Common tools:
statistical methods for combination of experimental results

JENNIFER2 meeting
(October 2018, Paris)

S.Bolognesi (CEA Saclay) and D.Tonelli (INFN Trieste)

Introduction

- The main deliverable will be **a document** detailing recipes on how to properly combine results from different experiments, in presence of multi-parameter analysis:

  emphasis on **combination of likelihoods** as a function of the parameter of interests (~5) and the nuisance parameters (~hundreds)

  (complete likelihood at their highest possible level of dimensionality to preserve coherence of information for further manipulation: profiling/marginalization...)

- Second optional deliverable (if personpower): **software tool** for storing and combination of user-provided likelihoods

A conceptual, technical (and sociological) challenge

Outline:

- Some previous examples
- Physics cases in T2K and Belle2
- A practical example
- Correlations, correlations, correlations!

How was solved in the past?

More frequent approaches:

- **condividing the full data** (in proper format) between different experiments

(e.g. Higgs search at LEP and at LHC, CMS+LHCb $B_s \rightarrow \mu\mu$)

- **just combining x±δx measurements** with proper correlation between the systematics (i.e. 1D likelihood with Gaussian behaviour)

(e.g. LEPEWWG)

- **combining multivariate likelihood** is more rare, is equivalent to sharing data in terms of correctness but more feasible

Most used in combination of different analysis or different datasets of a given experiment

A Combination of CDF and DØ Limits on the Branching Ratio of $B0_{s(d)} \rightarrow \mu^+\mu^-$ Decays
(http://arxiv.org/abs/hep-ex/0508058v3)

Using Likelihood for Combined Data Set Analysis
https://arxiv.org/pdf/1502.03081.pdf

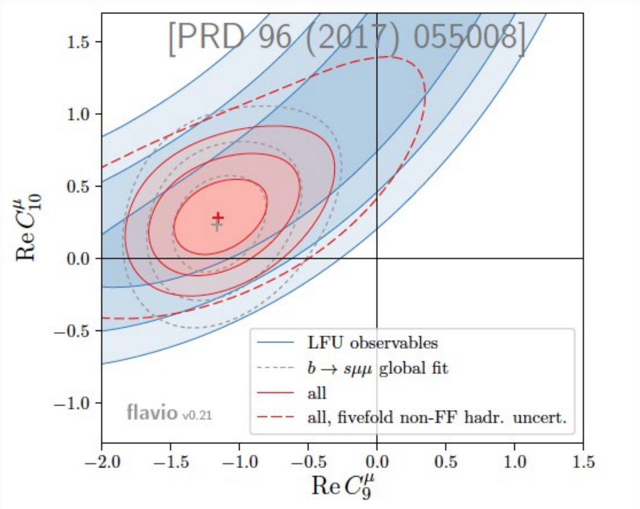
Physics cases (Belle II)

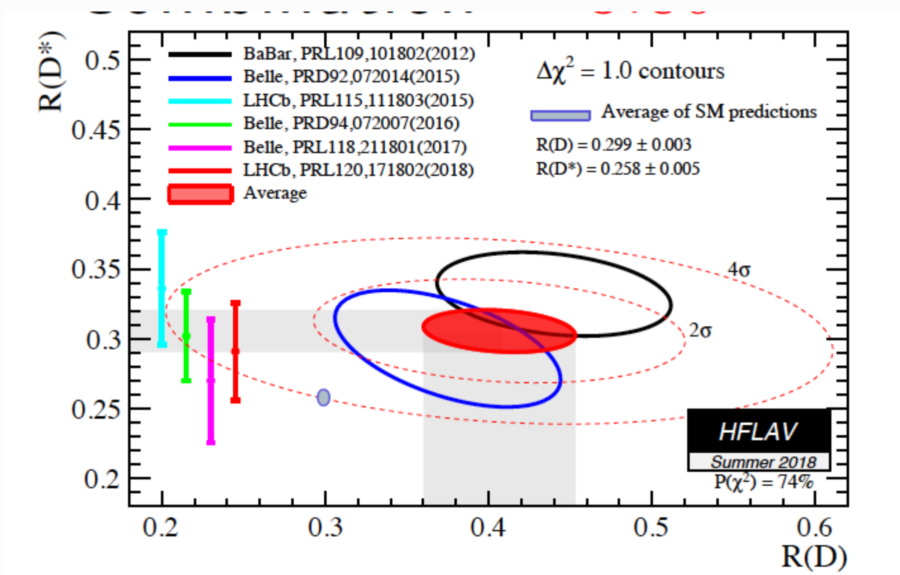
- **Combination with LHCb** for measurements limited by statistics. Same examples:

Lepton flavour universality

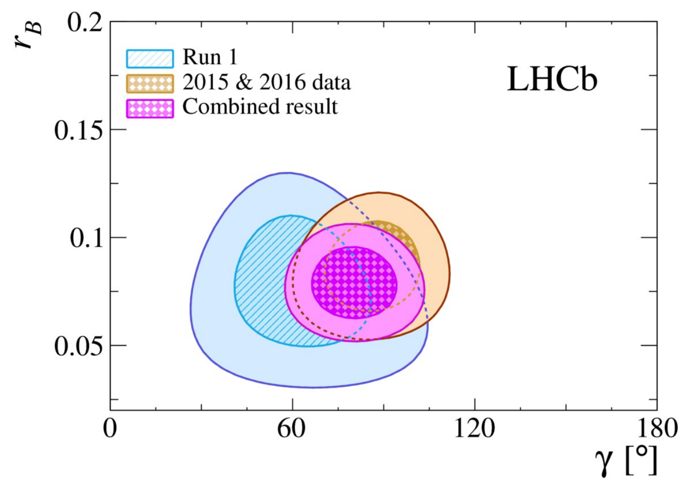
$$\mathcal{B}(\bar{B} \to D^{(*)}\tau^- \bar{\nu}_\tau)/\mathcal{B}(\bar{B} \to D^{(*)}\ell_\ell^- \bar{\nu}_\ell)$$

- Not significative (yet?) hints for SM departures with uncertainty largely dominated by statistics





- The CKM phase of the b → u transition
Uncertainty is purely experimental ($\sigma_{stat} \sim$ 3-4 x σ_{syst}) and it is the largest among CKM angles



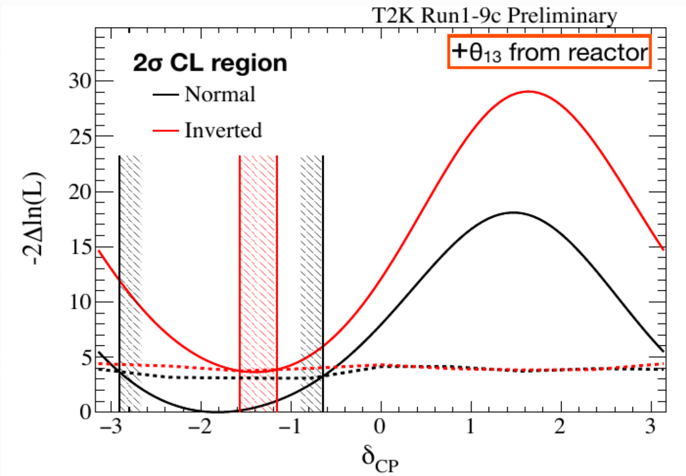
- All rare decays

Physics cases (T2K)

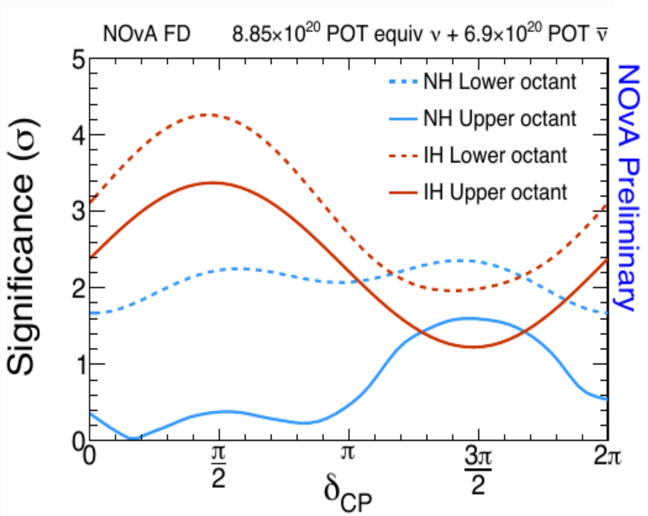
- **Neutrino oscillation results from T2K and NOVA** limited by statistics:

TODAY (ICHEP2018)

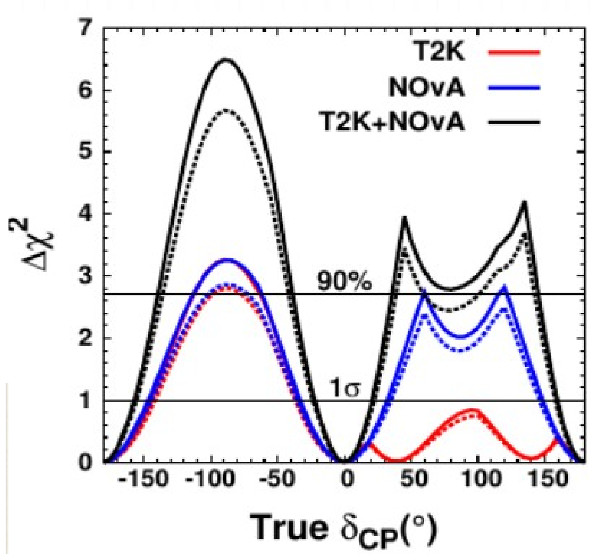
T2K



NOVA

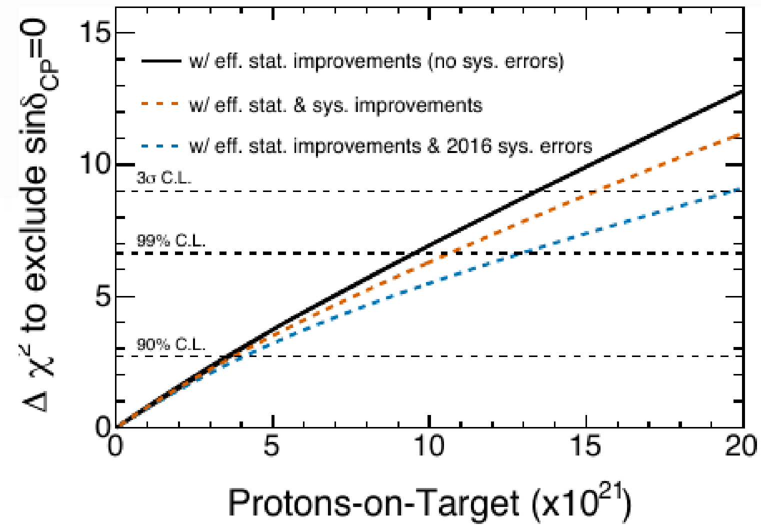


~2021

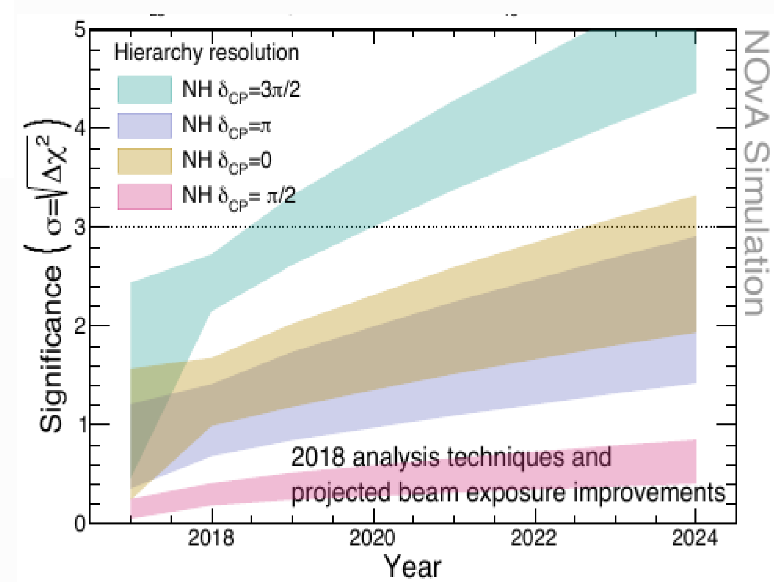


~2026

T2K



NOVA



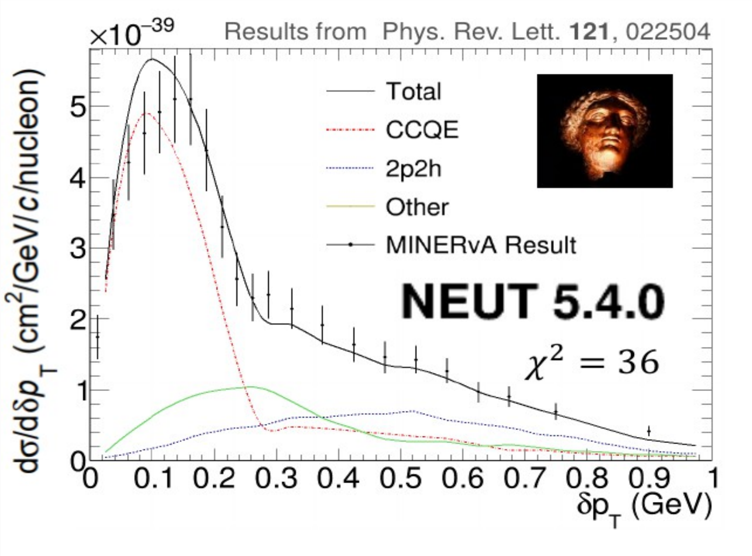
Physics cases (T2K)

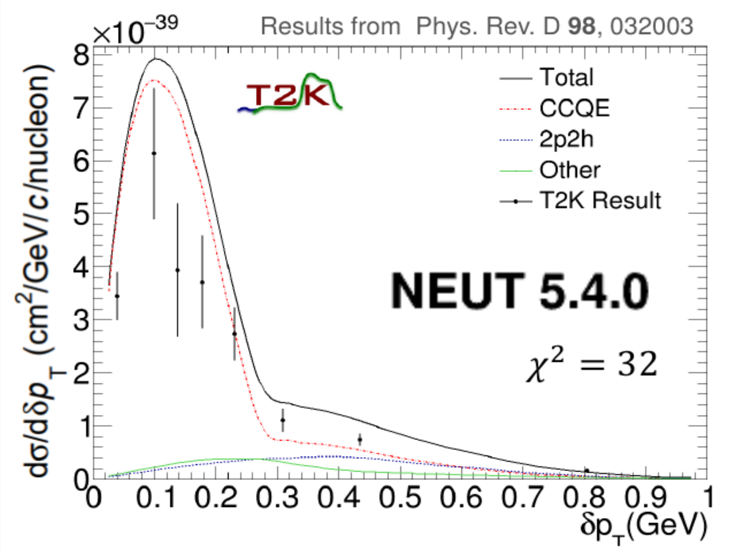
- **Neutrino oscillation results from T2K and NOVA** limited by statistics

- **Combination of T2K and T2K-2 datasets ?**

Not just more stat: new near detector and beam upgrade → not completely correlated systematics (availability of tools for re-analizing old data?)

- **Neutrino cross-section measurements** from ND280 and other experiments (e.g. Minerva)





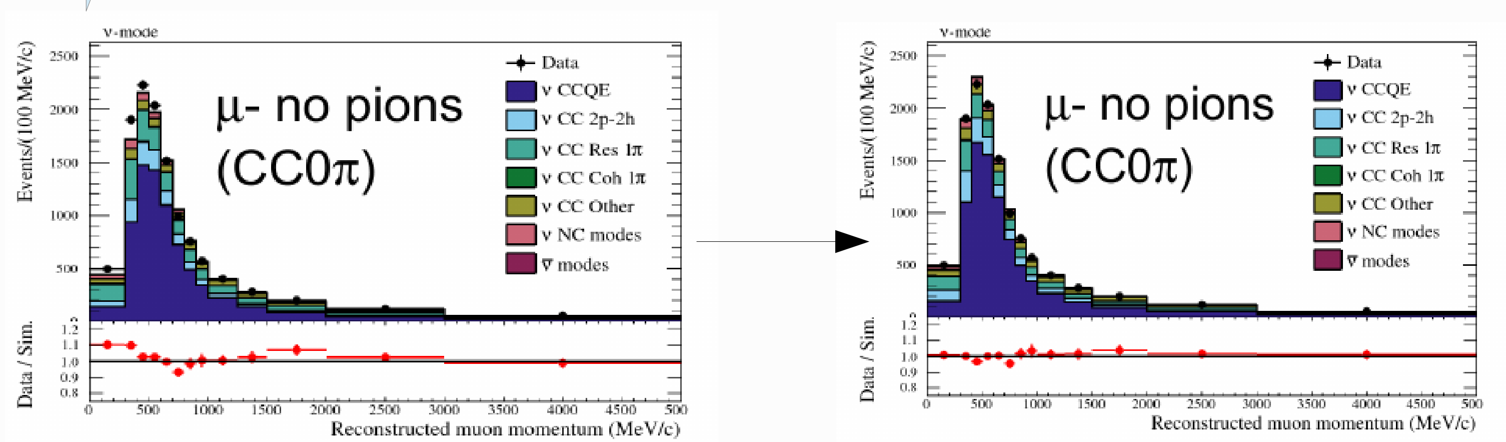
Plots from S.Dolan

The T2K oscillation analysis

MODEL WITH PRIOR UNCERTAINTIES

- **Neutrino flux** constrained from hadro-production experiments (NA61)
- **Neutrino interaction cross-section** constrained from measurements at bubble chambers, Minerva, ...
- **Detector simulation** (ND280 and SK) with systematics constrained in control samples

FIT THE MODEL TO THE ND280 and SK DATA



LIKELIHOOD as a function of

- **PMNS** parameters of interest (~6)
- **Nuisances:**

~10 for flux in each configuration (ν_μ,$\bar{\nu}_\mu$,ν_e,$\bar{\nu}_e$)

~20 for xsec in C, H and O

~hundreds for detector systematics

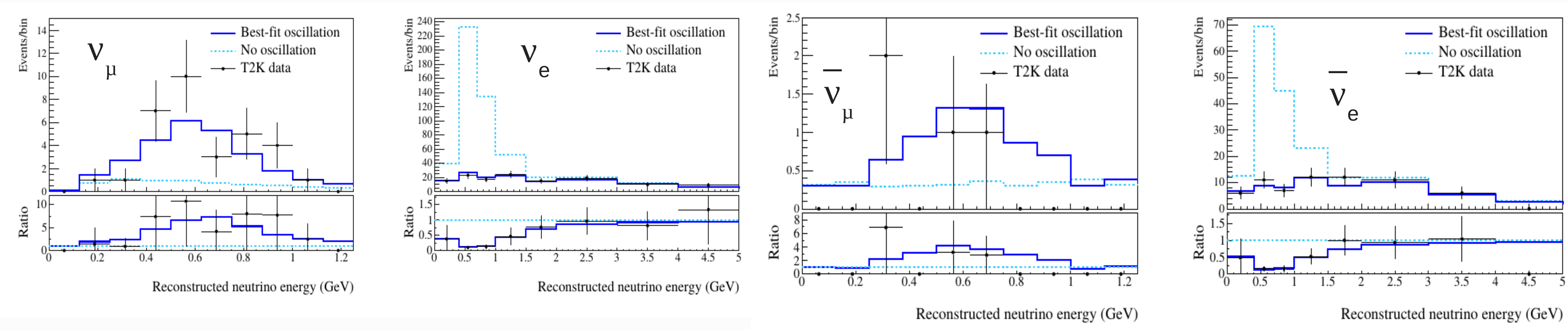
the technical challenge!

The T2K oscillation analysis

MODEL WITH PRIOR UNCERTAINTIES

- **Neutrino flux** constrained from hadro-production experiments (NA61)
- **Neutrino interaction cross-section** constrained from measurements at bubble chambers, Minerva, ...
- **Detector simulation** (ND280 and SK) with systematics constrained in control samples

FIT THE MODEL TO THE ND280 and SK DATA



LIKELIHOOD as a function of

- **PMNS** parameters of interest (~6)
- **Nuisances:**

~10 for flux in each configuration (ν_μ,$\bar{\nu}_\mu$,ν_e,$\bar{\nu}_e$)

~20 for xsec in C, H and O

~hundreds for detector systematics

the technical challenge!

Correlations!

Proper correlations between the nuisance parameters in particular for common external constraints and theoretical uncertainties.

Examples from T2K:

- Neutrino cross-section uncertainties: need to use the same interaction model to explicitate the correlations?
- Flux constraints in NOVA also partially from NA61 measurements...

The conceptual (and sociological) challenge!

Need a lot of communication in order to avoid 'hidden' manipulation of the likelihoods and agree on compatibel likelihood parametrizations ...

Summary

- Aim to document **recipes to properly combine results from different experiments** (and provide dedicated tools, if personpower available)

- This will be an **hot topic both for T2K and Belle2** during the JENNIFER2 lifetime

- Different solutions are available and were pursued in the past, we will promote the **combination of native likelihoods** in T2K and Belle2 communities

c) Statistical methods for combinations of experimental results

The reach of many crucial measurements of the T2K and Belle2 programs is severely limited by the small size of the event samples used. In this scenario, completely common for neutrino and quark flavour experiments, the combination of the statistical information from multiple measurements has significant potential to enhance the physics reach over the bare combination of the final results. Past results combination attempts have typically been conducted on an ad-hoc basis and after the individual measurements and their methodological choices and approximations had been consolidated. This results in suboptimal combinations limiting the statistical power of the outcomes.

Each individual measurement typically involves a large number of estimated parameters: the physics parameters of interest and many nuisance parameters correlated with them. While the former can be reasonably cast in an universal experiment-independent format and treated consistently in combinations, the latter are partly universal and partly experiment-dependent. This leads to a variety of possible options for the approximations and approaches needed to include their effect in the combination.

We propose a systematic and consistent plan for obviating the above pitfalls that consists in:

- A survey of the Belle2 and T2K physics topics and specific measurements where inter-experiment combinations (with NOvA, LHCb, etc.) have the potential to lead to significant reach enhancements.
- A survey of past and present combination efforts aimed at forming a global picture of the variance of the approaches adopted, the approximations made, and the possible pitfalls/inconsistencies encountered.
- A unified proposal for: (i) restricting the definition of the relevant physics and nuisance parameters for each measurement to one or few variants; (ii) restricting the approximations associated with the modelling of the interplay between nuisance and physics parameters to a few consistent variants. The proposal will be documented in a report that will serve as a reference for experimental groups willing to combine their results, which will be invited to conform to the selected prescriptions.

A possible development of such work could be the set up a software framework (e.g., a data base) explicitly suited and optimized for (i) accepting as inputs the values of multivariate likelihoods from each individual measurement and (ii) operating consistently the combination (likelihood multiplication) taking properly into account the commonalities between physics and global nuisance parameters and treating coherently experiment-dependent nuisance parameters. If successful, this work will enhance the physics reach of the single experiments both in neutrino and quark flavour physics.