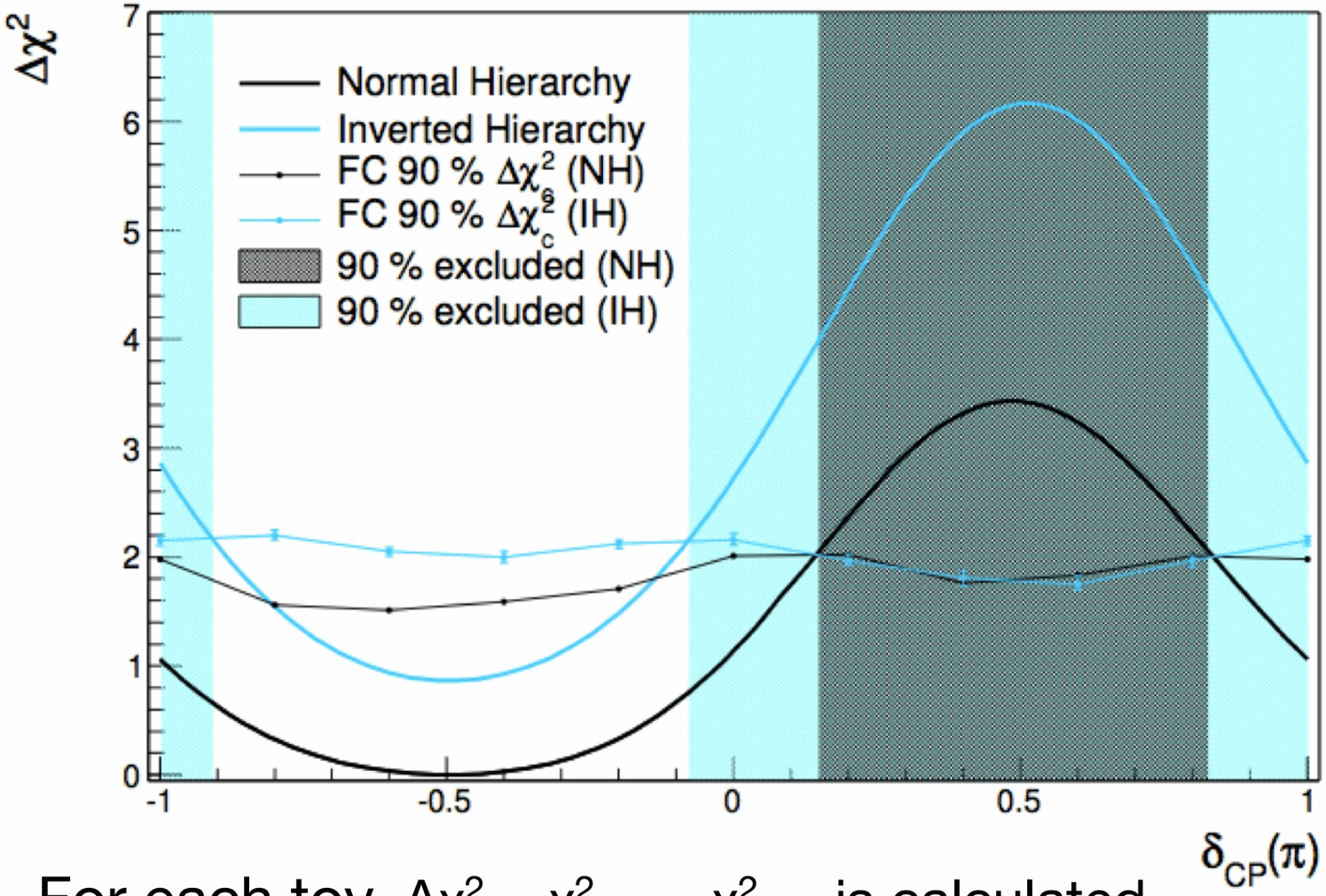


Again, critical $\Delta\chi^2$ values are smaller for FC, showing that the constant $\Delta\chi^2$ method overcovers due to boundary effects at maximal $\sin^2\theta_{23}$

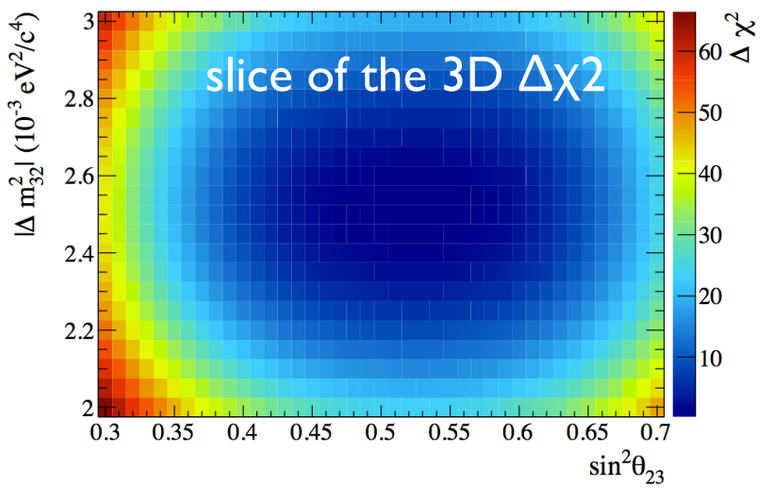
Parameter Estimation: Confidence Intervals

1D Feldman-Cousins confidence regions for δ_{CP}

T2K ν_μ disappearance + ν_e appearance analysis [3]



For the FC confidence regions, the toy MC expts are generated with $\sin^2\theta_{13}$, $\sin^2\theta_{23}$ and Δm^2 marginalised following the 3D $\Delta\chi^2$ surface from the data fit

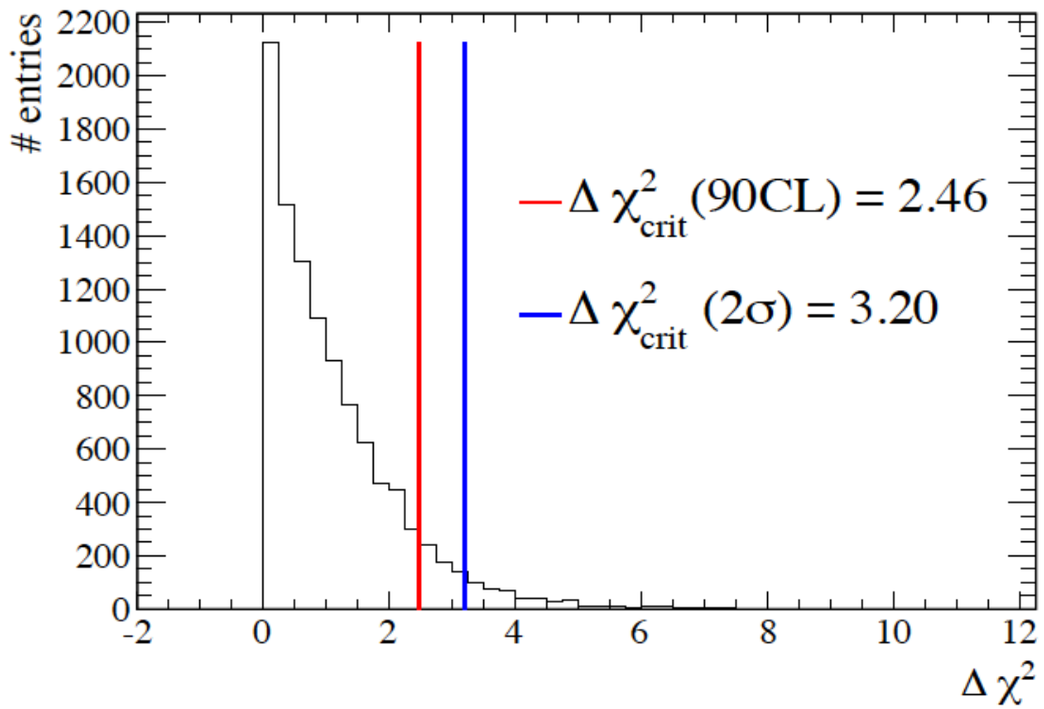


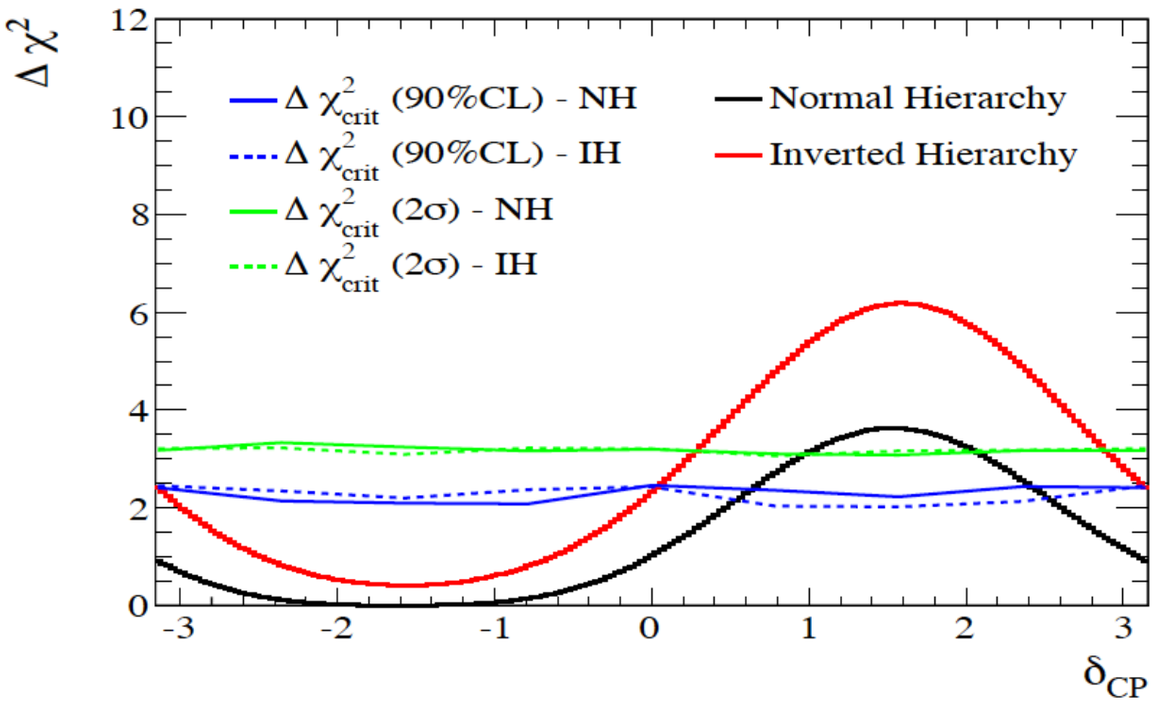
For each toy, $\Delta\chi^2 = \chi^2_{true} - \chi^2_{min}$ is calculated, where χ^2_{min} is the minimum using true MH (two fits: one in each $\sin^2\theta_{23}$ octant) and χ^2_{true} is the minimum fixing δ_{CP} (four fits: octant+MH combinations)

In more recent analyses, flat priors considered for nuisance oscillation parameters

Parameter Estimation: Confidence Intervals

Interplay of δ_{CP}-MH in the calculation of confidence regions





Joint ν + anti-ν oscillations [7,8]

Coverage studies for simultaneous fits of δ_{CP} and Mass Hierarchy show incorrect coverage when using the constant $\Delta\chi^2$ method due to different effects:

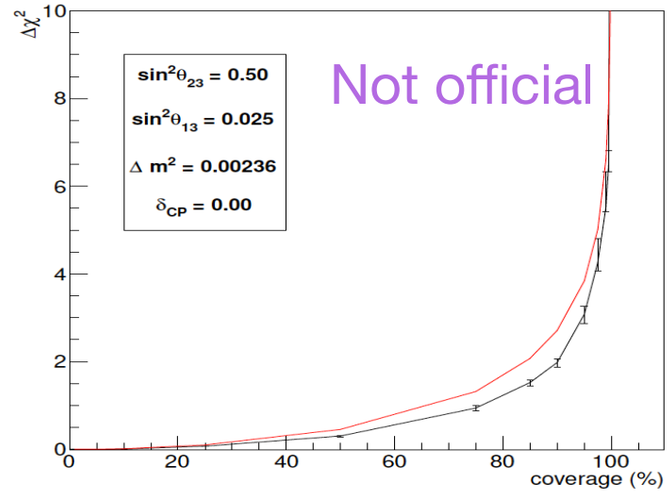
- Parameters of interest not Gaussian distributed, with no linear relation with likelihood
- "physical boundaries" for $\delta_{CP}=\pm\pi/2$, values maximizing/minimizing number of e-like events
- Degeneracy between δ_{CP} and Mass Hierarchy

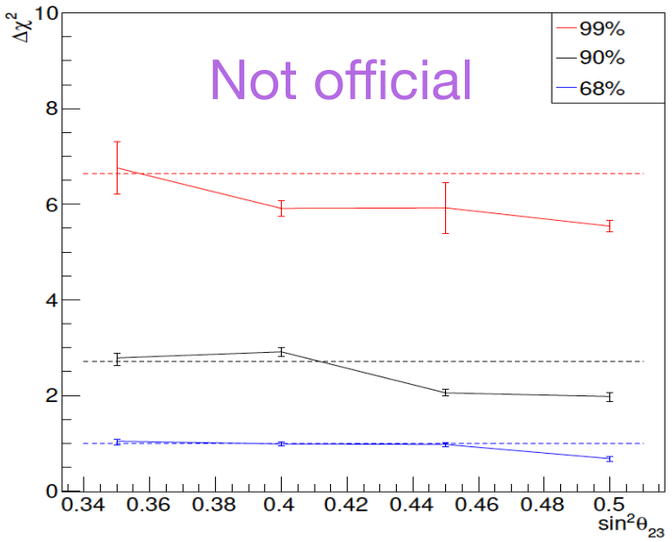
Coverage studies

Studies can be done to analyse the coverage of the constant Δχ2 method with a big number of toy MC experiments, fitting them, fixing the oscillation parameters under study (in1D or 2D) to their true values, then comparing the critical Δχ2 values to the canonical ones.

At least two effects can produce differences between the critical and canonical values of $\Delta\chi^2$ [1]:

- **Physical boundaries**: studied with toy MC experiments with statistical fluctuations only
- **Effects of systematics**: studied with toy MC exits with statistical fluctuations and systematic variations





Observations:

- Good agreement when away from physical boundaries
- Overcoverage appears due to pile up of events at physical boundaries when close to them
- If poor sensitivity to δ_{CP}, this parameter is not acting as a real degree of freedom, affecting critical $\Delta\chi^2$ values

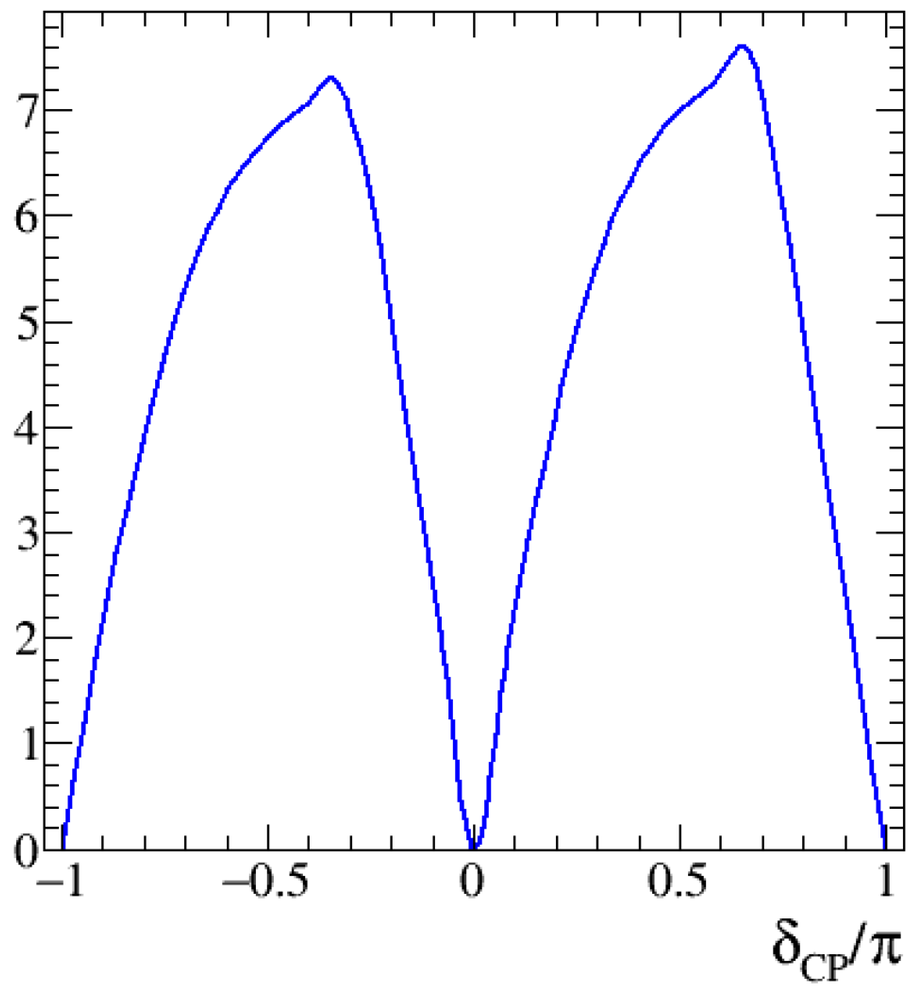
Not an official result, tests done for my PhD work

Sensitivity Studies

Example: δ_{CP} discovery sensitivity studies

Usually done by generating the special "Asimov dataset" (see [9]), toy MC spectra with nominal systematics without statistical fluctuations

Computing the confidence of rejecting the **sin(δ_{CP}) = 0 hypothesis** as a function of true δ_{CP}



With the Asimov dataset we find the median with which one would reject the hypothesised value ($\sin\delta_{CP} = 0$) under the assumption of the nominal model. But if the actual data will contain statistical fluctuations, and the observed significance is not in general equal to the median…

…is there more information that we should add in these discovery sensitivity studies? Error envelopes?

Summary/Conclusions

- In the last years VALOR has performed several neutrino oscillation analyses and extensively studied different options for
    - parameter elimination
    - construction of confidence levels
    - goodness of fit tests

- Now we are starting working on multi-detector, multi-experiment analysis

- The next steps for multi-detector, multi-experiment analyses will bring new challenges and studies!
- As well as moving from statistic limited measurements to systematic limited measurements.

- Feedback and ideas from experts are most welcome!

BACKUP

Likelihood Construction

VALOR analyses obtain measurements of the parameters of interest by comparing prediction and observation typically using a **binned likelihood-ratio method:**

The p.d.f of the total number of events (following a Poisson distribution) is

$$f_{\text{poisson}} = \frac{\left[n_{\text{tot}}^{\text{exp}}\right]^{n_{\text{tot}}^{\text{obs}}} e^{-n_{\text{tot}}^{\text{exp}}}}{n_{\text{tot}}^{\text{obs}}!}$$

with n^{obs}(n^{exp}) the number of observed(expected) events

For N kinematical bins, there are N possible ways to place an event in a bin (with a probability for the bin i given by n_i^{exp}/n_{tot}^{exp}), following a multinomial p.d.f distribution:

$$f_{\text{multinomial}} = n_{\text{tot}}^{\text{obs}}! \prod_{i=0}^{N-1} \frac{1}{n_i^{\text{obs}}!} \left[\frac{n_i^{\text{exp}}}{n_{tot}^{\text{exp}}}\right]^{n_i^{\text{obs}}}$$

The joint p.d.f is the product of both:

$$f_{\text{joint}} = e^{-n_{\text{tot}}^{\text{exp}}} \prod_{i=0}^{N-1} \frac{1}{n_i^{\text{obs}}!} \left[n_i^{\text{exp}}\right]^{n_i^{\text{obs}}}$$

and can be divided by a factor independent of oscillation parameters:

$$f_0 = e^{-n_{\text{tot}}^{\text{obs}}} \prod_{i=0}^{N-1} \frac{1}{n_i^{\text{obs}}!} \left[n_i^{\text{obs}}\right]^{n_i^{\text{obs}}}$$

giving the likelihood ratio:

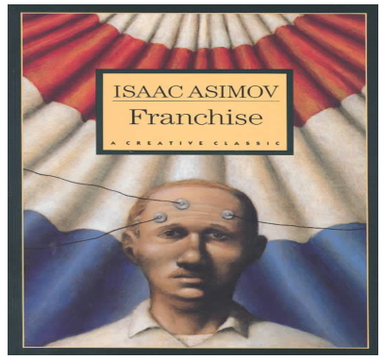
$$\lambda = e^{n_{\text{tot}}^{\text{obs}} - n_{\text{tot}}^{\text{exp}}} \prod_{i=0}^{N-1} \left(\frac{n_i^{\text{exp}}}{n_i^{\text{obs}}}\right)^{n_i^{\text{obs}}}$$

Toy Experiment Generation

A toy experiment for a given physic hypothesis is generated by calculating

- the predicted spectra $n^{pred}_{d;b;s}(r;\vec{\theta};\vec{f})$
- drawing random numbers $n^{obs;toy}_{d;b;s}(r)$ following a Poisson distribution with mean value equal to $n^{pred}_{d;b;s}(r;\vec{\theta};\vec{f})$
- if the exposure of the MC samples used in this calculation is not much larger than the experimental one, a second statistical fluctuation due to finite MC statistics must be included

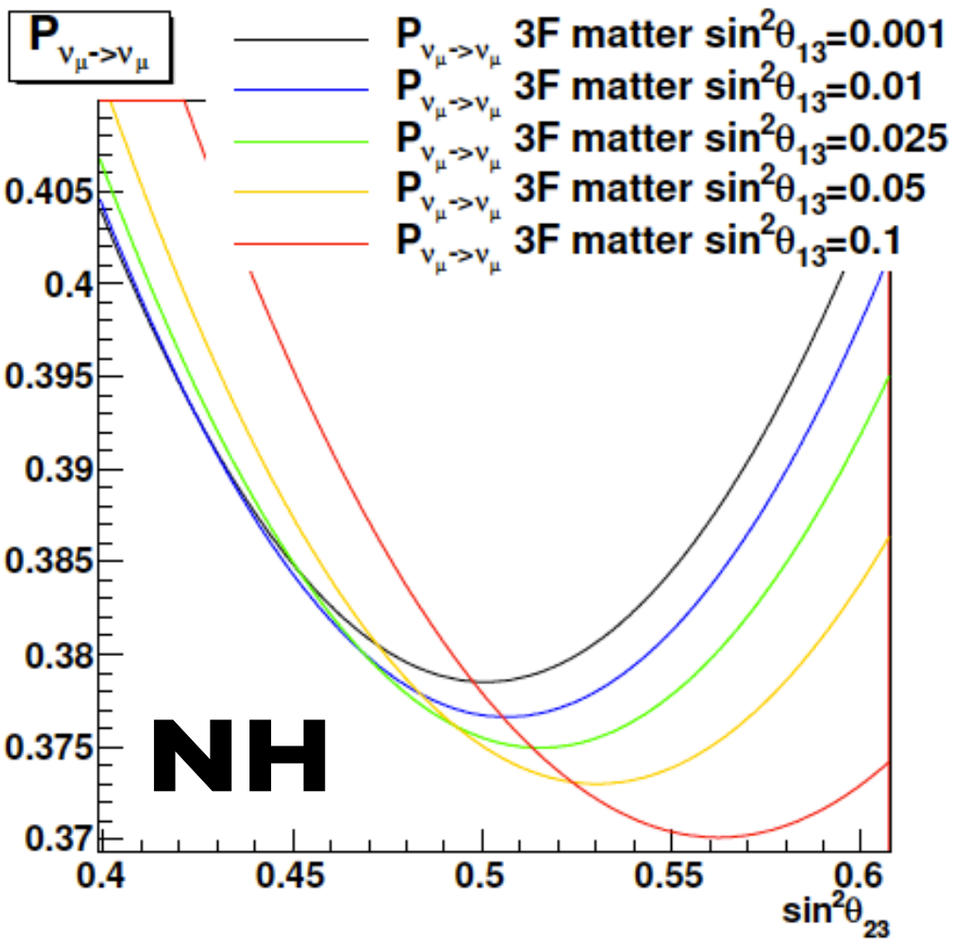
A special toy is often used in analyses, called the Asimov dataset, which is the dataset with all observed quantities equal to the expected values, and serves as median significance of many toys

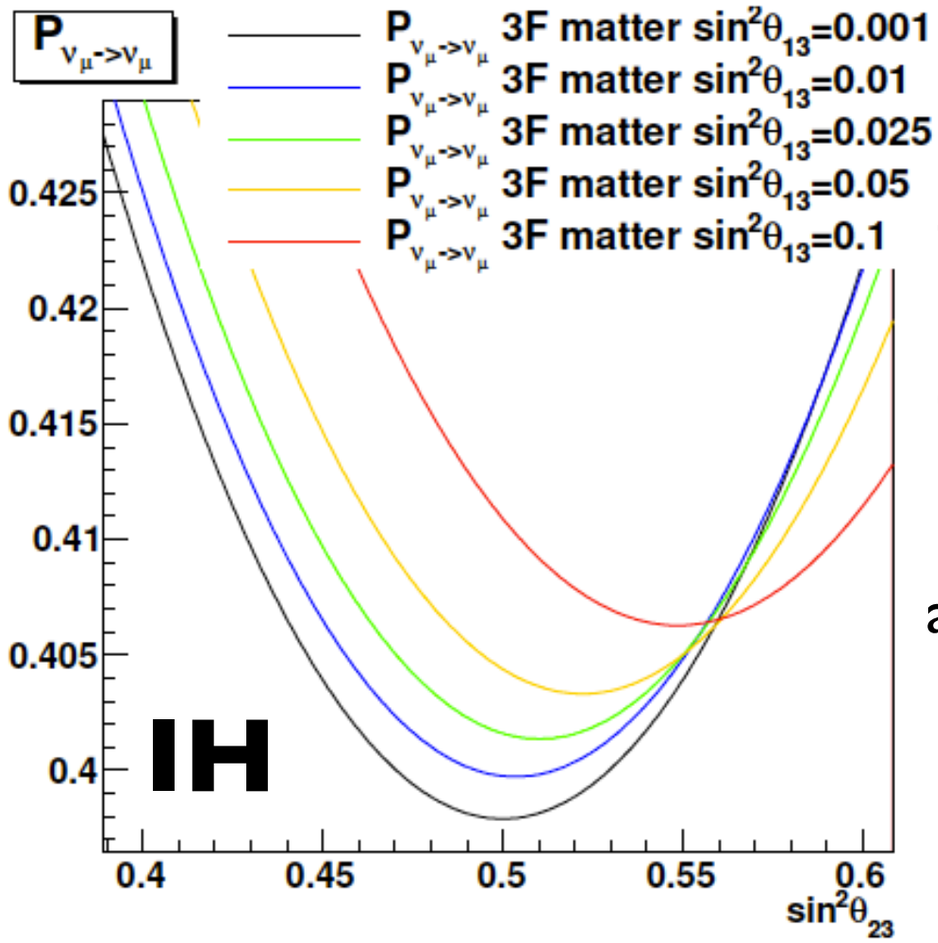


The procedure of generating toy experiments must **respect the correlations between the randomised parameters**. We do this by using the Cholesky decomposition, which is a special case of a LU factorisation of a matrix C done by finding a lower triangular matrix L such that: $C = L \cdot L^T$

Once L is found, a vector of correlated values (x) can be calculated from any vector of uncorrelated vales (u) $\vec{x} = L \cdot \vec{u}$

Maximal disappearance vs mixing





The value of maximal disappearance (minimum surv. prob) depends on the value of sin2θ13 and it is sin2θ23 > 0.5 if sin2θ13 > 0

$$P(\nu_\mu \to \nu_\mu) \approx 1 - 4\sin^2\left(\frac{\Delta m^2_{31} L}{4E}\right) c^2_{13} s^2_{23} \left[s^2_{12} c^2_{23} + c^2_{12} s^2_{13} s^2_{23} + c^2_{12} c^2_{23} + s^2_{12} s^2_{13} s^2_{23}\right]$$

$$= 1 - 4\sin^2\left(\frac{\Delta m^2_{31} L}{4E}\right) c^2_{13} s^2_{23} \left(s^2_{12} + c^2_{12}\right) \left[c^2_{23} + s^2_{13} s^2_{23}\right]$$

$$= 1 - 4\sin^2\left(\frac{\Delta m^2_{31} L}{4E}\right) c^2_{13} s^2_{23} \left[c^2_{23} + s^2_{13} s^2_{23}\right]$$

$$y \approx 1 - 4Ax(1-k)\left[(1-x) + kx\right] = 1 - 4Ax(1-k) + 4Ax^2(1-k)^2$$

$$\frac{dy}{dx} = -4A(1-k) + 8Ax(1-k)^2 \qquad x_{min} = \frac{1}{2(1-k)} \qquad \mathrm{k} = s^2_{13} = \sin^2(\theta_{13}).$$

Goodness of fit tests

Goodness-of-fit tests complement best-fit parameter estimation, analysing the agreement between data and (best-fit) model

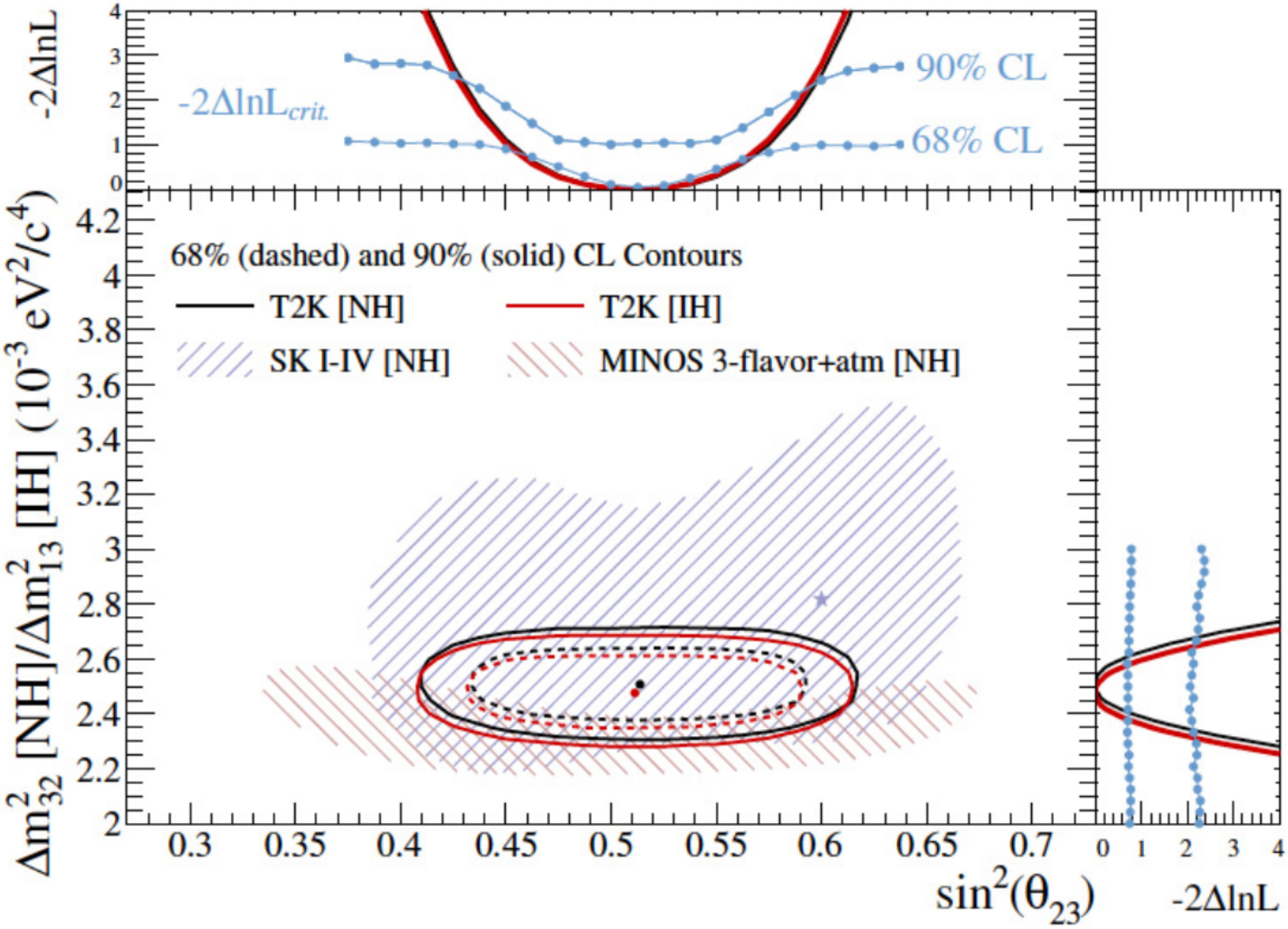
The advantage of the likelihood ratio method is that, in the large sample limit, $-2ln\lambda(\vec{\theta};\vec{f})$ has a χ^2 distribution and can be used as goodness-of-fit test

p-value calculation

Probability to obtain a measurement as or more extreme than the data with the null hypothesis (e.g. no anti-ve appearance or best-fit to the data tested)

- Generate many toy MC experiments with the null hypothesis (with statistical fluctuations and systematic variations)
- Defining a coarse binning to have enough statistics in each bin, calculate test statistic for each toy MC experiment: χ^2_{gof}
- Calculate test statistic for data: $(\chi^2_{gof})_{data}$
- Compare the distribution of test statistic for toys with the one for the data and find the proportion of experiments for which $\chi^2_{gof} > (\chi^2_{gof})_{data}$

VALOR T2K ν_μ disappearance



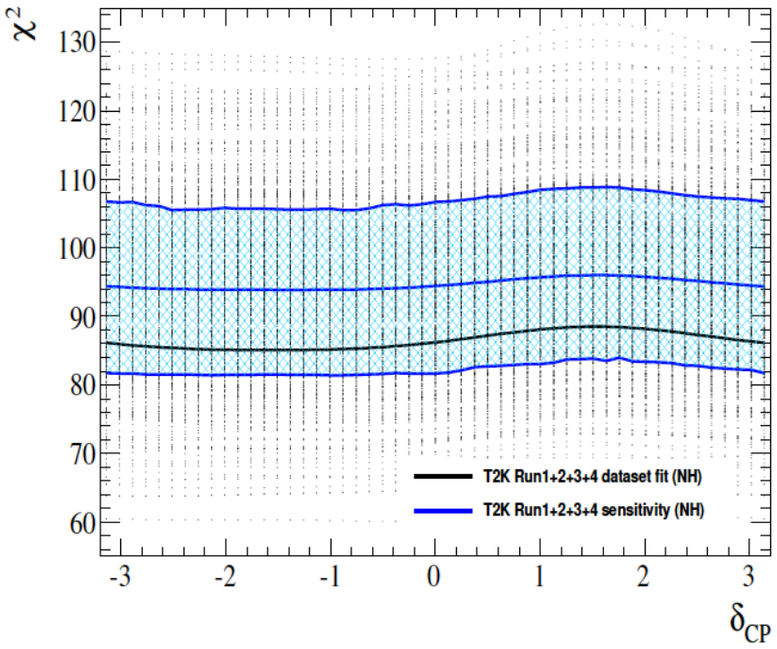
Phys. Rev. Lett. 112, 181801 (2014)

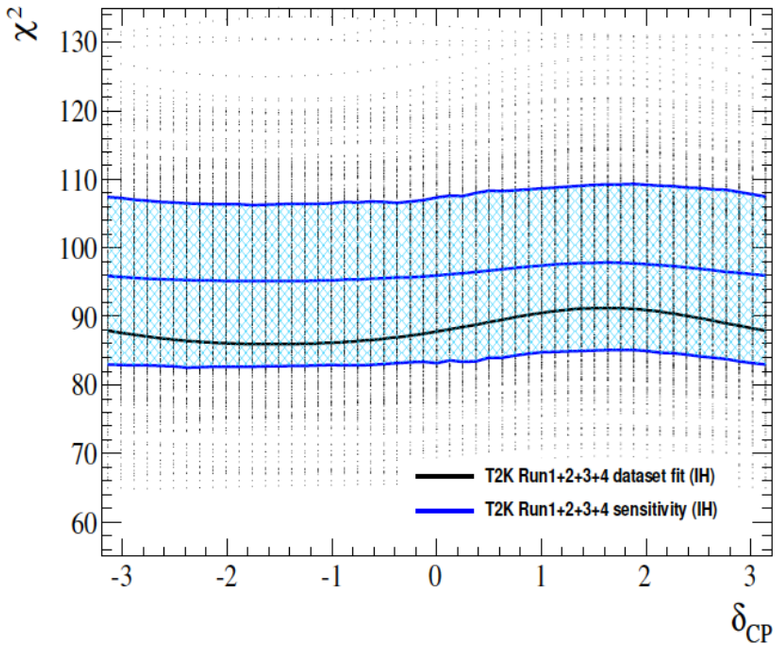
Tom Dealtry's PhD work, 2014

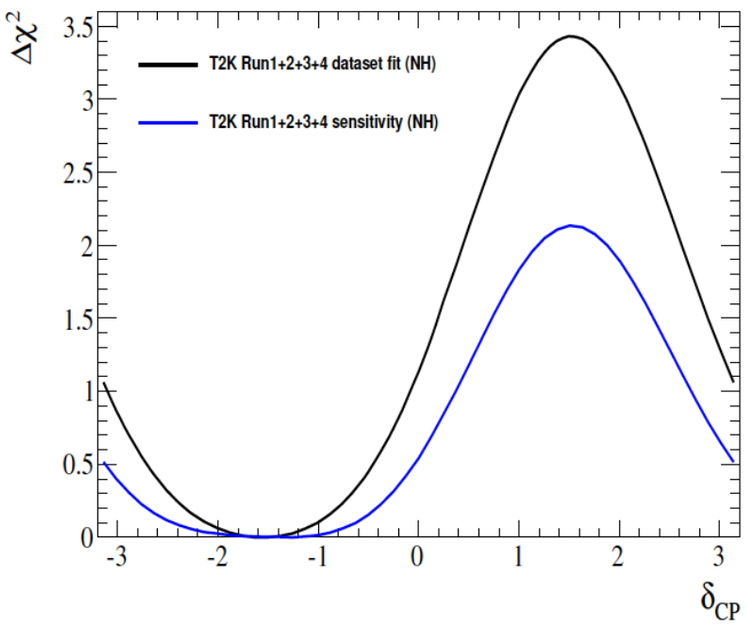
Parameter Estimation: Confidence Intervals

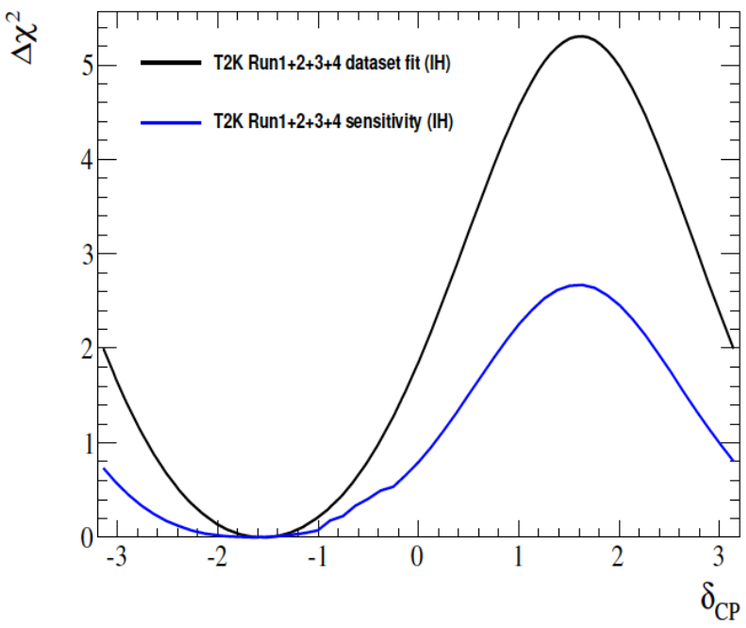
1D Feldman-Cousins confidence regions for δ_{CP}

T2K νμ disappearance + νe appearance analysis [3]









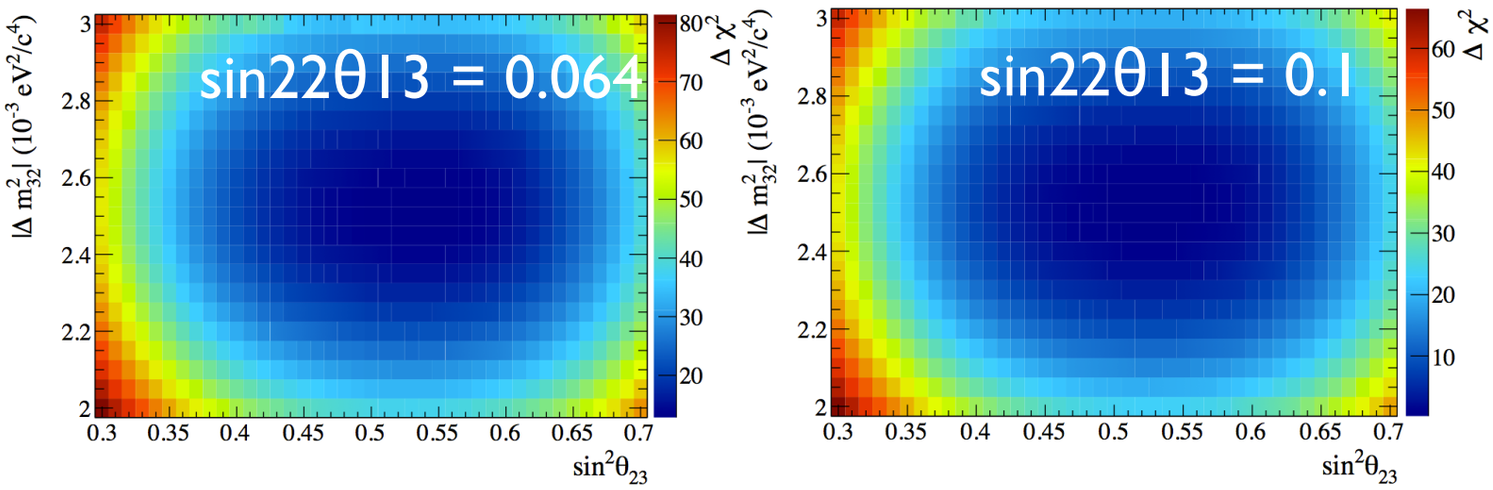
Profiled Δχ2 as a result of the data fit compared to the sensitivity as the averaged profiled Δχ2 and 1σ error bands calculated with many toy experiments

FC calculation for Δχ2 vs δCP

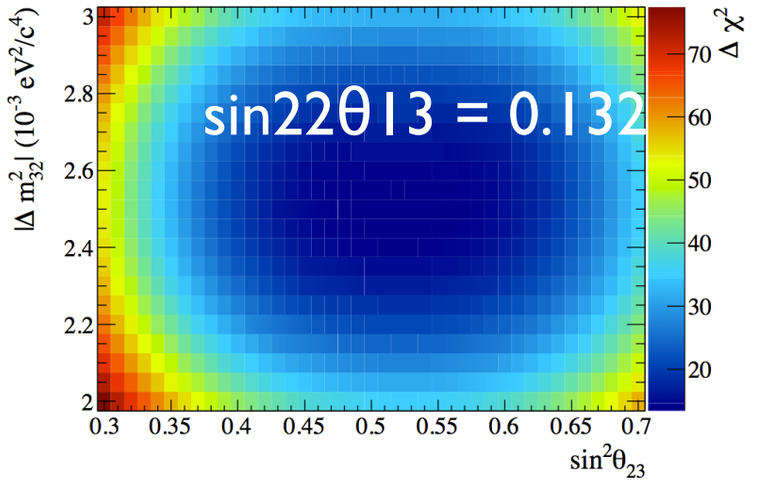
À la Feldman-Cousins and Cousins-Highland

Toy experiments created:
- at each value of δCP
- with stat. fluct.
- varying systematics
- varying sin2θ23-Δm2-sin2θ13

according to Lconst from 3D Δχ2 surface



2D slices of the 3D surface



Fit 1: 4+N fit, NH, 1st octant
Fit 2: 4+N fit, NH, 2nd octant

χ2min

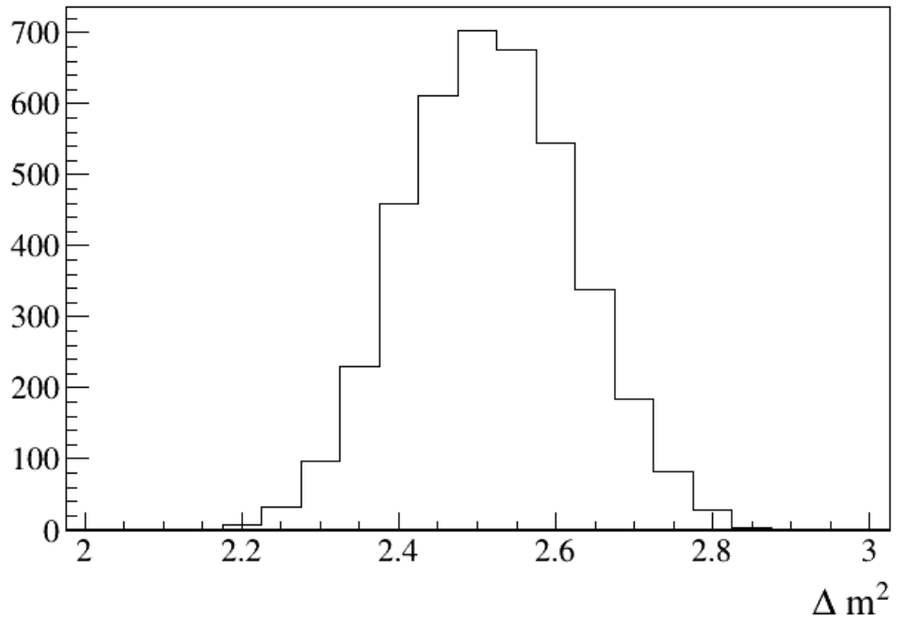
Fit 3: δCP fixed, 3+N, fit NH, 1st octant
Fit 4: δCP fixed, 3+N, fit NH, 2nd octant
Fit 5: δCP fixed, 3+N, fit IH, 1st octant
Fit 6: δCP fixed, 3+N, fit IH, 2nd octant

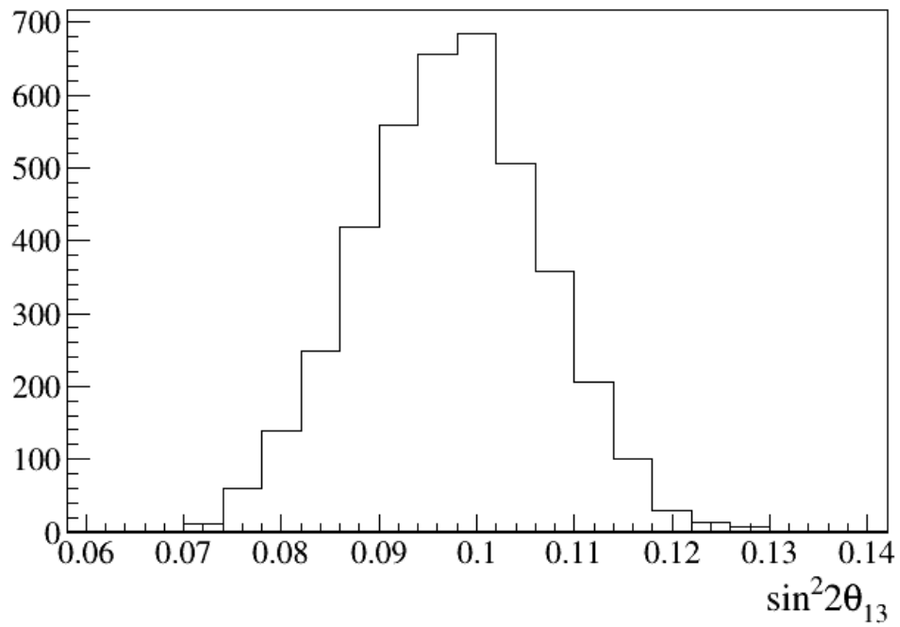
χ2true

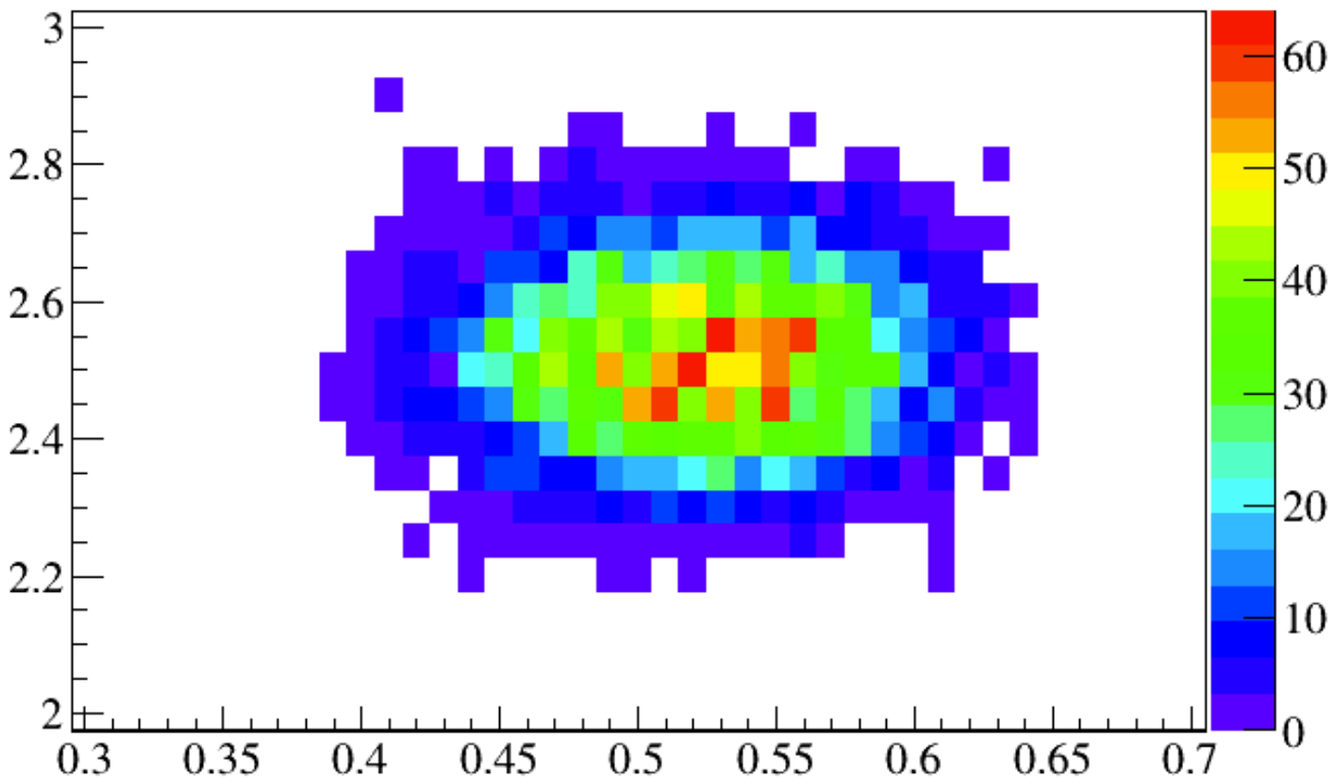
Calculate
Δχ2 = χ2 true - χ2 min
for each toy and
Δχ2 c with 68% and 90%
of the toys
with Δχ2< Δχ2 c

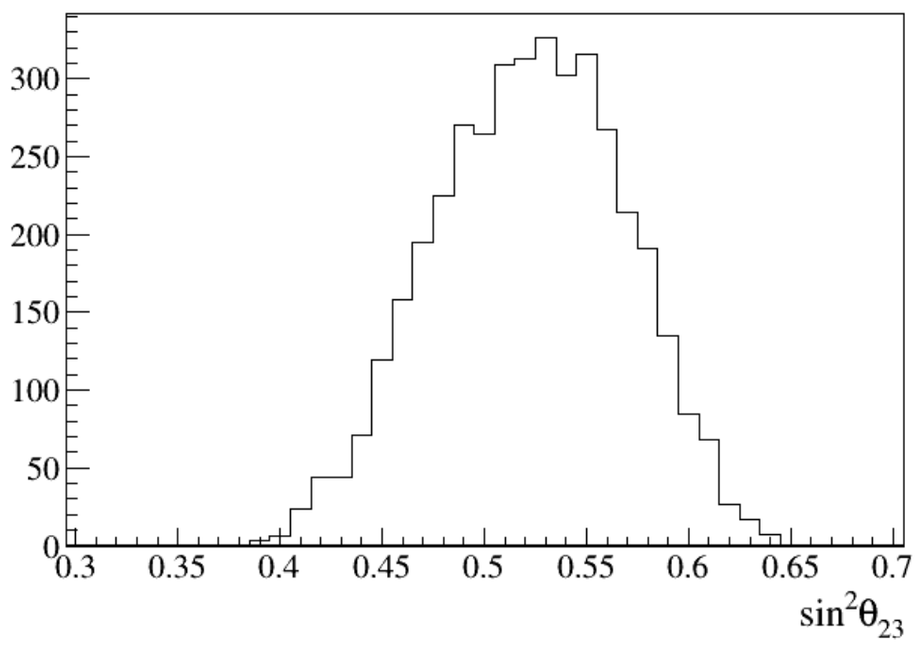
FC calculation for Δχ2 vs δCP

Example: distribution of values for 4k toys
for point 0 (δ=-π)
with the 3D Δχ2 surface









Goodness of fit tests

T2K anti-ν_e appearance analysis [9,10]

Two test statistics studied:

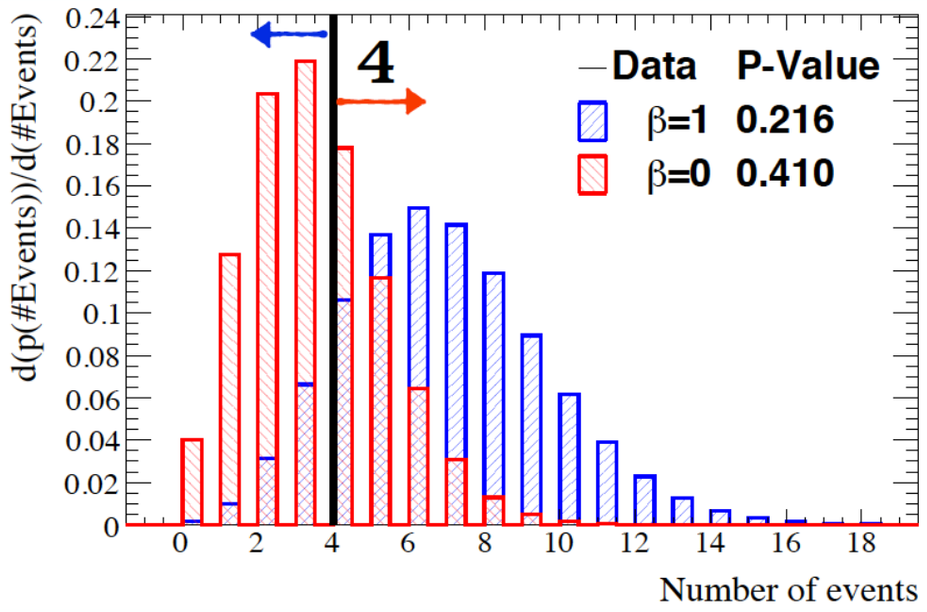
- 100k toy expts generated:
  - for null hypothesis: no anti-ν_e appearance, i.e. β=0 in
    $$P_{osc}(\bar{\nu}_\mu \to \bar{\nu}_e) = \beta P_{osc}(PMNS)$$
  - with systematic variations
  - weighted by T2K data samples
- For each toy expt, likelihood calculated and used to weight 10k statistical fluctuations of the toy expt
- Total distribution fully sampling statistical distribution and systematic and oscillation space

Bayes factor (marginal likelihood ratio):

B01 = 2.62, β=0 weakly preferred

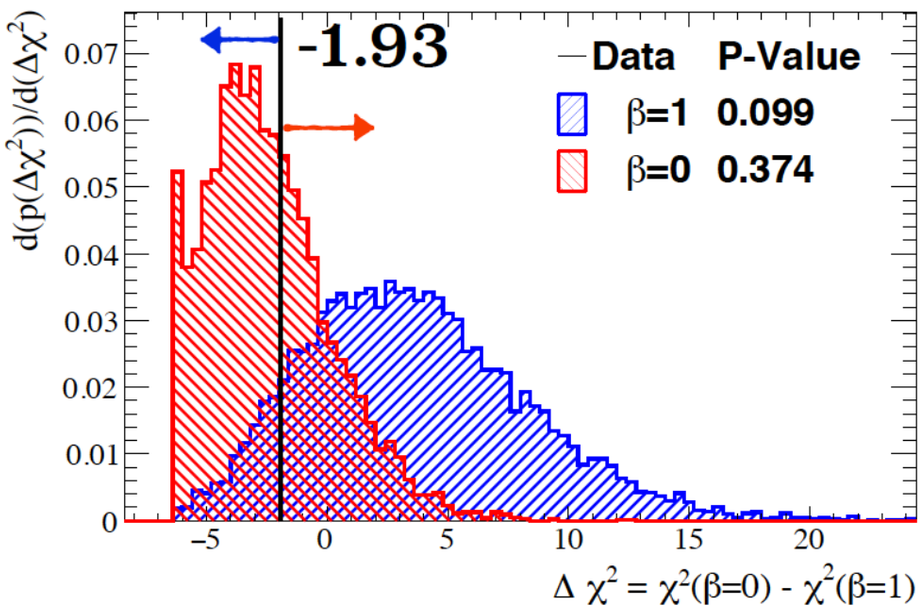
$$\frac{P(\beta=0|D)}{P(\beta=1|D)} = \frac{\pi(\beta=0)}{\pi(\beta=1)} B_{01} = \frac{\pi(\beta=0)}{\pi(\beta=1)} e^{-0.5\times\Delta\chi^2_{marg}}$$

Rate-only statistic: # of ν_e events



Rate+shape statistic:

$$\Delta\chi^2 = \chi^2(\beta=0) - \chi^2(\beta=1)$$



Goodness of fit tests

P Value:

The probability to make a measurement as or more extreme than seen in data given the null hypothesis is true.

Null Hypothesis:

No ν_eBar appearance ($\beta = 0$)

($P_{osc}(\nu_\mu Bar \rightarrow \nu_e Bar) = \beta\ P_{osc}(PMNS)$)

P Value

(1) Generate a fake data set T for null hypothesis
(2) Compute test statistic S for T
(3) Fill distribution with ensemble of statistics S_i
(4) Calculate data statistic S_D
(5) Compare S_D with S_i

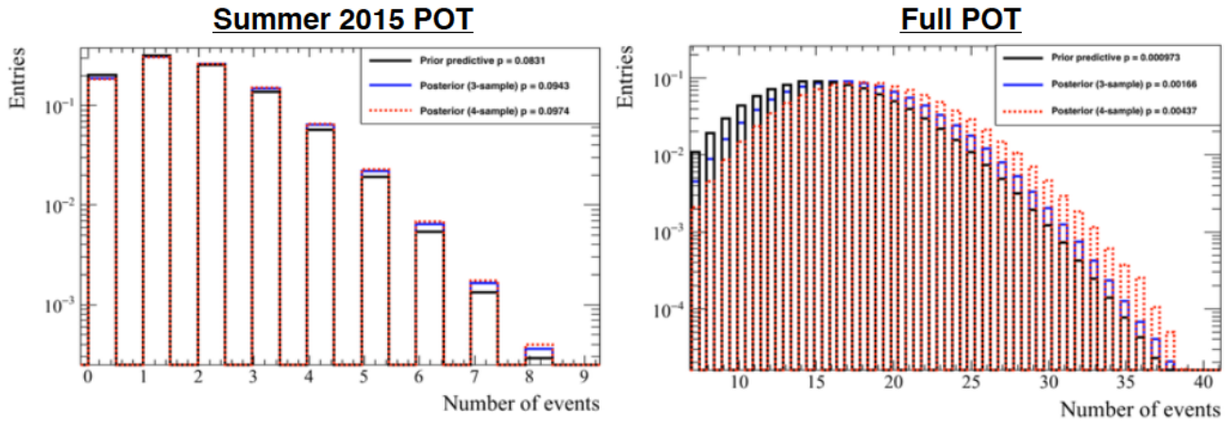
From [10]

Rate only analysis

"Data" = Asimov (MC) data

(1) Throw expectation T_{exp} from priors (nuisance parameter fluctuations)
(2) Likelihood weight $L = L(T_{data}|T_{exp})$
(3) Statistical fluctuation of T_{exp}: T_{obs}
(4) Distribution: 10k T_{obs} from 100k T_{exp}
(5) T_{obs} weighted by L

Statistic: #Events in ν_eBar sample



Rate + shape

"Data" = Asimov (MC) data

- Test: same as rate-only
- Toy weight from real data
- Marginalisation penalty?
  - "Real data"
  - MC throw data (T_{data})

Statistic: $\Delta\chi = \chi^2(\beta=1) - \chi^2(\beta=0)$ (marginalised)