# Forward Folding

## What? Why? How?

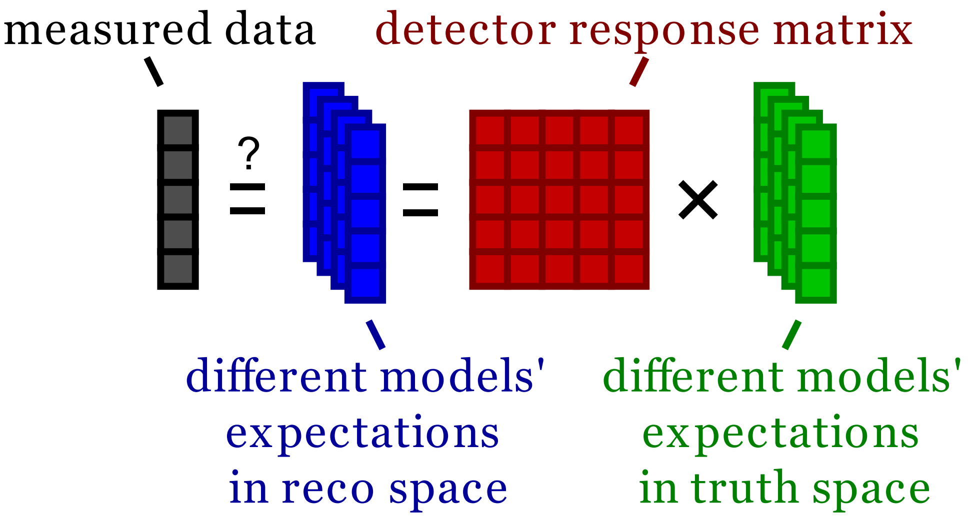
# The cross-section cave

- We do not record the events as they happen
  - We record only what we are able to see
- What we see is not what we are interested in
  - Lost events due to (in)efficiency
  - Added events due to background
  - Different event properties due to smearing

# The cross-section cave

- The canonical way: Unfolding
  - “Undo” the detector and selection effects
  - Challenging to do right without introducing bias
  - Can be wrong in very subtle ways
- Another way: Forward-folding
  - Apply detector effects to model predictions
  - Brings its own set of challenges

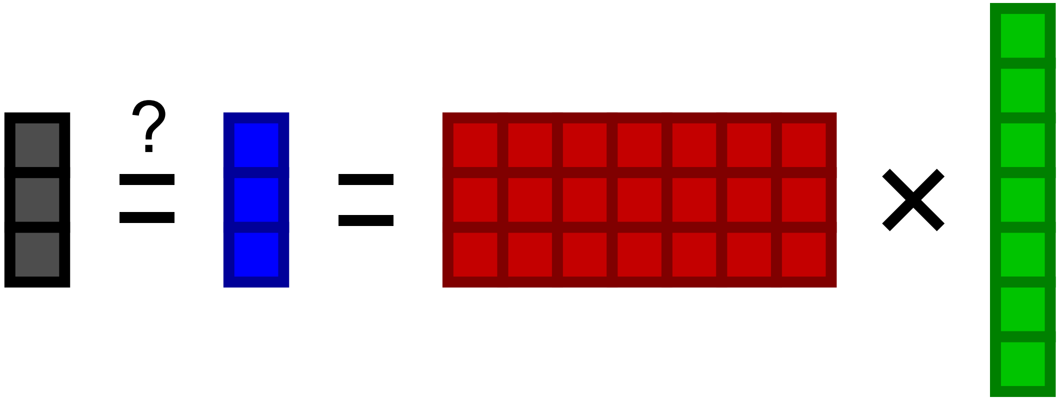
# How it works



- Every event belongs in exactly one truth bin and up to one reconstructed bin (if it gets reconstructed)
- P(reco bin = i | truth bin = j) = $R_{ij}$ = efficiency × smearing
  - Response matrix describes average detector response to true events
- reco expectation = response matrix × truth expectation
  - Can (and truth usually must) be binned in multiple variables

- The data is the data is the data
  - No uncertainty on the data points, 4 is exactly 4!
  - All systematics in response matrix or physics model
- All comparisons between data and theory (likelihoods, chi-squares, chi-by-eye) are done in reco space.
- Fast evaluations of many models possible

# Flexible binning



- Flexible number of bins
  - #reco bins ≠ #truth bins
- Combine coarse reco binning with fine truth binning
- Good for
  - low statistics (need large signal MC sample though)
  - difficult to constrain efficiency variations
- Admit we are not able to constrain truth completely

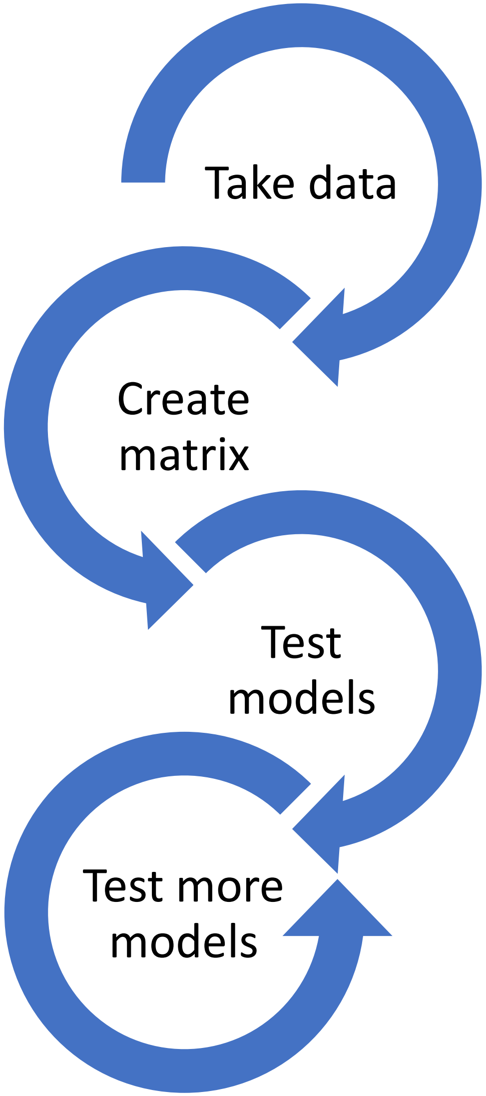
# Reco level data

- No data point correlation
  - Theory predictions will be correlated, but probably much less than what unregularised unfolding might do
  - Chi-by-eye
- Robert D. Cousins, Samuel J. May, Yipeng Sun, "*Should unfolded histograms be used to test hypotheses?*": [arXiv:1607.07038]

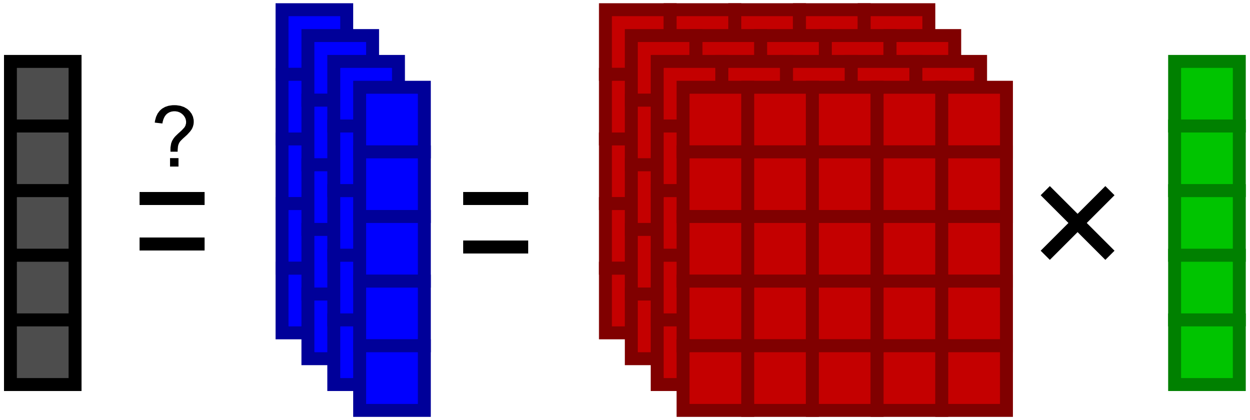
  "It seems remarkable that, even though unfolding by matrix inversion would appear not to lose information, in practice the way the information is used (linearizing the problem via expressing the result via a covariance matrix) already results in some failures of the bottom-line test of GOF. This is without any regularization or approximate EM inversion."

# Data life cycle

- Publication of only data and matrix probably not feasible (yet)
- Can do model comparisons or parameter fit as part of result
- Raw data & model independent response matrix ensure maximum usefulness of data in the future
  - New models can be easily compared against old data
  - Including all detector systematics



# Detector uncertainties

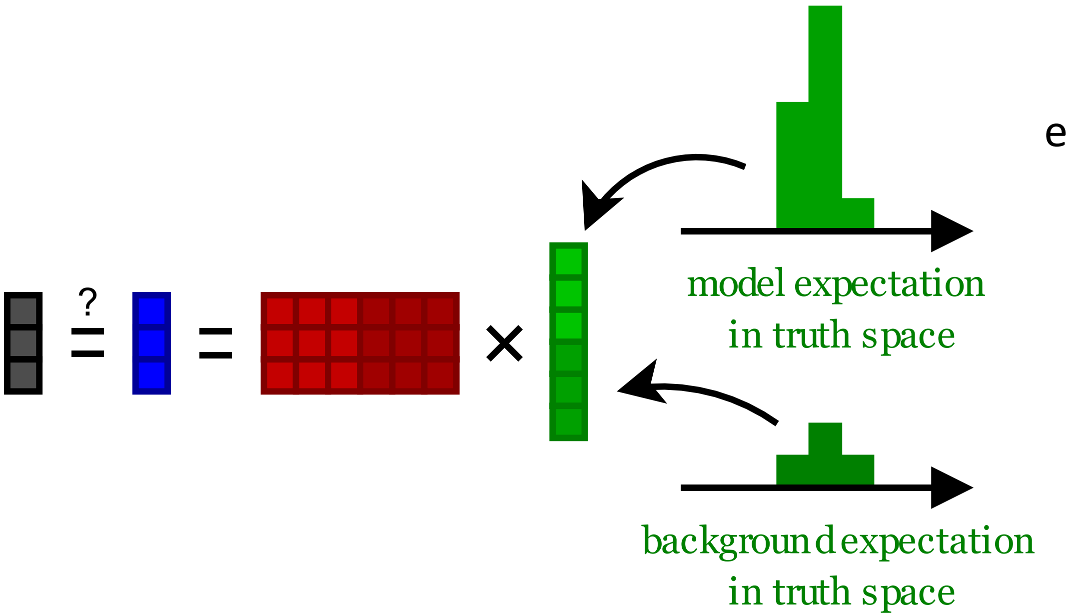


- One matrix only describes single possible detector
  - True detector probably behaves slightly differently
- Cover detector uncertainties with “toy simulations”
  - Variations and weights of same events
- Each toy yields own response matrix
- Each response matrix yields own reco prediction
- Compare to data w/ marginal, i.e. average, likelihood
  - No fitting of 1000 detector parameters!

# Background events

- Three kinds of backgrounds to distinguish:
- Irreducible background
  - Events that are indistinguishable from signal on truth level
  - E.g. CC-RES with pi lost in FSI in CC-QE selection (Don’t do this)
  - Must be added to the respective truth bins
  - Can be constrained with control samples (model dependent!)
- Physics-like background
  - Events with their own defined efficiency and smearing
  - E.g. CC-1pi events in CC-0pi selection
- Detector background
  - Events where it is difficult to define true events
  - E.g. certain kinds of OOFV

# Physics-like background

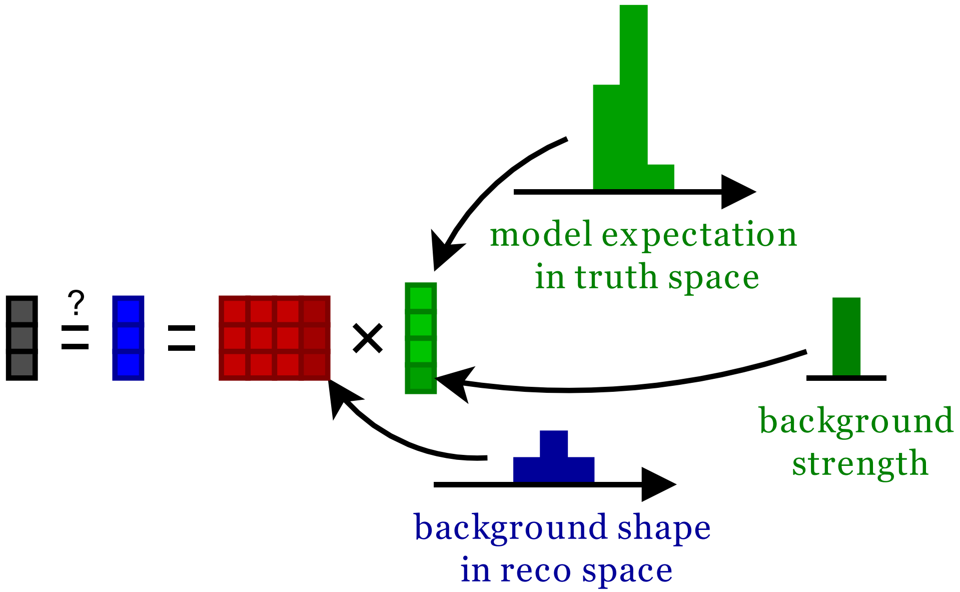


e.g. CC-0pi

e.g. CC-1pi

- Background where one can define a reasonable efficiency and smearing will be treated just like signal
- More columns in the response matrix
  - Equivalent to getting their own response matrix
- Users of the data might not care about those events
  - Provide templates or other simple models to go with the data

# Detector background

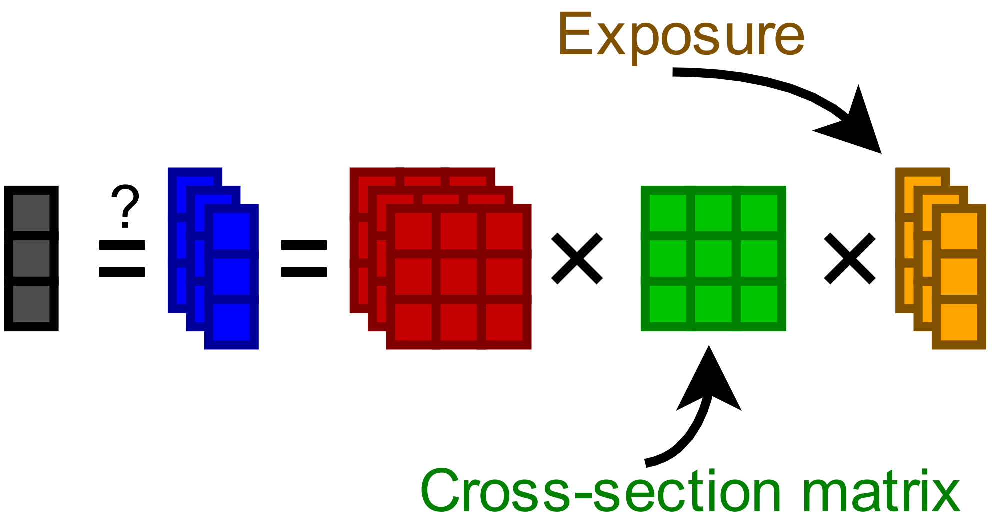


- Background without a reasonable truth definition can be added as pure reco shape in the matrix
- Background strength is single bin in truth vector
- Equivalent to using background template for physics-like background
  - Not possible to change BG model in the future

# From events to cross sections

- So far implicitly only talked about event rate predictions
- Models predict cross sections
- Need flux and target mass to go from cross section to event rates
- Problem: We do not know those numbers
  - Flux uncertainty
  - Fiducial mass uncertainty
- Cannot compare prediction using nominal flux with data “using” real flux!
- Give data users tools to vary the flux prediction

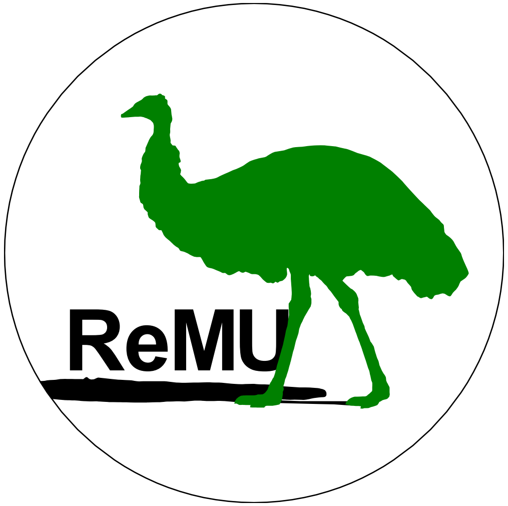
# Flux forward folding



- Model predicts cross section for each flux bin
- Provide set of flux exposures according to uncertainties
  - Exposure = flux × time × target mass
- Flux and detector uncertainties can be correlated
  - Make one response matrix correspond to one exposure vector

# Response Matrix Utilities

- Implements all of this
- Input:
  - Toy variations of selection (detector systematics)
  - Truth and reco binning
- Provides methods to:
  - Bin data in very flexible binning classes
  - Build matrix and evaluate uncertainty on elements
  - Forward-fold models to reco space
  - Compare to data (e.g. compute likelihoods, p-values, MCMC)
- Pure python (+ standard scientific packages numpy, etc)
  - Easy to install and use
  - $ pip install remu
- Tell me what you expect/want/need!



# Work in progress warning

- Exposure treatment not implemented yet
- Currently in process of refactoring a lot of code
    - API is changing
- Will be done by time of tutorial before Tokai meeting
- Feel free to play with frozen versions!
- Be aware that implementation details are changing

# Further reading

- ReMU documentation: [remu.readthedocs.io]
  - Short introduction to forward folding
  - Examples showing how to use the software
- Forward folding method paper: [arXiv:1903.06568]
  - In depth description of the concept and all the maths
- Talk to me

# Thank you!

# What to bin in

- Ideal:
  - Bin in all truth variables that affect reconstruction/selection
- This goes beyond the variables of physics interest, i.e. the reco variables
  - Measuring muon momentum distribution but true $\cos(\theta)$ affects the efficiency? Need to bin in $\cos(\theta)$.
  - Might lead down some weird rabbit holes (angular separation of tracks, total (charged) particle multiplicity, …)
- Realistic:
  - Bin in most important variables that affect reconstruction
  - Understand the detector!
  - Understand the selection!
  - Check the efficiencies!
  - Think of ways how nature could ~~screw you~~ break your analysis

# What not to bin in

- Never ever use truth variables that need a “physics” model to propagate to the reco level!
- Neutrino energy? Probably a bad choice.
    - Measurable effect depends on interaction model, nuclear model, FSI…
- Muon momentum? Good choice.
    - Directly accessible by detector (track curvature)
- HMN momentum? Even better choice!
    - Do you assume the muon to be selected as HMN?
    - What about confusion with high-momentum pions?
- Rule of thumb:
    - Bin in variables as “close” to reconstructed quantities as possible
- Must be predictable by external users
    - “Expected range in FGD2” not well defined outside of collaboration
    - “Kinetic energy” and “particle mass” are

# Are we doing cross-section model comparisons wrong?

On some level, yes.

# Status quo

- Cross section measurements are “flux integrated”
  - Measured cross sections are valid only for a specific neutrino flux
- Unfolding procedure uses flux uncertainty to evaluate effect on results → part of covariance matrix
- Models use nominal flux for cross-section predictions
  - As far as I know
- χ2 is calculated using nominal model prediction and covariance matrix
  - Assumption: Flux uncertainties are included in covariance matrix

# Problem: Flux shape uncertainty

- Claim: Flux shape uncertainty is not (fully) included in the covariance matrix of the unfolded result when doing model comparisons
- Instructive example: “perfect” 1-bin measurement
  - All efficiencies perfectly flat
  - No background
  - Monochromatic neutrino beam with perfectly known intensity
  - Only systematic uncertainty is neutrino energy

$$\sigma = \frac{N}{\epsilon \Phi T}$$

# Problem: Flux shape uncertainty

- Variation of beam energy does not vary the result!
  - Systematic uncertainty of result = 0
  - Result is still correct

$$\sigma = N\frac{N}{\epsilon\Phi T} = const$$

- We know the flux integrated cross section very well
- We just do *not* know the flux shape very well
- When using only nominal flux for model comparison the flux shape uncertainty is ignored
  - Simple example: cross section proportional to E
  - Should introduce an additional uncertainty proportional to neutrino energy uncertainty

# Re-phrasing the problem

- Measurement provides best guess at cross-section integrated over the *real* flux profile
- Model predictions are calculated using the *nominal* flux profile
- Difference between nominal and real flux shapes is not taken into account when comparing the two
- Perfect monochromatic beam example:
  - Measurement: $\sigma(E_{real})$, well known
  - Model: $\sigma(E_{nominal})$, perfectly known
  - $\Delta E = E_{real} - E_{nominal}$, not well known, ignored in comparison

# Reality check

- Efficiencies are not perfectly flat
- Flux shape has some influence on result
  - Adds “something” to covariance matrix
- Flux shape is not dominant systematic (probably?)
  - Has flux shape effect on model predictions/comparisons been tested?
- Reality somewhere between “effect of flux shape is completely negligible” and “our χ2 are completely wrong”
  - How do we know where we are?