Systematic uncertainties in (some) neutrino experiments

C. Bronner
Nov. 2nd, 2021

Introduction

- Original request was about the incorporation of systematic uncertainties in neutrino experiments. However, that’s a vast domain and difficult to talk about technical details of analysis or experiments I’m not part of or don’t have regular contact with

- As a result, focusing on methods for systematic uncertainties in two oscillation experiments I know well (T2K and Super-Kamiokande atmospheric) with some points about NOvA.
  Apologies to the large number of neutrino experiments not considered here

- The experiments covered are currently limited by statistics, and systematic uncertainties have had a limited impact so far. As a result, details of systematics implementation did not always matter that much.

- Less and less true as more statistics is being accumulated and will definitely not be true for high statistics next generation experiments (Hyper-Kamiokande, DUNE)
  → proper evaluation and implementation of systematic uncertainties is becoming an important topic in this field

Introduction

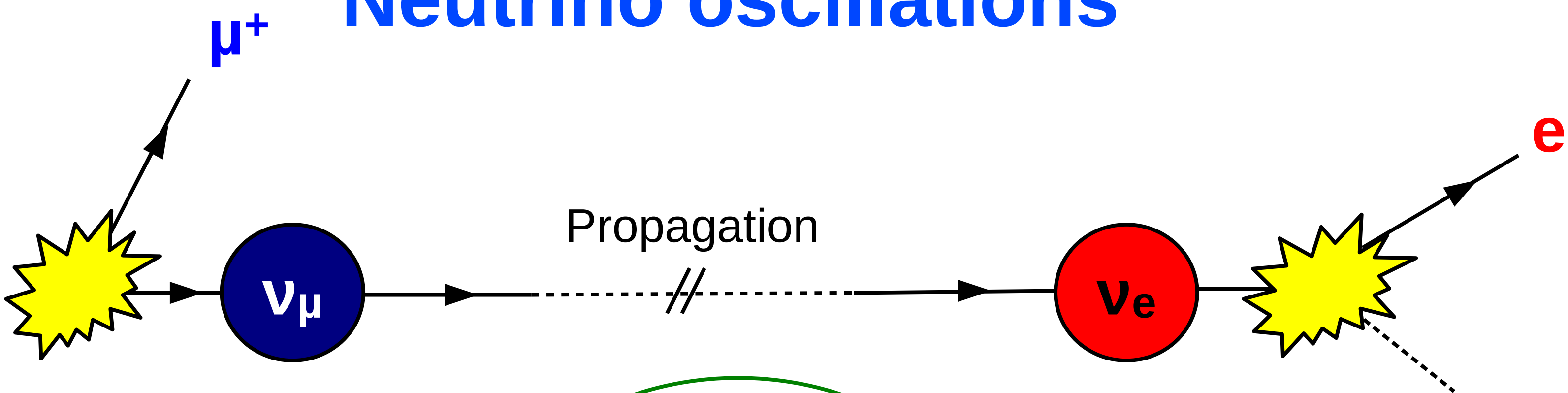
- Will present methods used to incorporate systematic uncertainties into neutrino oscillation analyses, problems we're dealing with and ideas to go beyond. Comments and suggestions of alternative methods very welcome

- A number of those methods likely to be used in other fields as well, with potentially different names

- Will be using names commonly used in T2K/Super-K (but might be our jargon), trying to introduce them when they first appear

- Also no particular formal training in statistics, point of view will be the one of a physicist having worked on those analyses and run into a number of statistical issues as a result

Neutrino oscillations



Flavor eigenstates (interaction)

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \times \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

Mass eigenstates (propagation)

Mixing (or Pontecorvo-Maki-Nagawa-Sakata) matrix
link between the two sets of eigenstates

$P(\nu_\alpha \rightarrow \nu_\beta)$ oscillates as a function of distance L traveled by the neutrino

- Amplitude of oscillations depends on the mixing matrix U
- Phase of the oscillation depends on energy and difference of mass squared: $\mathbf{\Delta m^2_{ij} L/E}$

($\Delta m^2_{ij} = m^2_i - m^2_j$)

Standard neutrino oscillations
What do we measure?

In practice, for neutrino oscillations:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

"Atmospheric" "Reactor" "Solar"

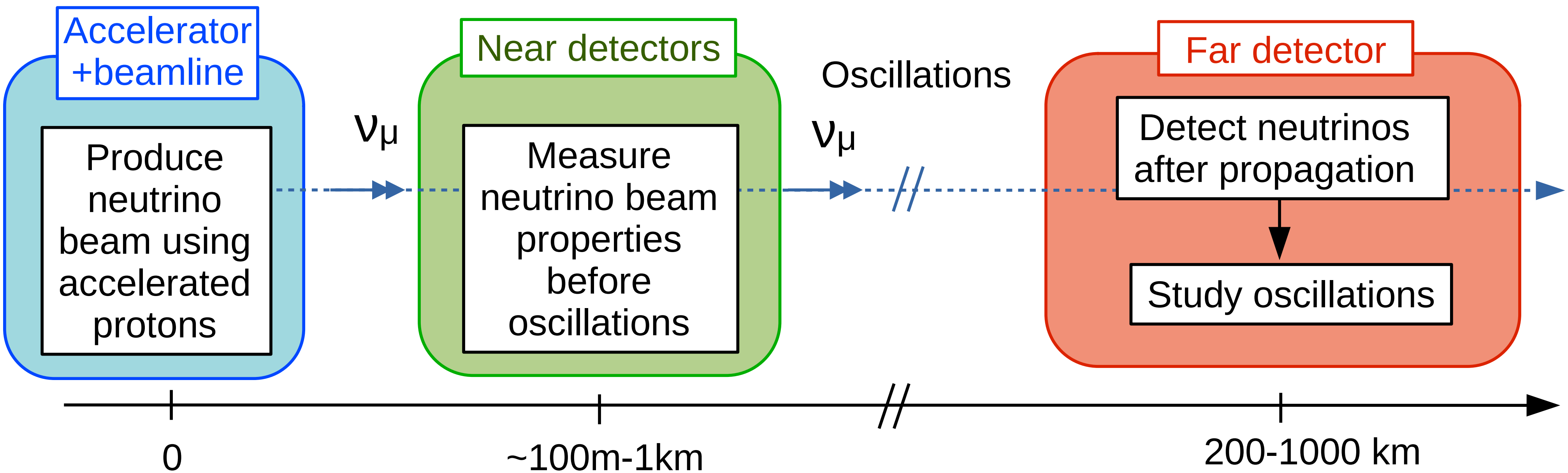
($c_{ij} = \cos(\theta_{ij})$, $s_{ij} = \sin(\theta_{ij})$)

$P(\nu_\alpha \rightarrow \nu_\beta)$ depends on **6 parameters**:
- 3 mixing angles $\boldsymbol{\theta_{12}}$, $\boldsymbol{\theta_{23}}$, $\boldsymbol{\theta_{13}}$
- 2 independent mass splittings $\boldsymbol{\Delta m^2_{ij}}$
- 1 complex phase, the CP phase **δ**

3 main open questions:
- **Mass hierarchy**: sign of Δm^2_{32}
- **Octant of θ_{23}**: $\theta_{23}=45°$?
- **CP symmetry**: $\sin(\delta)=0$?

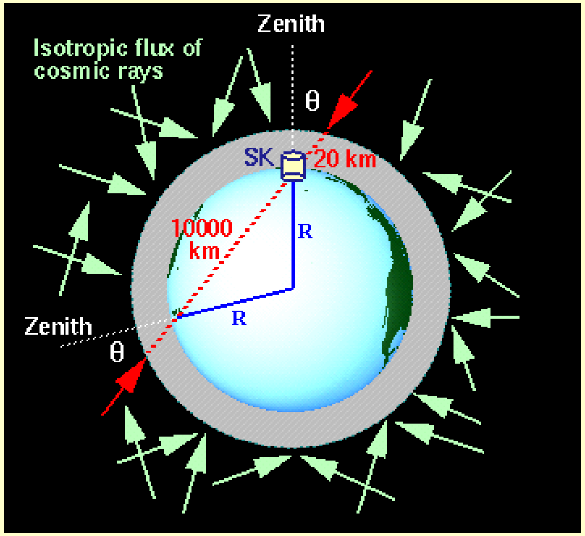
Experiments concept

Long baseline oscillation experiments (T2K, NOνA)



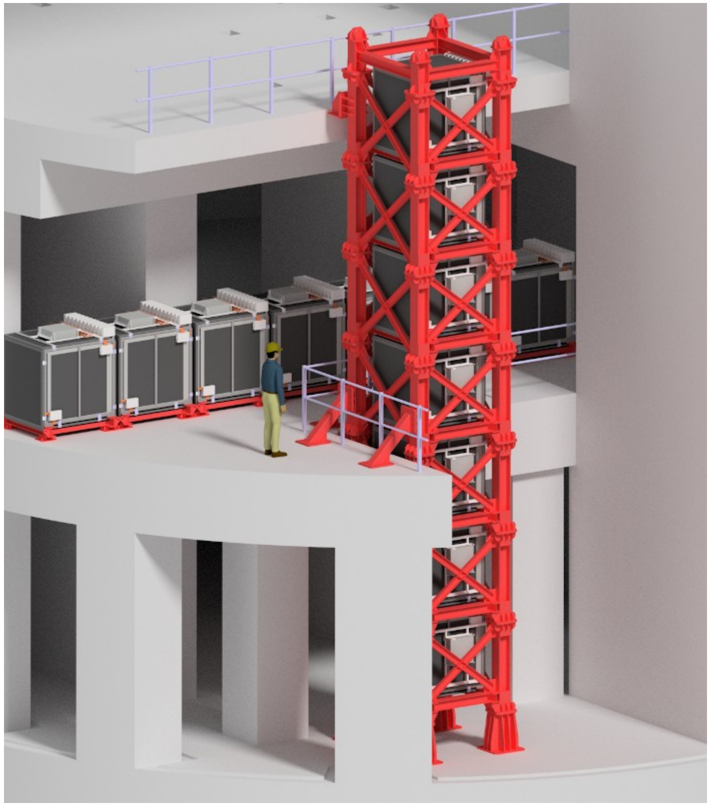
Atmospheric neutrino oscillation (Super-K)

- Atmospheric ν produced isotropically around the Earth by collisions of cosmic ray with the atmosphere
- Detector on the Earth surface: propagation distance (L) depends of direction of arrival of neutrino
- Wide range of energies
- Can study number of events detected as a function of reconstructed L and E to measure oscillation patterns
- Down-going (small L) events did not have time to oscillate for high enough E: provide reference



Neutrino oscillation experiments
Systematic uncertainties

- 3 types of systematics: neutrino flux, neutrino interactions and detector systematics



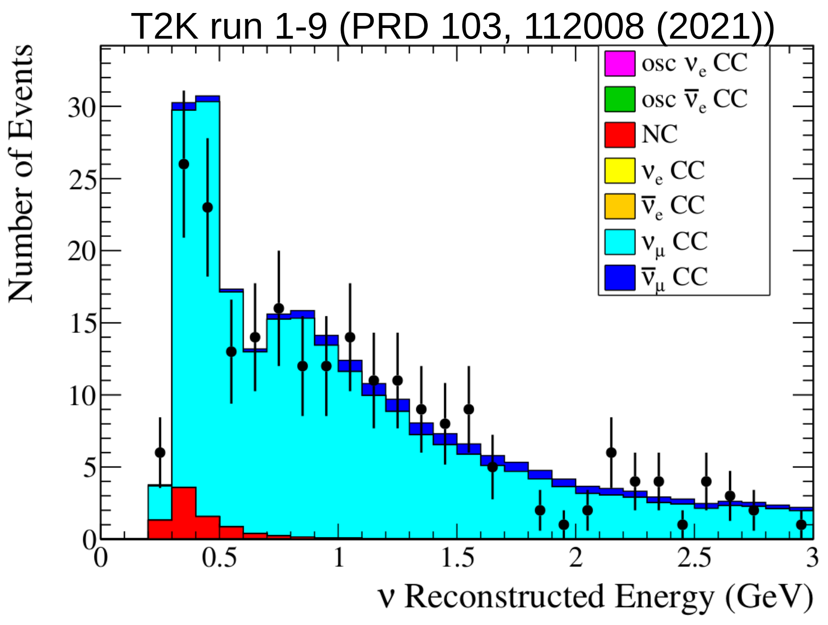
- Difficult to know precisely initial neutrino flux: no electric charge and very low interaction probability. Even for beam produced by accelerator, predicting neutrino flux from flux of proton beam hitting target is non-trivial.
  → relatively large initial flux uncertainty

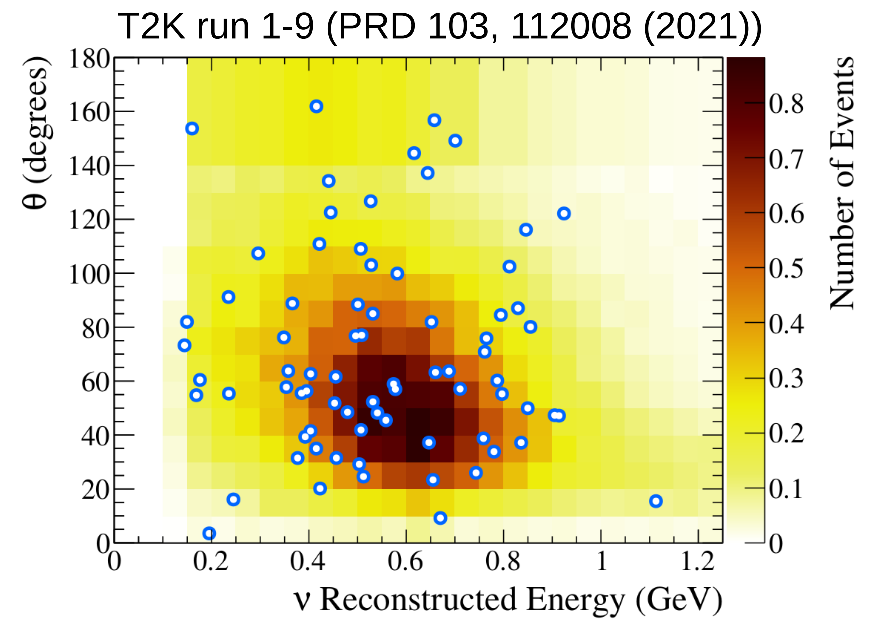
- Measurements done before oscillations can reduce them, but measures number of events
  → flux and interaction uncertainties become correlated

- Significant uncertainties on models for interactions of neutrino on nuclear target as well.
  Have complicated effects on predicted kinematics for the interactions, and its relationship with the true neutrino energy on which oscillation depends
  → cannot be modeled by simple uncertainties on number of events

Analysis method
Likelihood

Maximum likelihood analysis. Generally:
- bin events as function of some reconstructed quantities ("analysis bins")
- likelihood as product of probability of observed number in each bin according to Poisson probability based on predictions





$$\mathcal{L} = \prod_{bins} \left(\frac{Poisson(N^i_{obs}, N^i_{pred}(o,f))}{Poisson(N^i_{obs}, N^i_{obs})} \right) \times \mathcal{L}_{penalty}(o,f)$$

N^i_{obs}: observed nb of events in bin i
N^i_{pred}: predicted nb of events in bin i

o: parameter(s) of interest
f: nuisance parameters

Analysis method
Likelihood evaluation

- To compute L(o,f), essentially needs prediction for (o,f) in each bin $N^i_{pred}(o,f)$
- Nominal predictions from MC, with (mostly) multiplicative weights for effect of variations of parameters (systematics and oscillation probabilities)
- So mainly loop over MC to apply effect of variation of the parameters

Event by event

Loop over individual MC events to apply effect of parameter variations

- ✔ can track events migrating from one analysis bin to another as a result of systematic effects
- ✔ can have precise effect of parameter variation for each event
- ✘ Generally expensive in terms of computing resources (=slower)

Bin by bin

- Group MC events in bins
- Different from analysis bins of previous slide, have more dimensions (neutrino flavor, true neutrino energy, interaction mode, reconstructed variable bin,...)

- ✔ Generally quite faster (less objects to loop over)
- ✘ all MC events in a bin get same effect from parameter variation
- ✘ harder to implement systematics that can results in “migrations”

Analysis method
Reducing likelihood dimensionality

To be able to do statistical analysis in practice, need likelihood as a function of only the parameter(s) of interest

$$\mathcal{L}(o, f) \rightarrow \mathcal{L}(o)$$

Marginalization

Integrate the likelihood over the nuisance parameters

$$\mathcal{L}_{marg}(o) = \int \mathcal{L}(o, f) df$$

- MCMC or numerical integration
- Used in:
  - T2K far detector analyses
  - T2K near detector analysis 1

Profiling

Maximize likelihood with respect to the nuisance parameters

$$\mathcal{L}_{prof}(o) = \max_{f} \mathcal{L}(o, f)$$

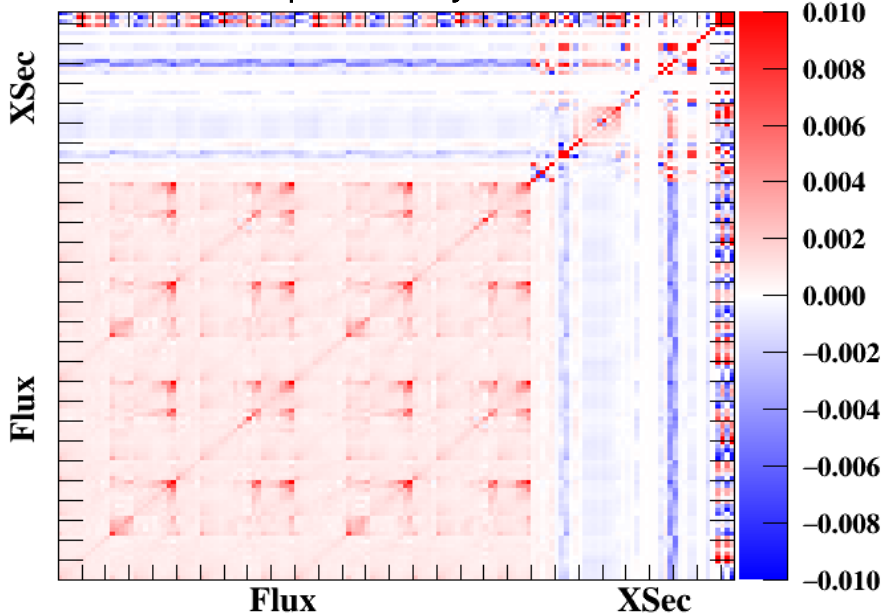
- Generally with MINUIT, but sometimes custom minimization methods
- Used in:
  - NOvA
  - Super-Kamiokande
  - T2K near detector analysis 2

Systematic implementation
Penalty term / prior

- Generally assume (multivariate) gaussian form for the penalty term in the likelihood.
- Fully defined by a set of nominal values for all parameters and covariance matrix
- In T2K, a few parameters don't have this gaussian constraint, and can vary freely on a certain range ("flat prior")

$$-2\ln\mathcal{L}_{penalty}(f) = (f - f_0)^T M^{-1}(f - f_0)$$

T2K run 1-10 preliminary



Constraint for flux and interaction systematics in T2K far detector analysis, from near detectors measurements

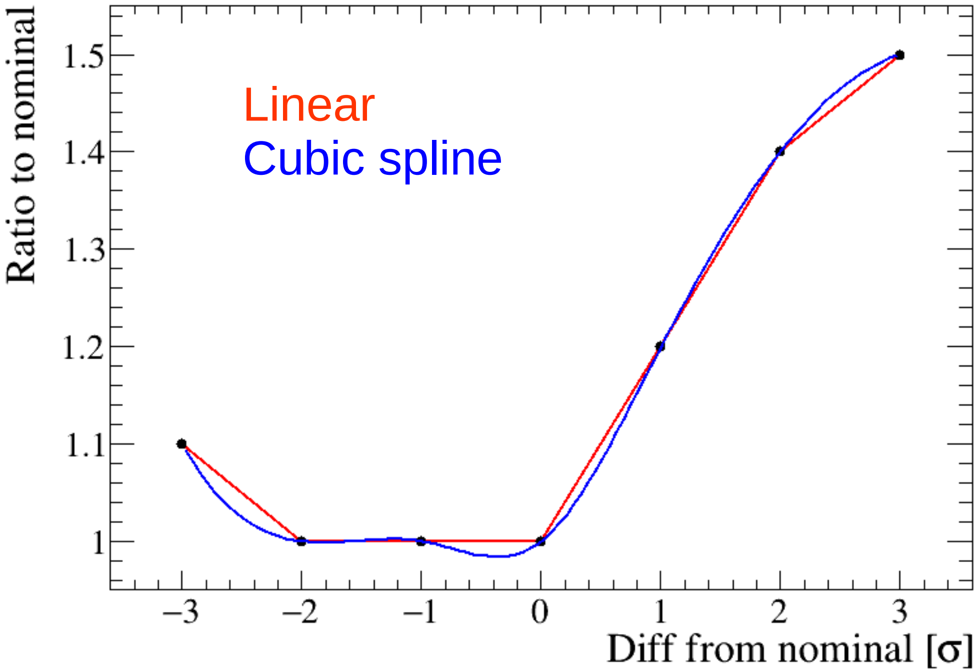
Nice that this allows for correlations between systematic parameters (but that cannot depend on values of those parameters)

Systematic implementation "Splines"/"Response functions"

- Systematic parameters being gaussian does not mean resulting uncertainty on the number of events is.
- Only some systematic parameters ("normalization" systematics) have a simple linear effect on number of events.
- "Spline"/"response function" parameters can have different effects on different MC events, and possibly non linear effect as a function of parameter value

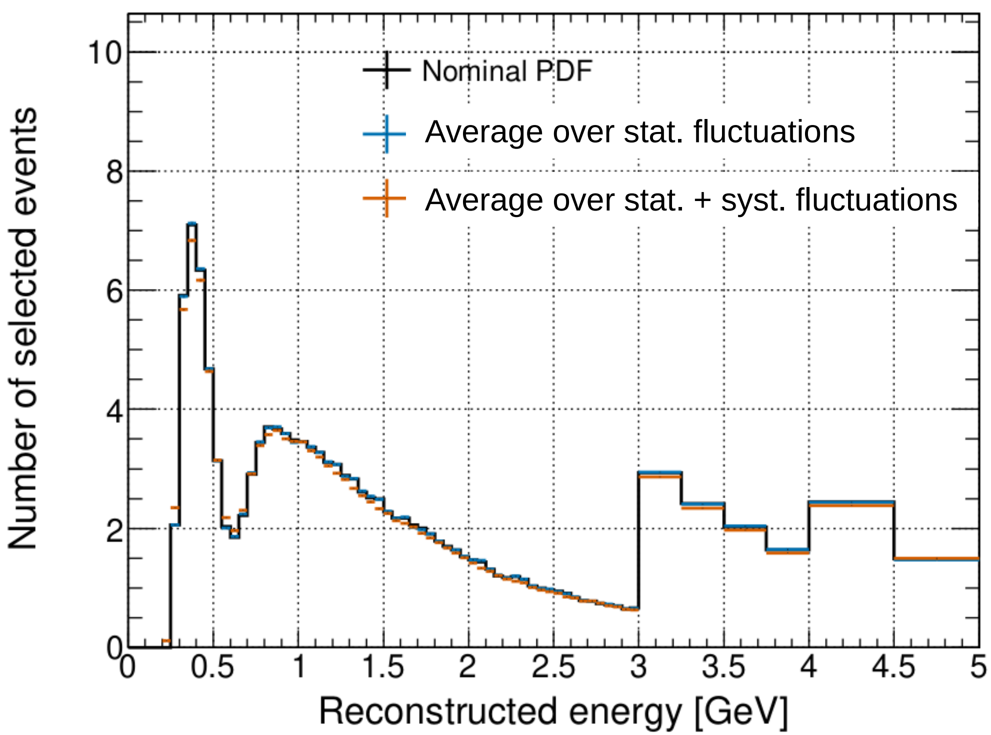
Implementation

- Pre-compute effect for a nb of values of the parameter ("knots")
- Done by looping over MC events to evaluate effect, either on bins or individual MC event
- Interpolate between knots to get effect between knots



"Splines"/"Response functions" Average vs nominal

- With those more complicated parameters, no guarantee that the average over many toy datasets will correspond to the MC dataset for nominal values of systematic parameters
- Have been wondering about the consequences for the "Asimov dataset" method:
  - nominal dataset is probably not an Asimov dataset contrary to gaussian uncertainties case
  - how to define the Asimov dataset for marginal (and not profile) likelihood?



We define the Asimov data set such that when one uses it to evaluate the estimators for all parameters, one obtains the true parameter values. Consider the likelihood function for the generic analysis given by (6). To simplify the notation in this section we define

$$\nu_i = \mu' s_i + b_i \,. \tag{22}$$

Further let $\theta_0 = \mu$ represent the strength parameter, so that here θ_i can stand for any of the parameters. The ML estimators for the parameters can be found by setting the derivatives of $\ln L$ with respect to all of the parameters equal to zero:

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{N} \left(\frac{n_i}{\nu_i} - 1 \right) \frac{\partial \nu_i}{\partial \theta_j} + \sum_{i=1}^{M} \left(\frac{m_i}{u_i} - 1 \right) \frac{\partial u_i}{\partial \theta_j} = 0. \tag{23}$$

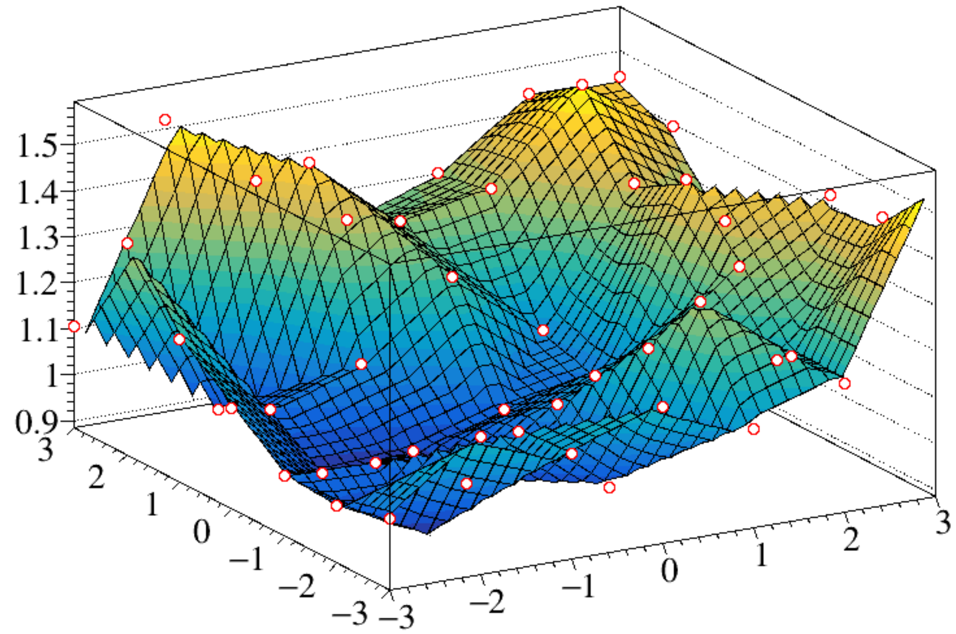
This condition holds if the Asimov data, $n_{i,\mathrm{A}}$ and $m_{i,\mathrm{A}}$, are equal to their expectation values:

G. Cowan et al.Eur. Phys. J. C (2011) 71: 1554

"Splines"/"Response functions"
Multi-parameters

- Main limitation of the method is that we pre-compute variations assuming all other parameters stay at their nominal values
- In a number of cases, effect of a parameter could depend on the values of other ones
- How to extend this method to those cases is something we're actively thinking about in T2K, and interested in the experience of other experiments/people

- Re-evaluating systematics at run time for each MC event too long
- Sometimes can be parameterized in an easy to compute at run time way (simple function of MC event variables and systematic parameters)
- Considering extending spline method to multi-dimensions for other cases
- Also has a cost in terms of computing:
  - nb of knots increase exponentially
  - interpolation more complex



A viable approach?

Systematic implementation
Linearisation vs factorisation

- Generally factorize the effect of systematics: $N^i_{pred}(f)=N^i_0 * (1+\Delta_{flux}) * (1+\Delta_{xsec}) * (1+\Delta_{det})$
- Super-K atmospheric uses a different approach, summing all the effects

$$N^i_{pred}(\vec{\epsilon})=N^i_0\left(1+\sum_j^{Nsyst} F_{ij}\,\epsilon_j\right)$$

F_{ij}: effect of systematic i in nb of events in bin j ("response functions" with 2 knots (nominal, 1σ) using linear interpolation)
ε_j: variation from nominal of systematic parameter j, in units of standard deviations

Makes it possible to write minimization of the likelihood as a linear matrix equation in ε, that can be solved by (fast) iterative inversion method

$$\boldsymbol{M}\,\vec{\epsilon}=\vec{V}(\vec{\epsilon})$$

- ✔ Significantly faster than MINUIT minimization for large number of parameters
- ✘ Approximation of linear effect of parameters and summing the different effects can become problematic in systematics dominated regime
- ✘ Does not give a post-fit error for nuisance parameters, just best fit values

Migration systematics

Systematics creating “migrations” of MC events harder to handle with multiplicative weights:

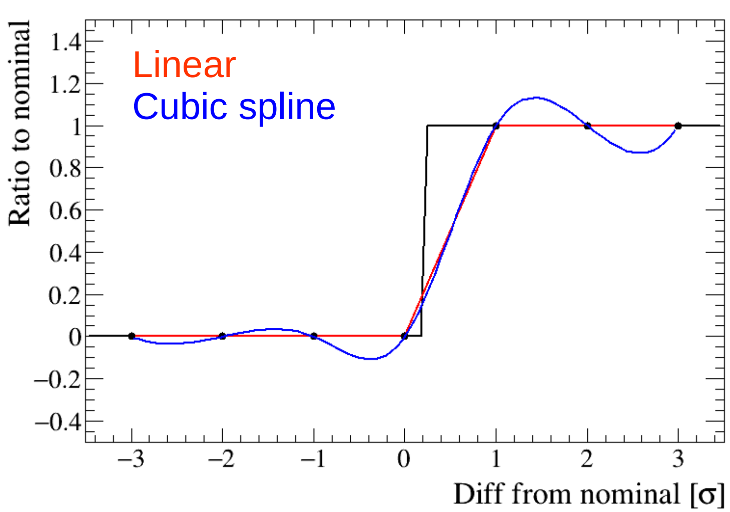
- change of analysis bin as a result of a change of the value of the reconstructed variable used to bin events
- change whether event passes event selection cuts or not

Standard spline/response function method could a priori handle this, but challenges in practice

Event by event

✔ Change of analysis bin easy to handle when looping over MC events if effect can be computed at run time

✘ Interpolation might be problematic for effect on event selections: binary effect



Bin by bin

✔ Can pre-compute bin splines for those systematics (with enough MC events, interpolation between knots can be OK)

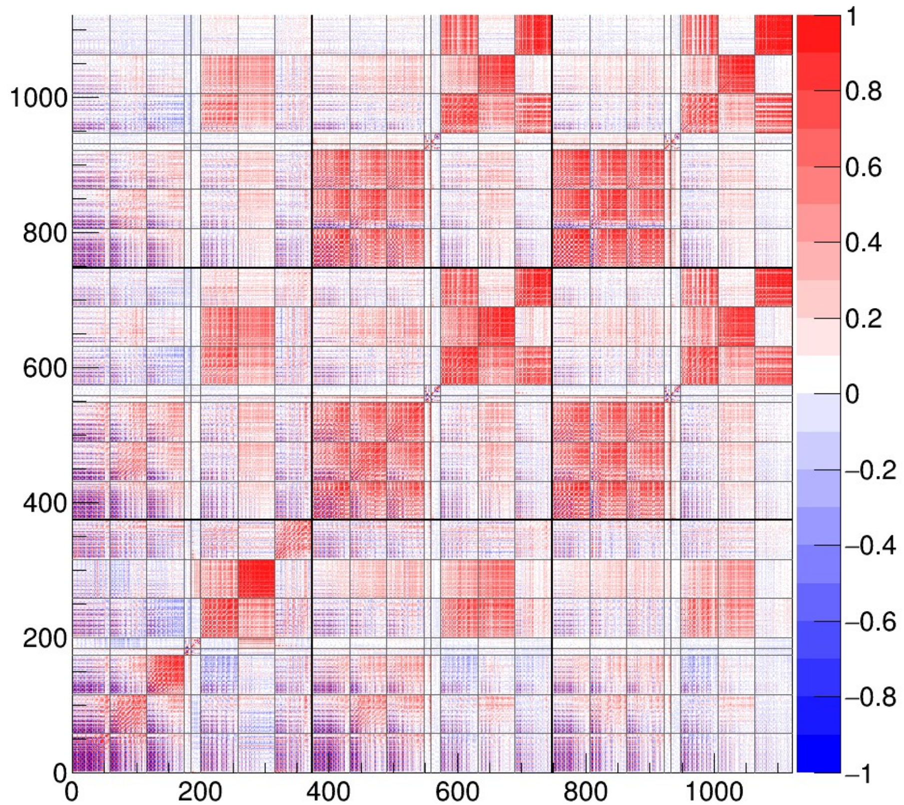
✘ If arrival bin is empty in nominal prediction, cannot reweight

✘ Creates problems when >1 such systematics create significant migrations into a given bin

✘ Does not keep track of which events migrate (for effect of other systematics)

Systematic implementation
Covariance matrix method

- Systematics affecting whether events pass event selections or not likely to be difficult to compute at run time, and have dependence on values of other parameters
- Often end up using the “covariance matrix” method:
  - generate many sets of variations of the systematic parameters
  - compute predictions for each analysis bin for each set ($N^i_{pred}(f)$)
  - build covariance matrix of those variations
  - Later analysis uses this to incorporate systematics: 1 parameter per analysis bin, covariance matrix as constraint



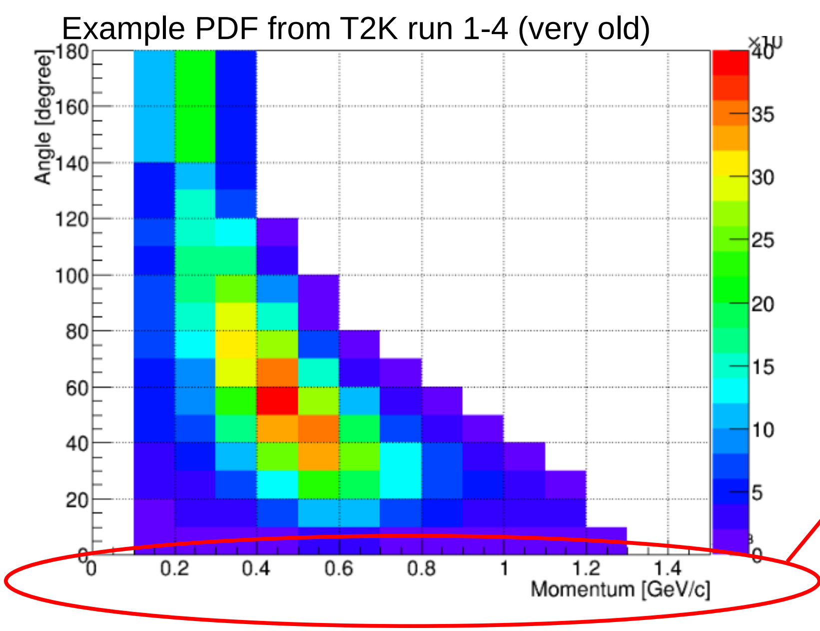
- ✔ Allows us to do the analysis
- ✔ Keeps correlations between analysis bins
- ✘ Large number of parameters: makes fits harder
- ✘ Obtained uncertainty on N^i_{pred} can only be gaussian
- ✘ Loose link to original systematics, and cannot see their individual effect on analysis

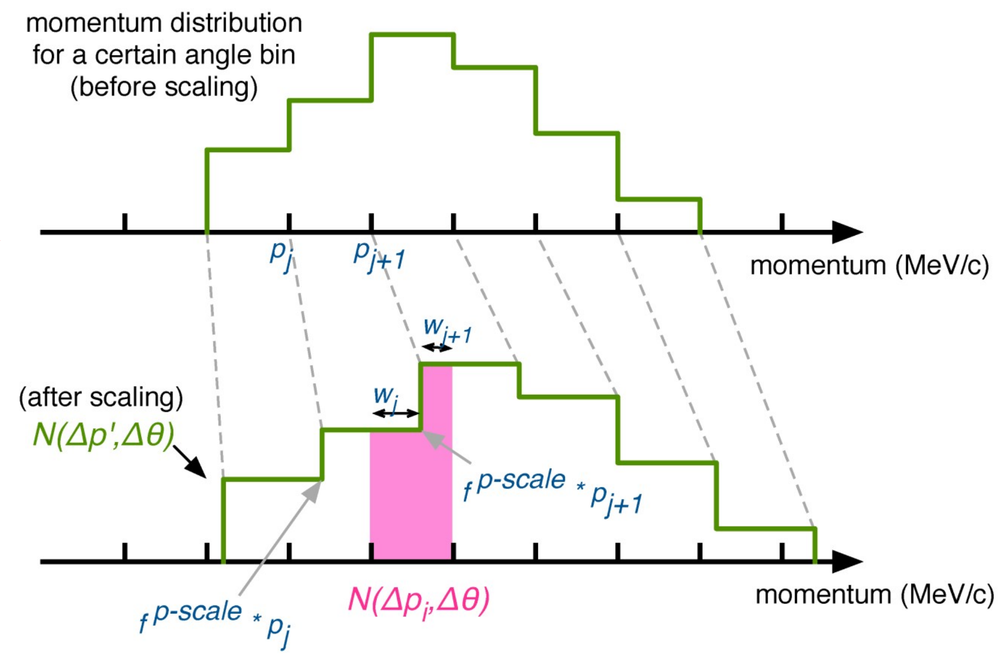
Can try to reduce nb of bins (PCA, merging bins with similar relative error size), but would generally like to move away from this method

Energy scale

- Energy scale: absolute level of the energies measured in detector (eg momenta measured are all wrong by 3%)
- In practice, migration systematics that can change analysis bin an MC event is in
- Easy to handle for event by event analyses

For bin by bin analyses, previously used a method where edges of bins in the relevant variable were scaled to determine migrations





Works, but:

- assumes uniform distribution of events in a bin
- binning variable not necessarily the one to which scale should apply

Decided to move to “additive splines” method for the future

Migration systematics
Binned splines: additive method

- For bin by bin analysis, try to avoid problem of empty target bin with “additive splines”: add to predictions in bin i instead of applying weight to nominal predictions.
- Not so easy in practice, 3 constraints:
  - conserve total number of events
  - commutative (order in which systematics are applied should not matter)
  - computationally feasible

Lukas Berns (Tokyo Institute of Technology) came up with a method to implement this:

$$\tilde{N}_a := \sum_z \tilde{N}_{az} := \sum_z \left[\left(\prod_s \frac{\sum_{a'} N^s_{a'z}}{N^0_z} \right) \frac{\overline{W}_{az}}{\sum_a \overline{W}_{az}} \right] N^0_z$$

with

$$\overline{W}_{az} := \left[\delta_{az} + (1 - \delta_{az}) \sum_s \frac{N^s_{az}}{\overline{N}^s_{zz}} \right]$$

N_a: #evts in bin a after migration
N_{az}: #evts in bin a coming from bin z
N^s_{az}: #evts in bin a coming from bin z as a result of systematics s
N^0_z: nominal #evts in bin z

(Derivation is not trivial, would be a talk in itself)

✔ Found to work well when migrations are all following the same binning variable

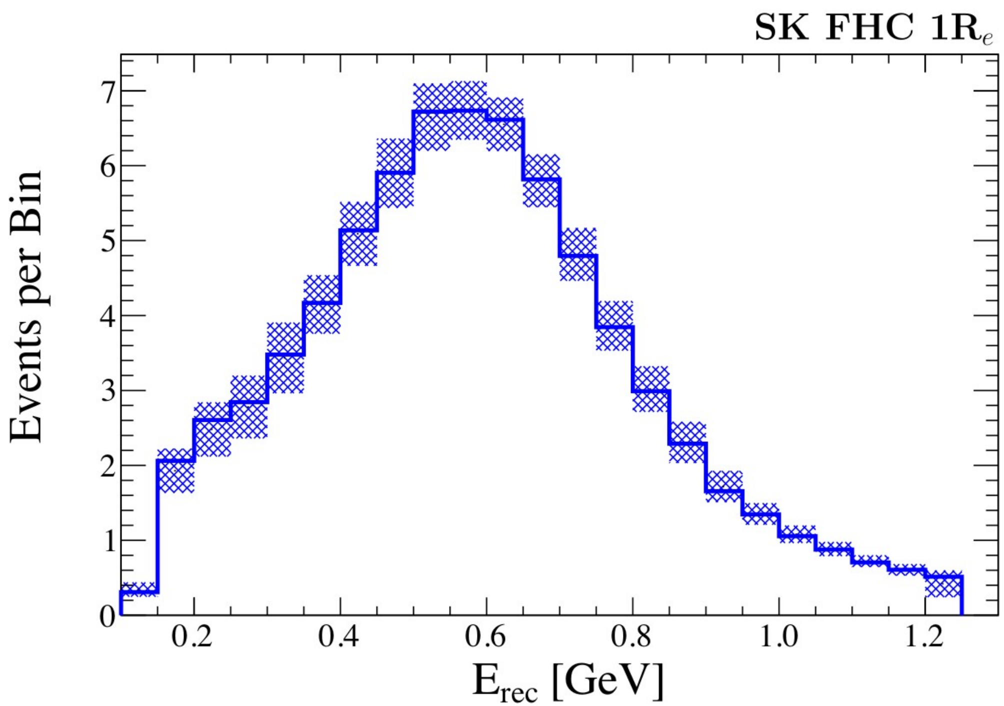
✘ Some doubts whether it would work well if there are significant migrations along different variables in multi-dimensional PDFs

Migration systematics
MC statistics

- Due to finite amount of MC, problems can happen when migrating MC events in event by event approach or additive spline method
- Can create discontinuity of the likelihood as a function of the migration systematic parameter when an event moves to another bin

Toy example:

- Take a predicted PDF for a sample in an analysis
- Generate underlying MC for it, to which a migration systematics can be applied
- Use an energy scale-like parameter, which scales the variable used to bin sample (Erec)
- Applied with an event by event approach
- Look at the likelihood as a function of this parameter for various number of MC events



T2K PDF for FHC 1Re sample from Phys. Rev. D 103, 112008 (2021)

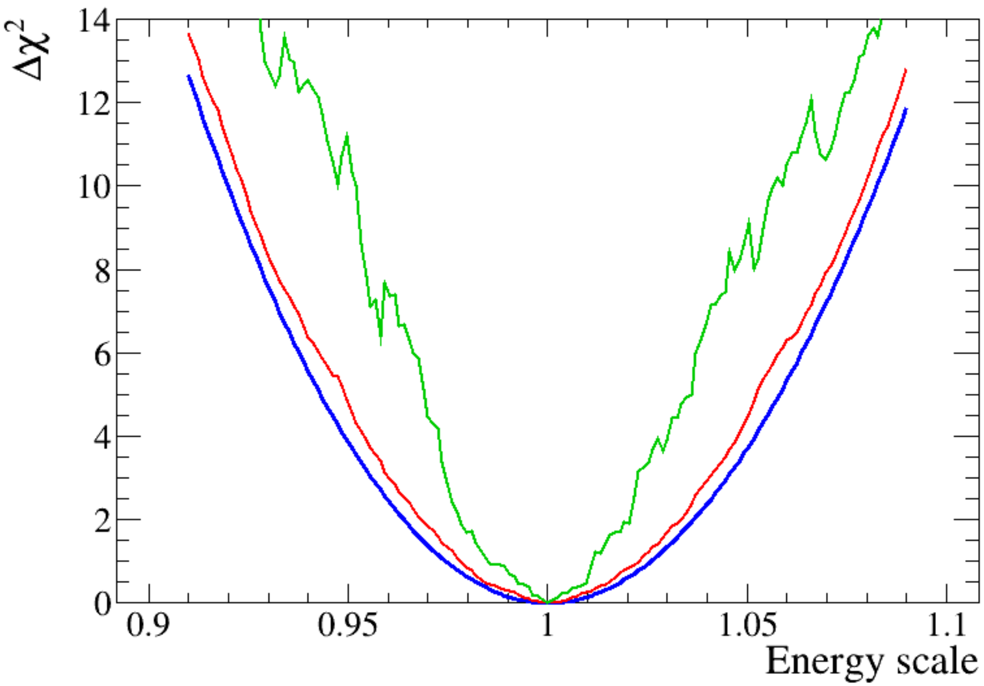
Migration systematics
MC statistics

- Look at the likelihood as a function of this scale parameter when fitting an "Asimov" dataset, and a toy dataset generated from the underlying PDF
- Number of MC events used has a strong impact, even more in the case of a toy dataset
- Usual rule of thumbs "10 times more MC than data" does not work with that kind of systematic parameter

500 MC events
5k MC events
50k MC events

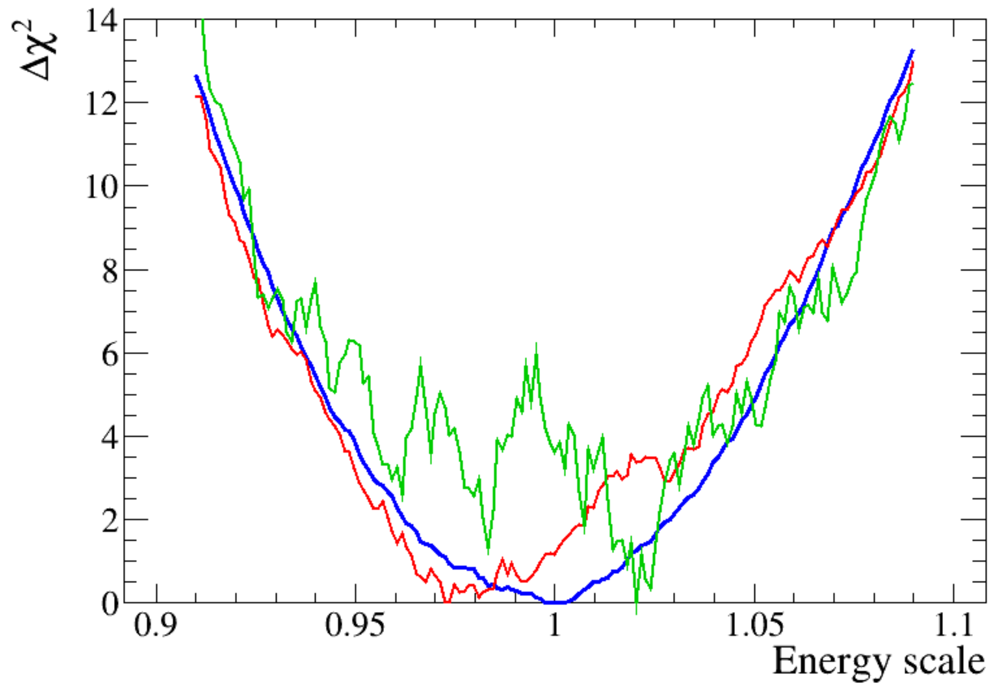
Asimov dataset

N_{obs}=72.9611



Toy dataset

N_{obs}=65



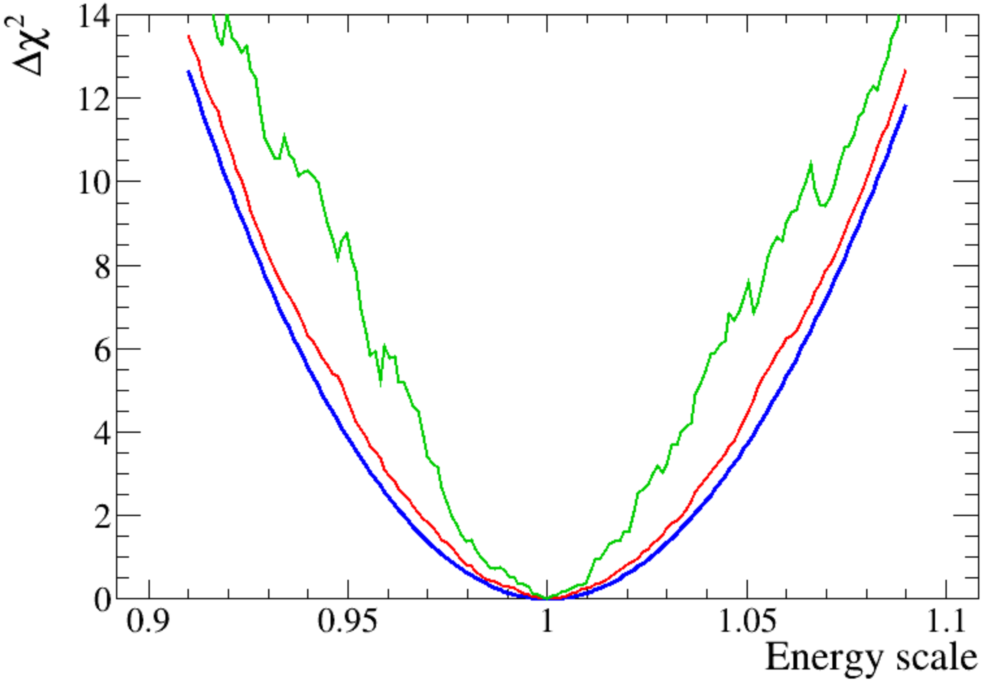
Migration systematics
MC statistics – Barlow-Beeston method

- Received comment that this could be due to MC statistical error on prediction in each bin
- Implemented method by Barlow and Beeston (https://inspirehep.net/record/35053/) that takes this into account
- Improves a bit for lowest number of MC events, but does not solve the problem
- Indicates that fundamental problem is not simple MC statistical error, but discontinuity of predictions as a function of systematic parameter (or that I did not implement method properly)

500 MC events
5k MC events
50k MC events

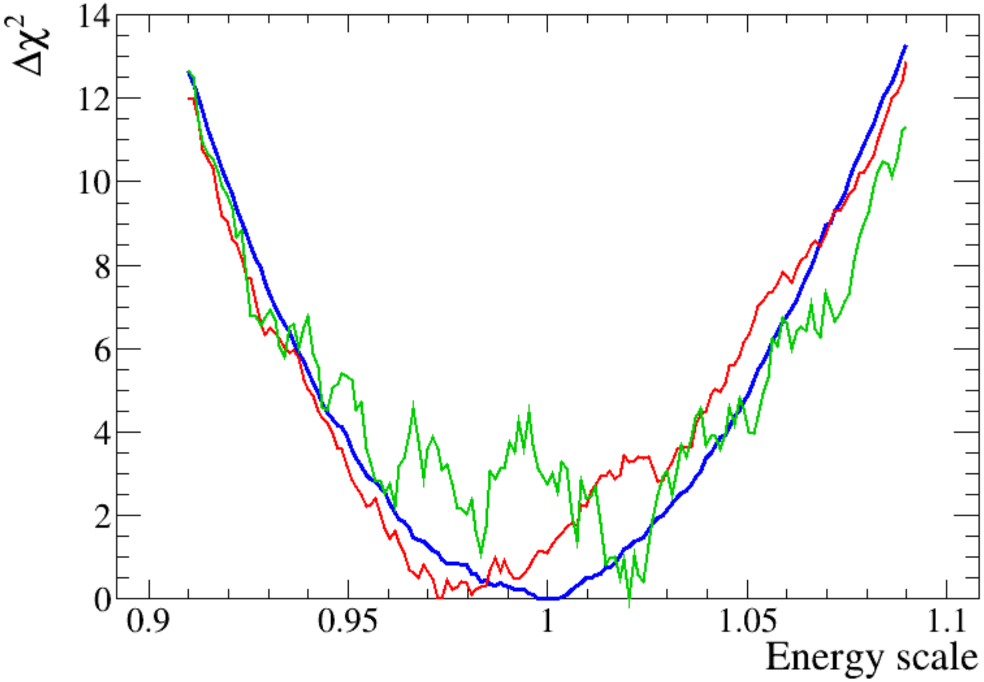
Asimov dataset

N_{obs}=72.9611



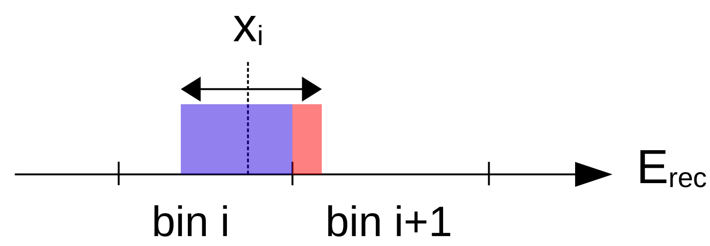
Toy dataset

N_{obs}=65



Migration systematics
Regularization

- Can use regularization methods to avoid abrupt changes of the numbers of events in bins
- One that has been used in T2K is to give a “width” to events, and put fractions of the weight of this event in different bins



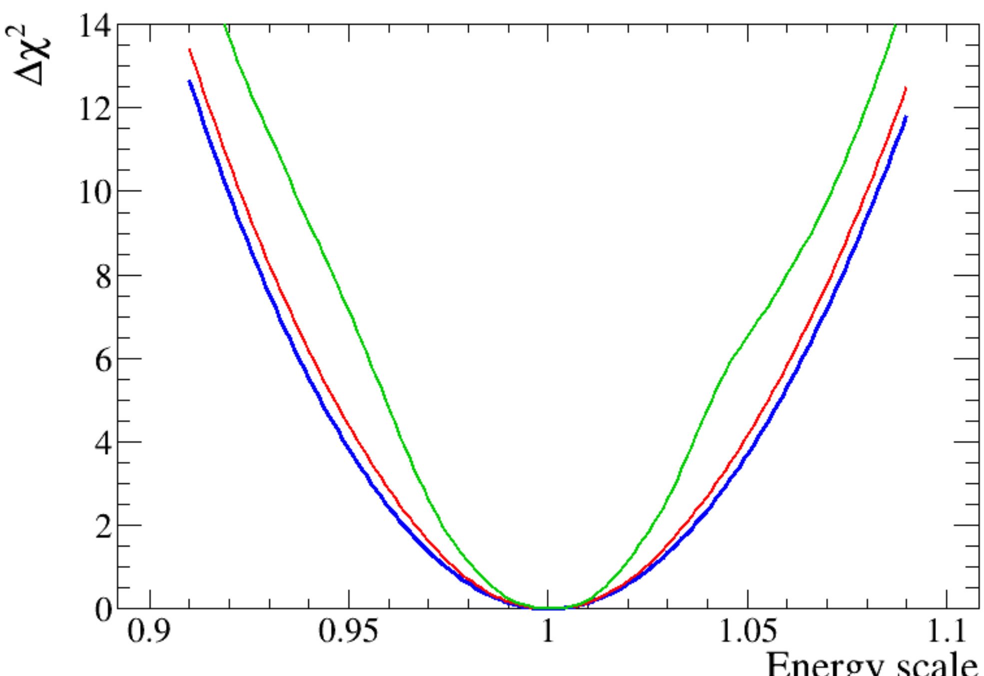
Can also use other shapes (gaussian, triangles)

- Tried it on my toy example (width=bin width), removes discontinuities, but MC stat still matters

500 MC events
5k MC events
50k MC events

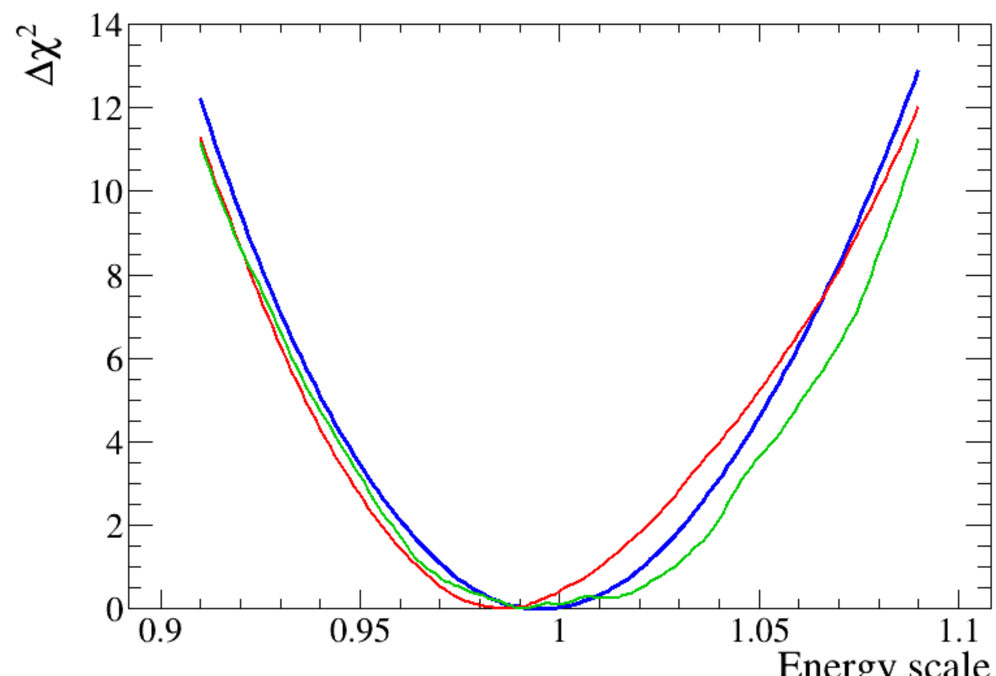
<u>Asimov dataset</u>

N_{obs}=72.9611



<u>Toy dataset</u>

N_{obs}=65



Near/Far extrapolation in LBL experiments Overview

- In long baseline experiments, near detectors sample neutrino beam before oscillation, comparing data to predictions from the experiment model
- Can be used to predict more precisely what should be observed at the far detector for various neutrino oscillation hypotheses
- Different experiments use different methods

Common systematic model

- A number of systematic uncertainty parameters (flux, xsec) are common or directly correlated between near and far detector model
- Fit of near detector data used to tune predictions and constrain uncertainties for the far detector analysis, by changing central values and constraint on those parameters

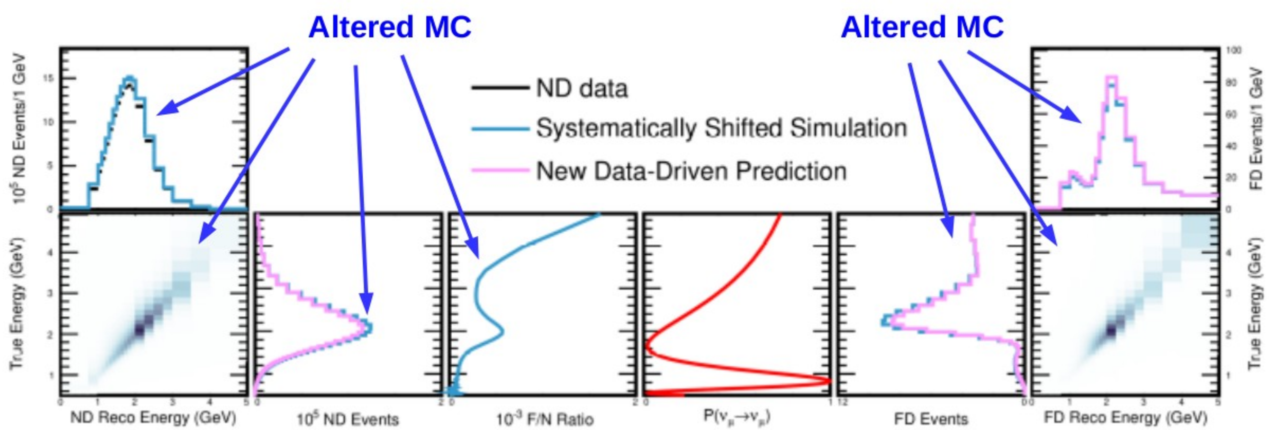
Direct extrapolation

- Turn observation at the near detector into a data driven prediction for the far detector
- Models used to go from reconstructed quantities to true neutrino energy at each detector, and for differences between the 2 detector

Near/Far extrapolation in LBL experiments
Direct extrapolation (NOvA)

- Method affects how systematic uncertainties have to be evaluated for oscillation fits at the far detector
- Can then be incorporated to the analysis as nuisance parameters using methods described previously such as binned splines, so will not go into details

"Extrapolation" and uncertainties



We simulate the effect of our cross section systematics' *residual* effect after extrapolation

by re-doing the entire analysis for each systematic
(each of which can affect multiple both signal & bknd)

and use the difference to extrapolated nominal MC as nuisance parameters in our oscillation fits

J. Wolcott @ NuInt 2018

Near/Far extrapolation in LBL experiments
Model based (T2K)

- T2K uses near detector data to constrain systematic uncertainties on flux and neutrino interaction models
- General problem of propagating an auxiliary measurements of some systematic parameters to an analysis
- 2 different approaches used

Sequential fits

- Fit near detector data only (using profiling)
- Best fit values and covariance matrix from MINUIT for the relevant systematic parameters used as prior constraint in far detector analysis

✔ Can use different fitting methods at near and far detector
✔ (T2K case) faster far detector fits
✘ Can only propagate gaussian constraint

Simultaneous fits

Joint analysis of near and far detector data

✔ Proper constraint (=non-gaussian) on the systematic parameters from ND data
✘ Many parameters in the fit: limits possible fitting methods (MCMC used in T2K)
✘ Can be slower

Near/Far extrapolation in LBL experiments
Model based (T2K) – beyond gaussian

- So far, constraint from near detector fit is close to gaussian for most (but not all) systematic parameters propagated to far detector analysis
- Gaussian assumption nevertheless not very satisfactory, considering other method for sequential fit approach for the future

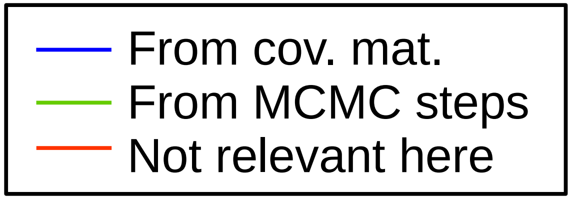
In sequential fit approach, FD fitters compute L_{marg} by numerical integration

$$\mathcal{L}_{marg}(o) = \int \mathcal{L}(o, f)df = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(o, f_i)$$

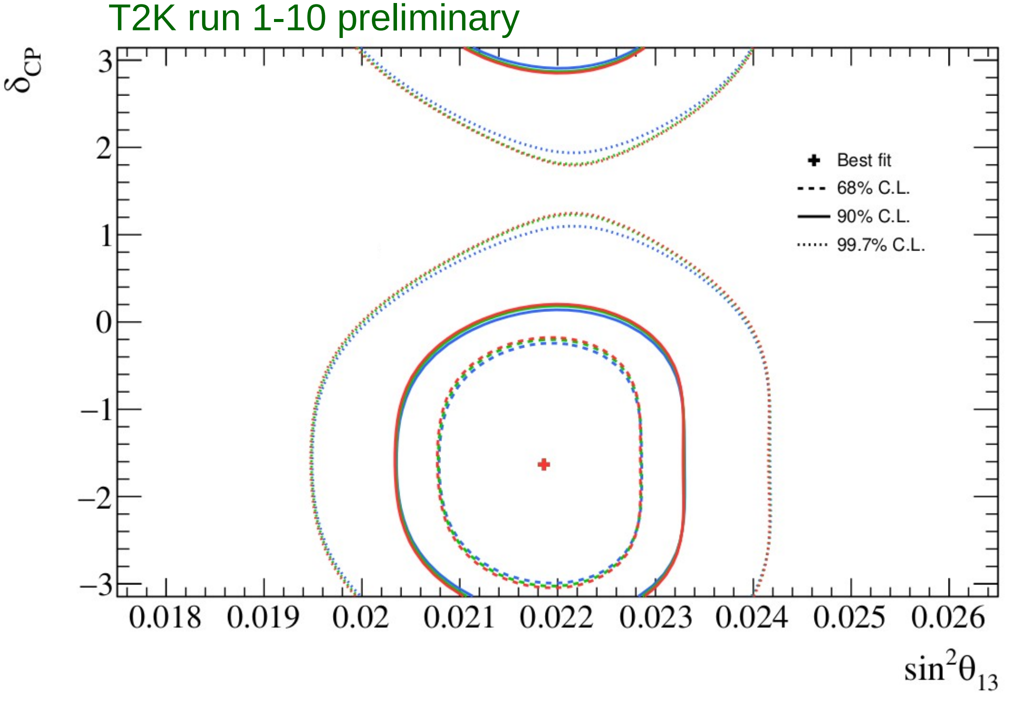
- f_i are sets of values of the systematic parameters
- In T2K, f_i generated following prior distributions of the nuisance parameters (with penalty term removed from likelihood)
- For the parameters constrained by ND fit, uses multivariate gaussian prior based on the covariance matrix from ND fit

Near/Far extrapolation in LBL experiments
Model based (T2K) – beyond gaussian

- Can instead use sets of values of the parameters corresponding to steps of the chain from MCMC ND fit
- Provides sets of values of the parameters that follow the constraint from ND data, without gaussian assumption
- Was tested in T2K to check effect of gaussian assumption in sequential fit



Example: T2K sensitivity to (δ, $\sin^2\theta_{13}$), using reactor constraint on $\sin^2\theta_{13}$



- Method works, but MCMC produces correlated sets of values: less statistical powers
- Interested in testing other methods as well
- Current idea is generate sets of values following covariance matrix, weight them by likelihood in ND analysis (importance sampling)

Summary

- Analyses considered in this presentation use a maximum likelihood approach, comparing observations to predictions for different values of the parameters
- Systematic uncertainties incorporated as parameters modifying the predictions compared to the nominal prediction obtained by MC simulation
- Relies mostly on multiplicative weights assigned to MC events or group of events
- Systematic parameters can have either linear effect on number of events ("normalization") or more complicated effect varying from an MC event to another ("spline"/"response functions")
- Latter ones are implemented as pre-computed functions of the parameter value, giving weights for MC events or group of events.
  Allows us to implement more complex effects, but difficulties arise if those effects depend on the value of more than one parameter, or the parameter can make events migrate to another analysis bin or in and out of the selected samples
- In long baseline oscillation experiments, near detectors allow to constrain systematic uncertainties, but there can be challenges to propagate this constraint to the final analysis

BACKUP

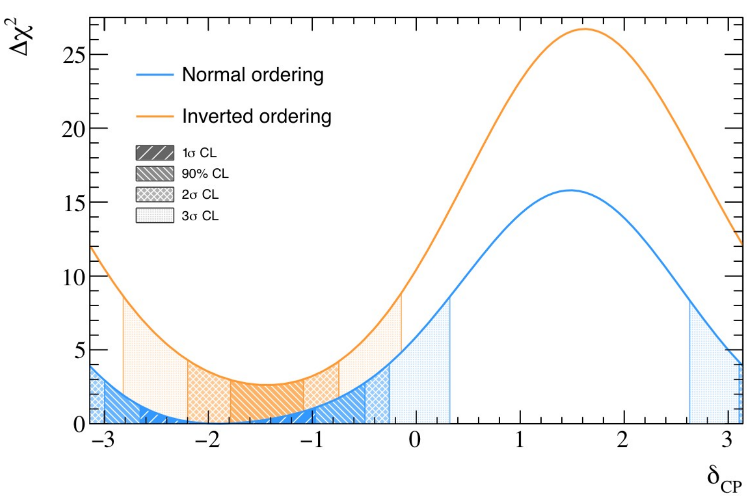
Standard neutrino oscillations
What do we measure?

- Phenomenon now well established, parameters measured to a certain precision
- 3 (degenerate) open questions in standard oscillation framework
- Precise measurements allow to test this standard oscillation framework, in search of physics beyond the standard model

CP symmetry

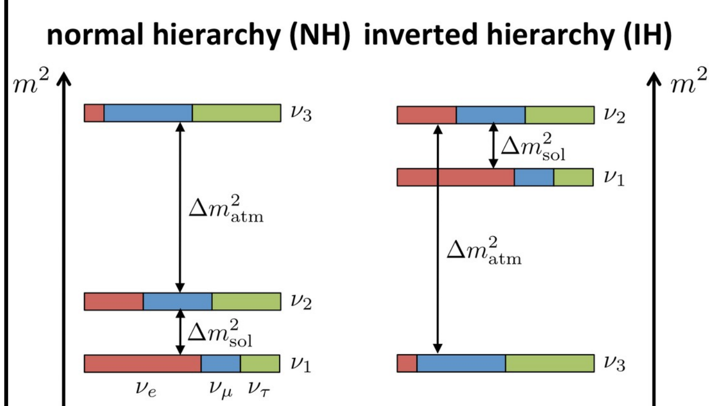
- difference matter/anti-matter

~~CP~~ ⇔ sin(δ) ≠ 0



Mass hierarchy

- Neutrino mass models
- input for other experiments (0νββ, supernova)

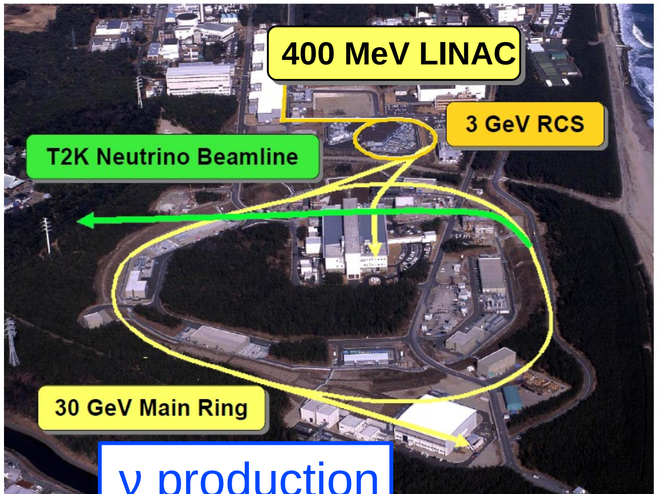


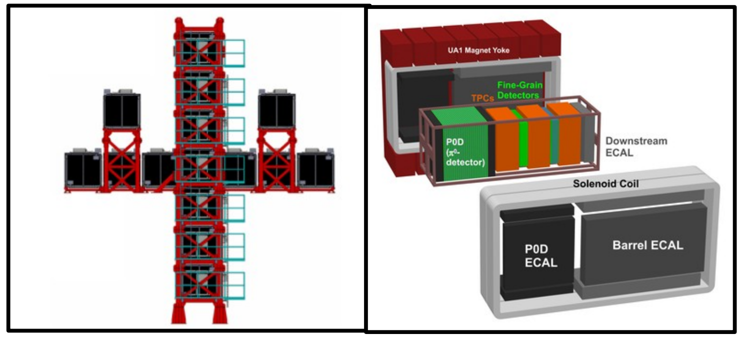
Octant of θ_{23}

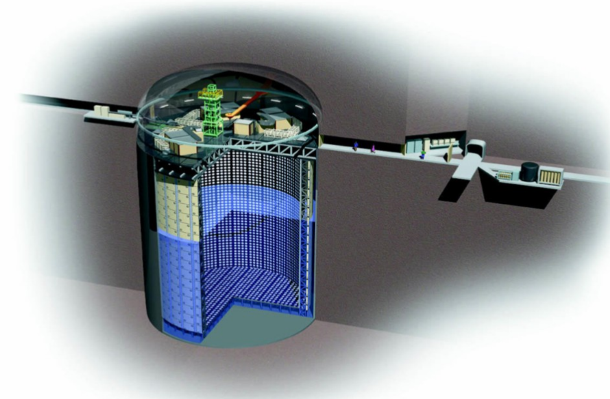
- symmetries in lepton sector

$\theta_{23} > \pi/4$?
$\theta_{23} = \pi/4$?
$\theta_{23} < \pi/4$?

Tokai to Kamioka (T2K)







ν production

J-PARC accelerator complex and neutrino beamline

0

ν_μ

2.5°

Near detectors

On-axis: INGRID

Off-axis: ND280

280m

ν_μ

Far detector

Super-Kamiokande

295 km

Study neutrinos before oscillations

$\nu_\mu \rightarrow \nu_e$ appearance
$\nu_\mu \rightarrow \nu_X$ disappearance

- Baseline: 295 km
- Off-axis beam

Oscillation analysis
General method

Likelihood analysis:
- Compare predictions to observations
- Look for model or values of the parameters which give best agreement
- And range of values which give an agreement within a certain distance from this best fit

$$\mathcal{L}(data, osc, syst) \approx \prod_{bins} (Poisson(N_{obs}, \underline{N_{exp}}(osc, \underline{syst}))) \times \underline{\mathcal{L}_{penalty}(syst)}$$

Build a model of the experiment to evaluate N_{exp}

Properly model and try to constrain systematic uncertainties

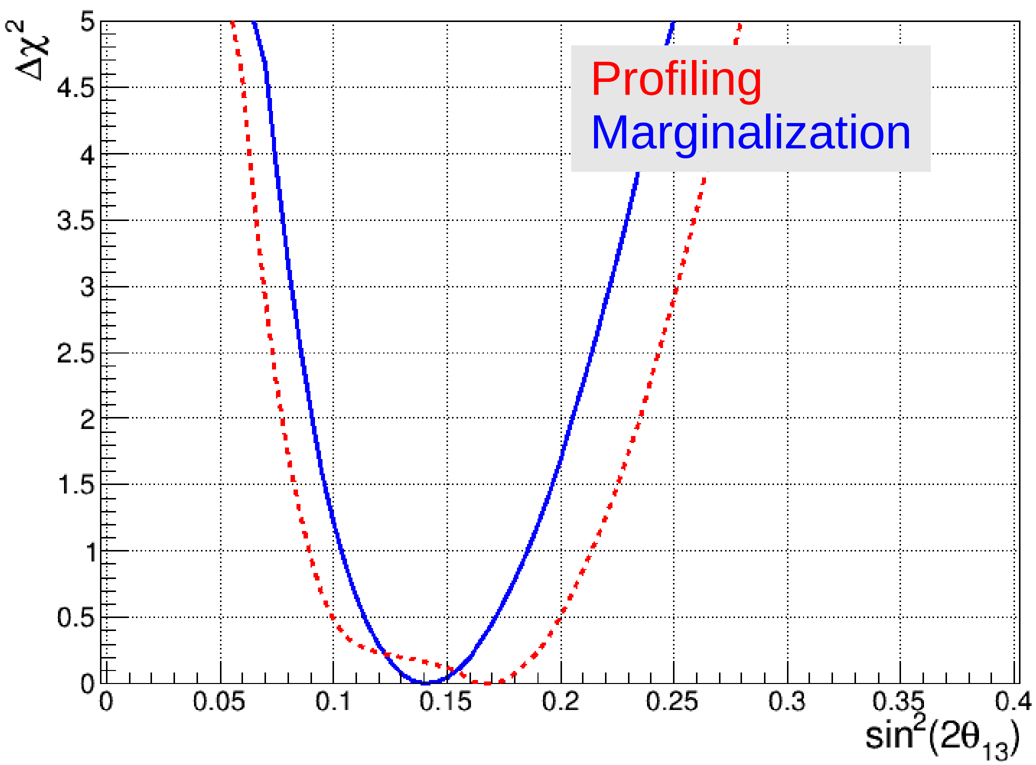
Keep analysis manageable in terms of computing

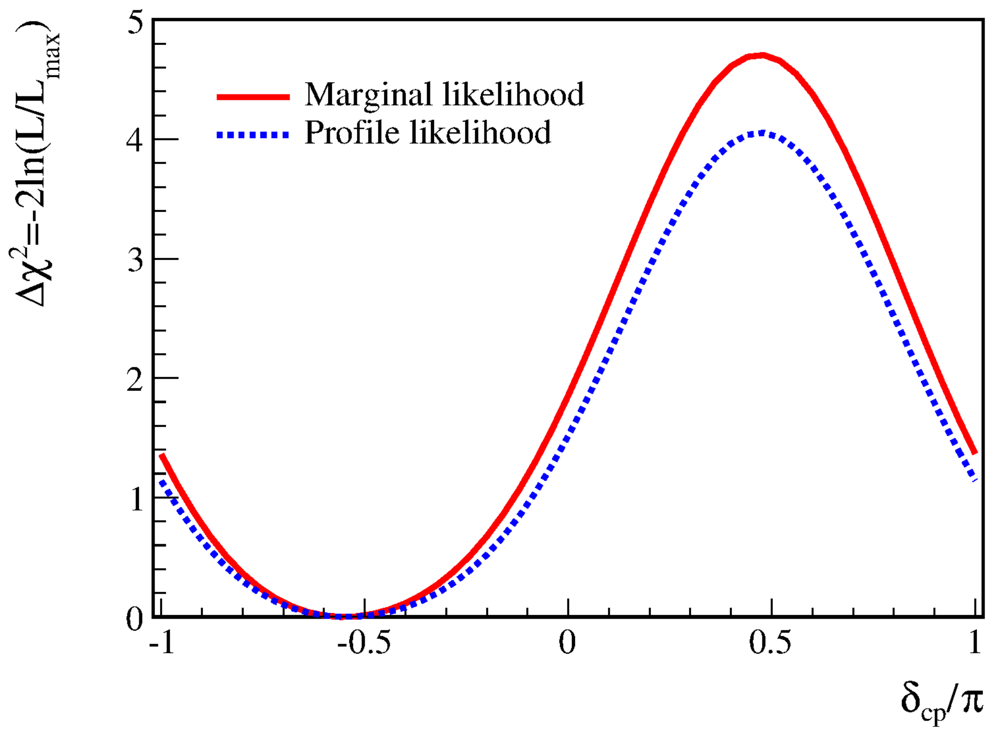
(in practice use likelihood ratio instead of Poisson in each bin)

Common points
"Reducing dimensionality of likelihood"

"In the case of linear parameter dependence and where the nuisance parameters appear in a Gaussian form, the profile and marginal likelihood functions will be identical and can be used to produce intervals with correct frequentist coverage. For the neutrino oscillation analysis, the parameter dependence is nonlinear, and as a result the profile and marginal likelihoods differ and frequentist coverage is not guaranteed"

PRD 91, 072010 (2015)

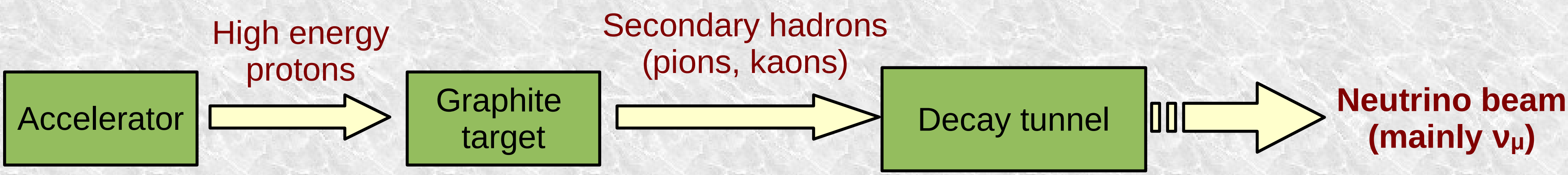


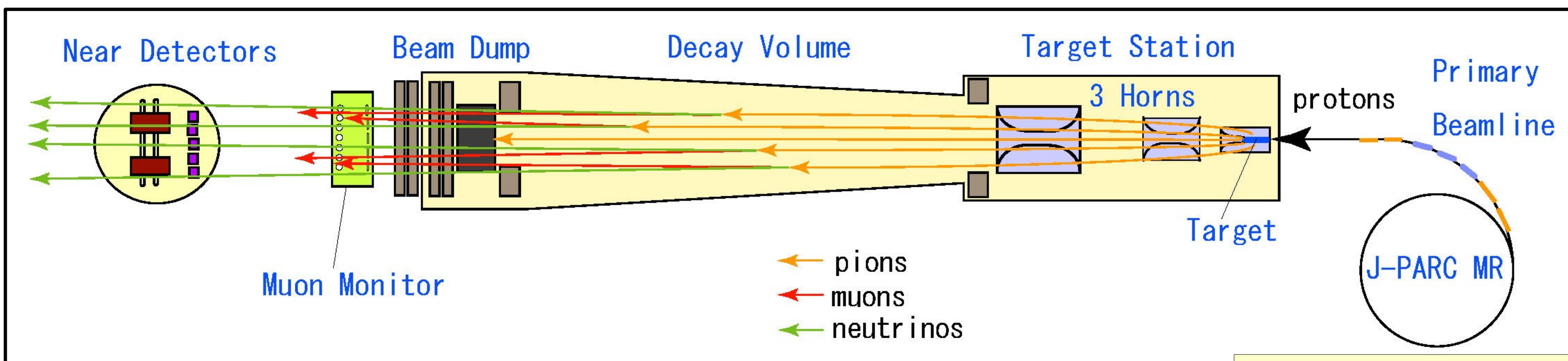


T2K Run1-4 (old) data fit assuming NH

Long baseline experiments
Beam production

Neutrino beam produced by an accelerator (basic concept)





Example: T2K beam

Obtain an **almost pure ν_μ beam**, with an intrinsic ν_e component of a few %