Profile likelihood unfolding with large number of bins

France-Berkeley PHYSTAT Conference on Unfolding

10–13 Jun 2024, LPNHE, Paris

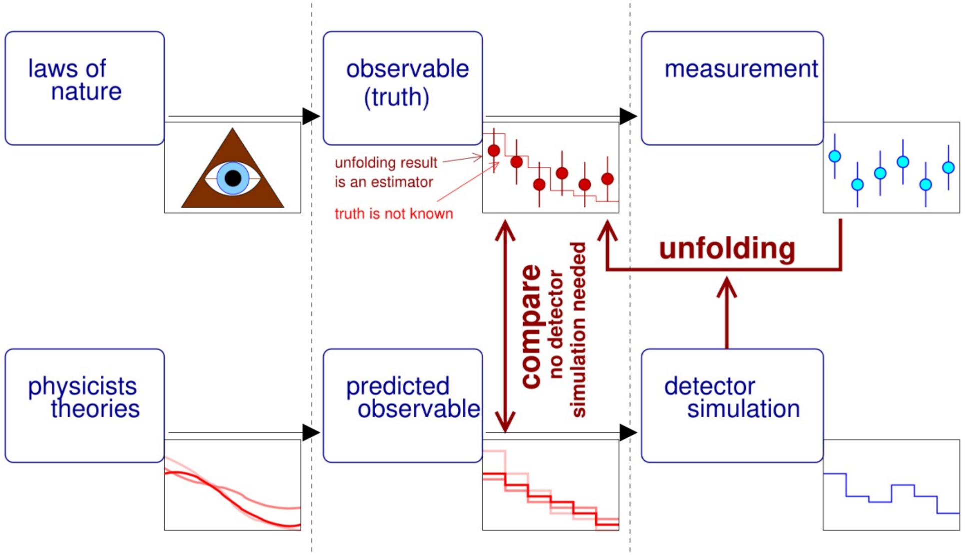
Josh Bendavid[1], Matteo Defranchis[2], David Walter[2]

2) CERN

1) MIT

CMS Compact Muon Solenoid

Overview



Binned maximum likelihood unfolding is the genuine solution to Poisson nature of counting experiments

$$\vec{n}|\vec{\lambda} \sim \mathrm{Poisson}(\mathbf{K}\vec{\lambda}+\vec{b})$$

- It gives the smallest unbiased variance

Binned profile likelihood unfolding

- 👍 Background subtraction accounted for directly in likelihood
- 👍 Systematic uncertainties accounted for directly during unfolding as nuisance parameters
- 👍 Profile nuisance parameters during unfolding to make most use of data
- 👍 Simultaneous fit across categories and all bins
- 👎 Expensive numerical minimization

Binned maximum likelihood unfolding

Any template shape fit can be expressed as a many-channel counting experiment, negative log likelihood can be written as

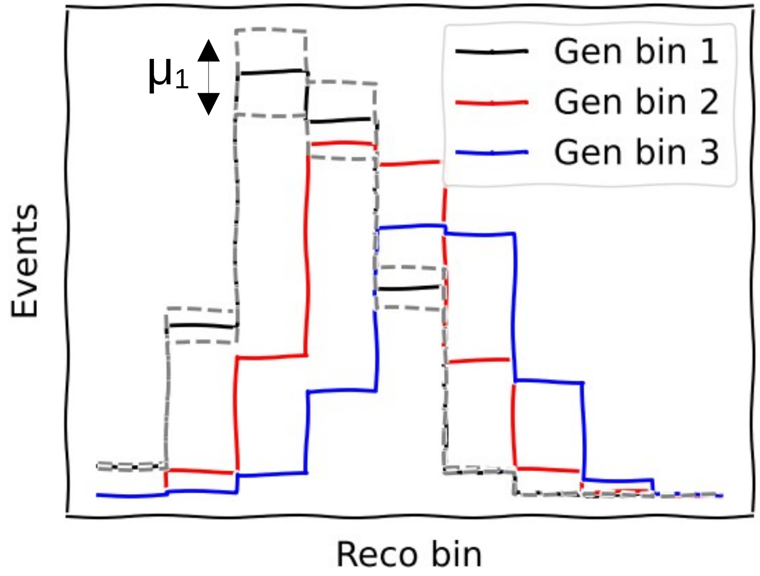
$$L = -\ln\left(\mathcal{L}(\mathrm{data}|\vec{\mu}\quad)\right) = \sum_{i}^{N^{\mathrm{reco}}} \left(n_i^{\mathrm{obs}} \ln n_i^{\mathrm{exp}}(\vec{\mu}\quad) + n_i^{\mathrm{exp}}(\vec{\mu}\quad)\right)$$

$n^{exp}_{i,p}$: Yield for each reco bin and process

$$n_i^{\mathrm{exp}} = \sum_{p}^{N^{\mathrm{procs}}} \mu_p n_{i,p}^{\mathrm{exp}}$$

μ_p: Signal strength modifier for each process

- Each gen bin is represented by a separate process (template), scaled by the μ_p



Binned profile maximum likelihood unfolding

Any template shape fit can be expressed as a many-channel counting experiment, negative log likelihood can be written as

$$L = -\ln\left(\mathcal{L}(\text{data}|\vec{\mu},\vec{\Theta})\right) = \sum_{i}^{N^{\text{reco}}}\left(n_i^{\text{obs}}\ln n_i^{\text{exp}}(\vec{\mu},\vec{\Theta}) + n_i^{\text{exp}}(\vec{\mu},\vec{\Theta})\right) + \sum_{k}^{N^{\text{syst}}}\frac{1}{2}\left(\Theta_k - \Theta_k^0\right)^2$$

$$n_i^{\text{exp}} = \sum_{p}^{N^{\text{procs}}} \mu_p n_{i,p}^{\text{exp}} \prod_{k}^{N^{\text{syst}}} \kappa_{i,p,k}^{\Theta_k}$$

$n^{\text{exp}}_{i,p}$: Yield for each reco bin and process

μ_p: Signal strength modifier for each process

- Each gen bin is represented by a separate process (template), scaled by the μ_p

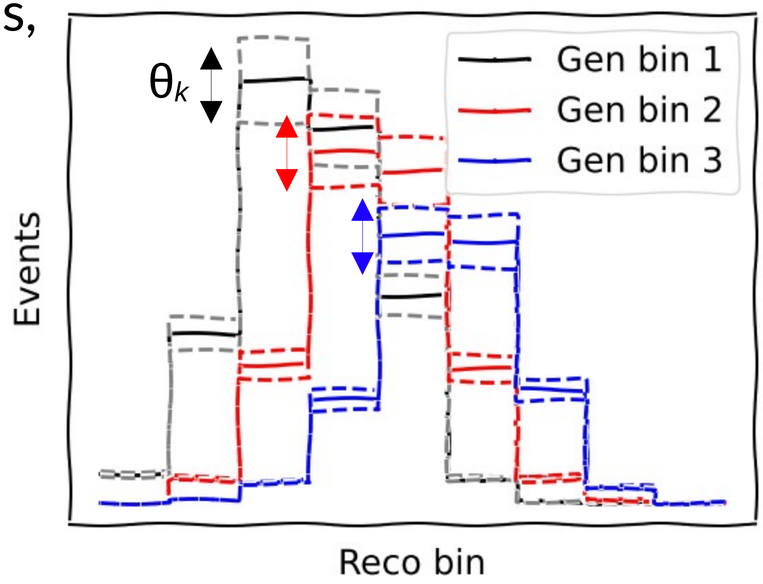
Θ_k: Nuisance parameter constrained to unit Gaussian for each systematic uncertainty

κ: size of systematic, 3D tensor of reco bins, processes, and nuisance parameters (log normal variations)

- The templates are scaled correlated for each systematic uncertainty

$$\kappa_{i,p,k} = 1 + \frac{v_{i,p,k}^{\text{exp}}}{n_{i,p}^{\text{exp}}}$$

- #Histograms ~ #gen bins · #systematics
- Potentially 10’s of thousands of histograms



Binned profile maximum likelihood unfolding

Uncertainties/ covariances can be inferred from likelihood function

$$L = -\ln\left(\mathcal{L}(\text{data}|\vec{\mu},\vec{\Theta})\right) = \sum_{i}^{N^{\text{reco}}} \left(n_i^{\text{obs}} \ln n_i^{\text{exp}}(\vec{\mu},\vec{\Theta}) + n_i^{\text{exp}}(\vec{\mu},\vec{\Theta})\right) + \sum_{k}^{N^{\text{syst}}} \frac{1}{2}\left(\Theta_k - \Theta_k^0\right)^2$$

- Exact: Likelihood scan – computationally expensive/slow
- In Gaussian limit: Covariance = Inverse Hessian matrix (second derivative)

In summary

- Challenging minimization problem
- Convergence to global minimum required
- Uncertainties/ covariances need to be computed accurately
- Time/memory of minimization must be kept under control

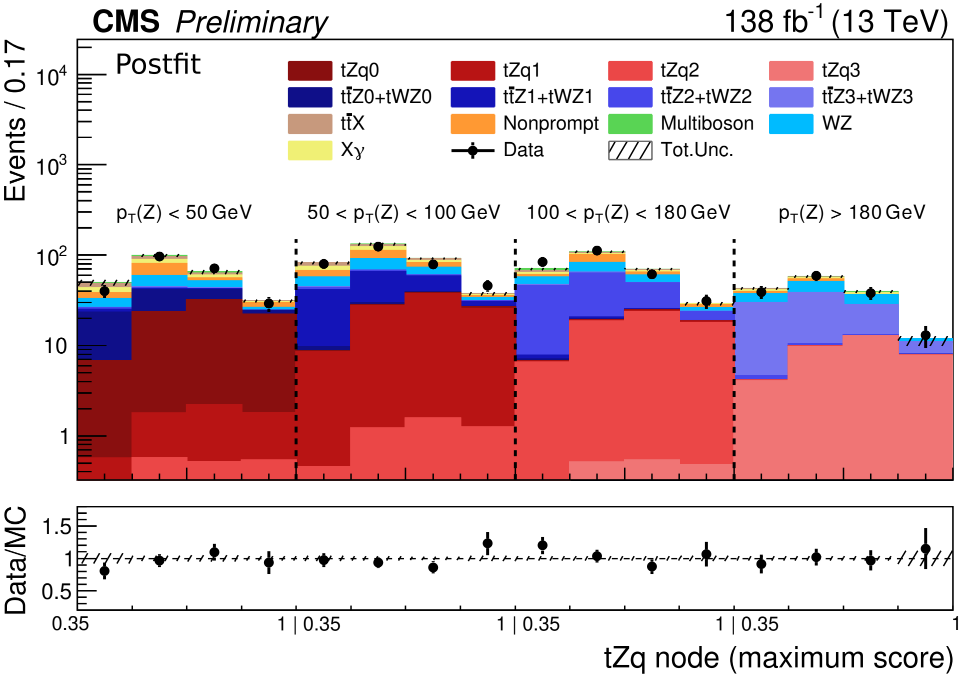
$$\mathbf{H}_f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

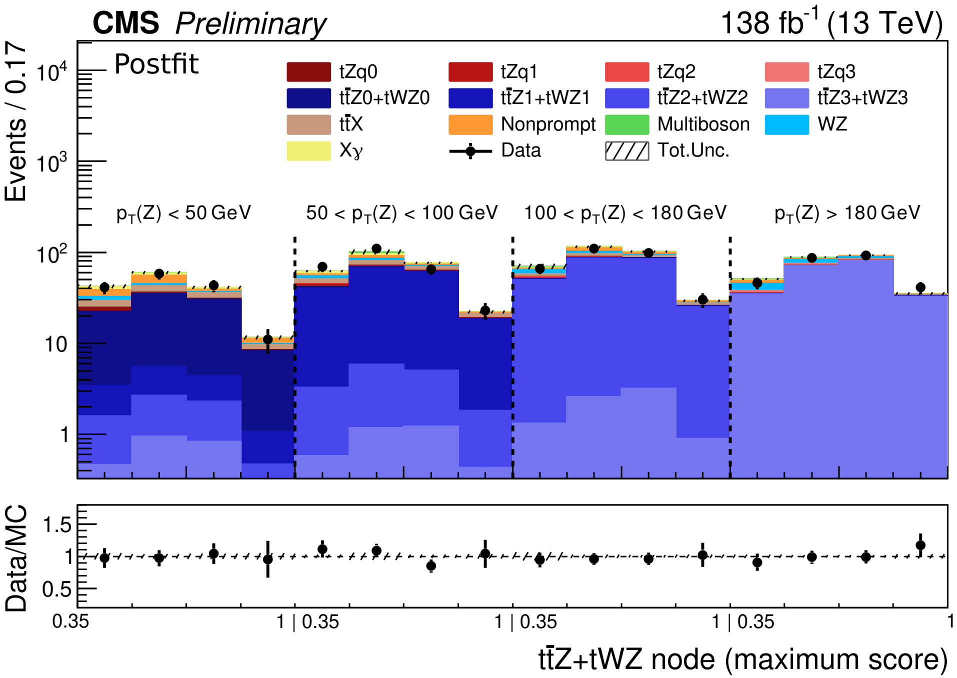
[CMS-PAS-TOP-23-004]

Example 1: multi process unfolding in CMS

tZq, ttZ, and tWZ are mutual backgrounds

- Simultaneously unfold differential cross sections
- Combined fit of tZq and ttZ enriched selection
- Fit reco variable and event classifier to separate gen bins and processes



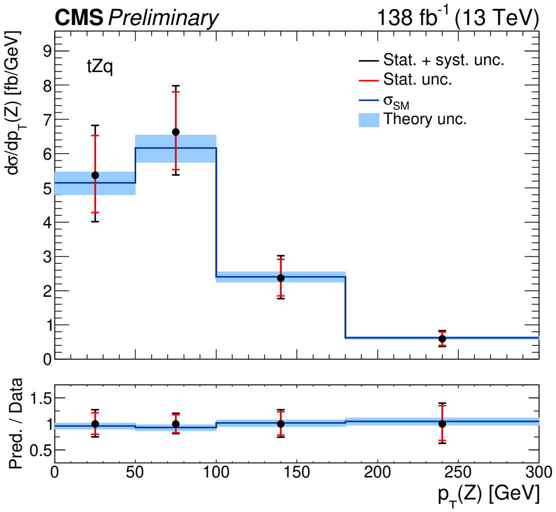


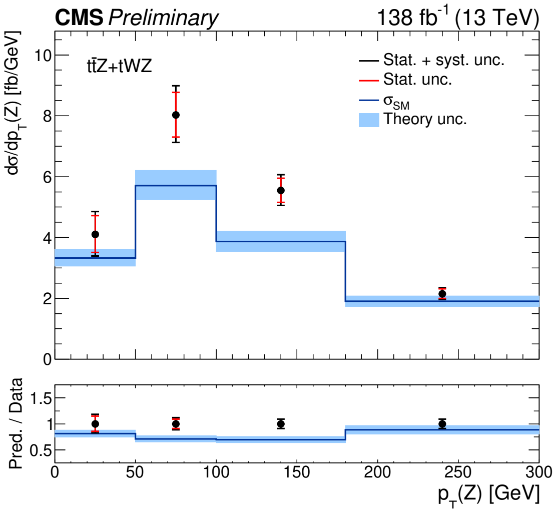
[CMS-PAS-TOP-23-004]

Example 1: multi process unfolding in CMS

Obtain unfolded cross sections together with full covariance matrix

- Allow consistent re interpretation e.g. in EFT (operators effect both processes)





Specifications: 144 reco bins, 8 gen bins 13 processes in total, ~400 systematics

- O(1,000) histograms
- Computationally “easy” with today's hardware
- Done on traditional CMS way with Combine (roofit via minuit)

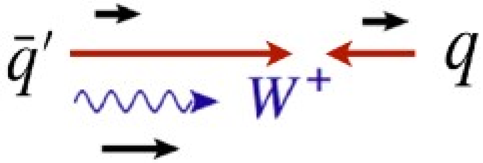
[PRD 102 (2020) 092012]

Example 2: W helicity in CMS

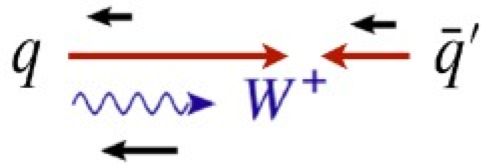
At LO at LHC, W produced via $q\bar{q}$

Due to pure left handed coupling, W helicity determined by its direction relative to incoming quark

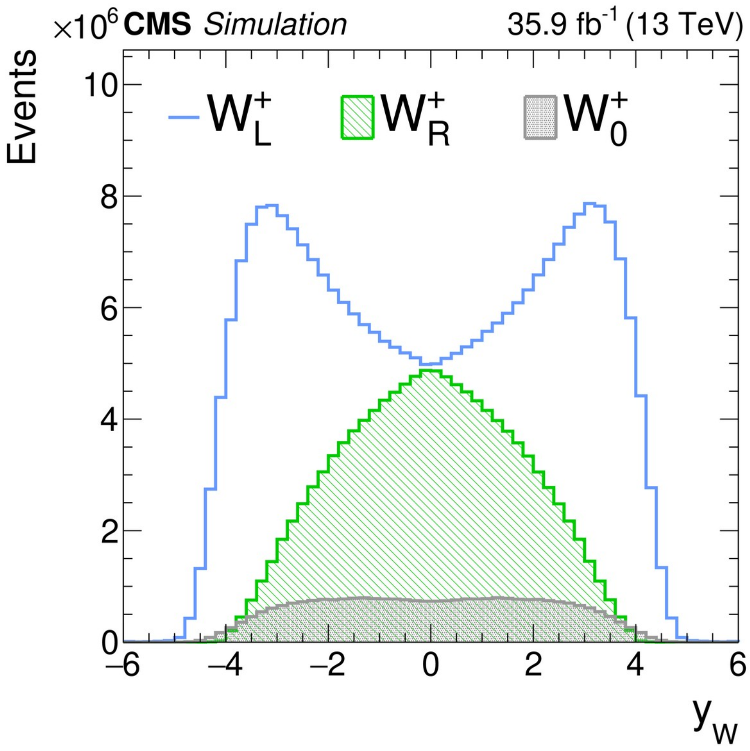
- W helicity contains information about PDFs

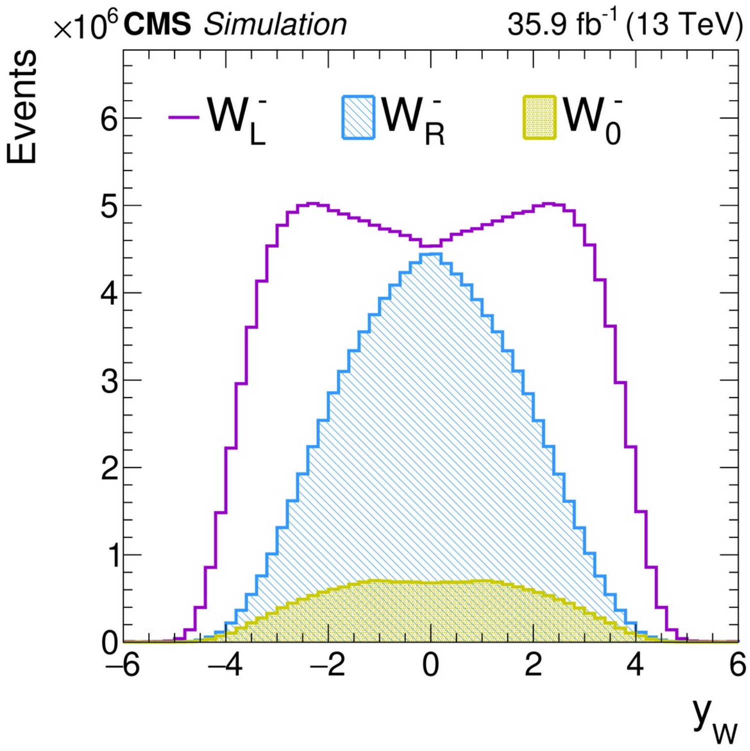


Left-handed



Right-handed





[PRD 102 (2020) 092012]

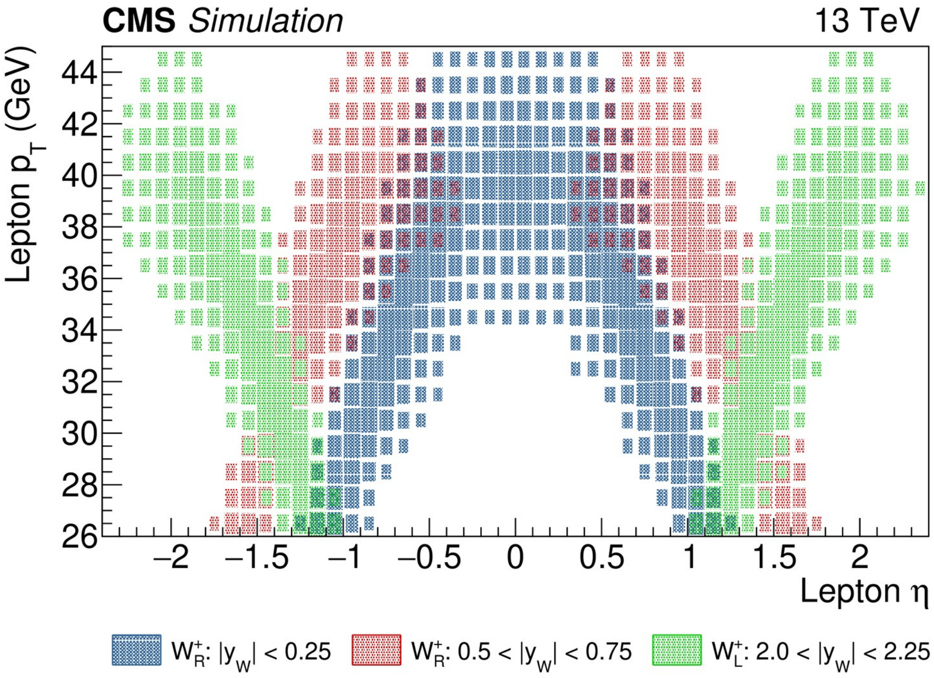
Example 2: W helicity in CMS

Decay (anti) lepton prefers to travel (alongside) against direction of W spin

- Polarization states can be extracted from charged lepton |η| - p_T distribution
- Avoid dependence on less precise MET

Measure transverse polarization states for W+ and W- in bins of boson rapidity

- Longitudinal component fixed to theory prediction with inflated uncertainty
- Separate signal template for each gen bin

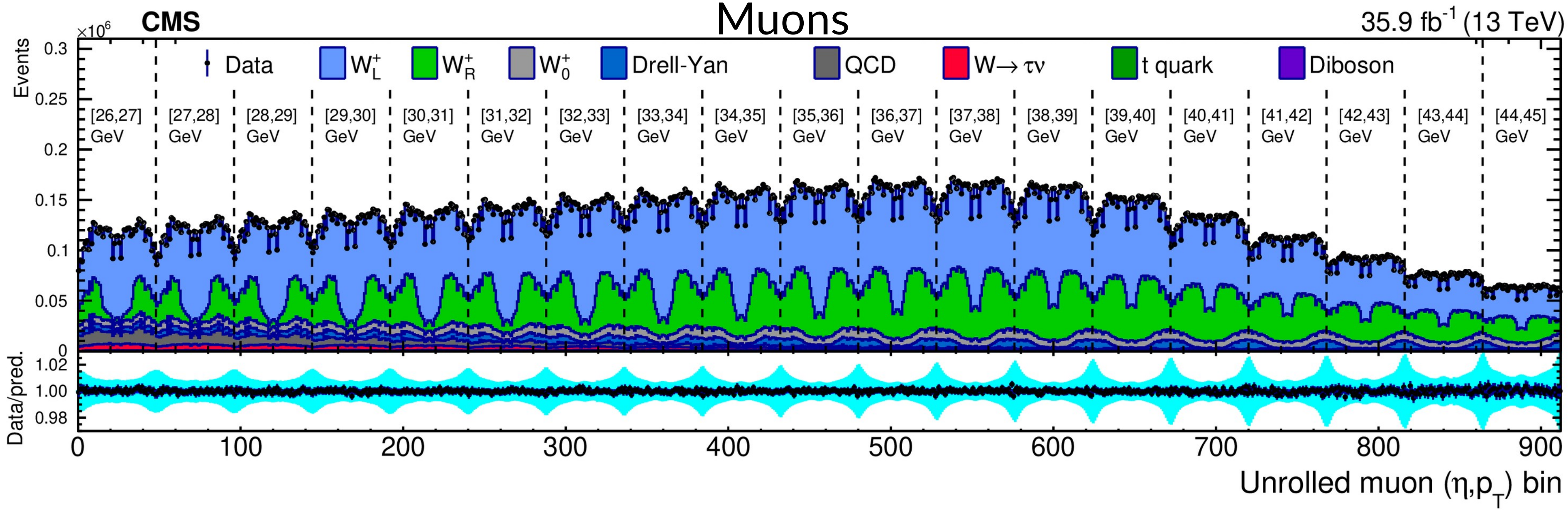


[PRD 102 (2020) 092012]

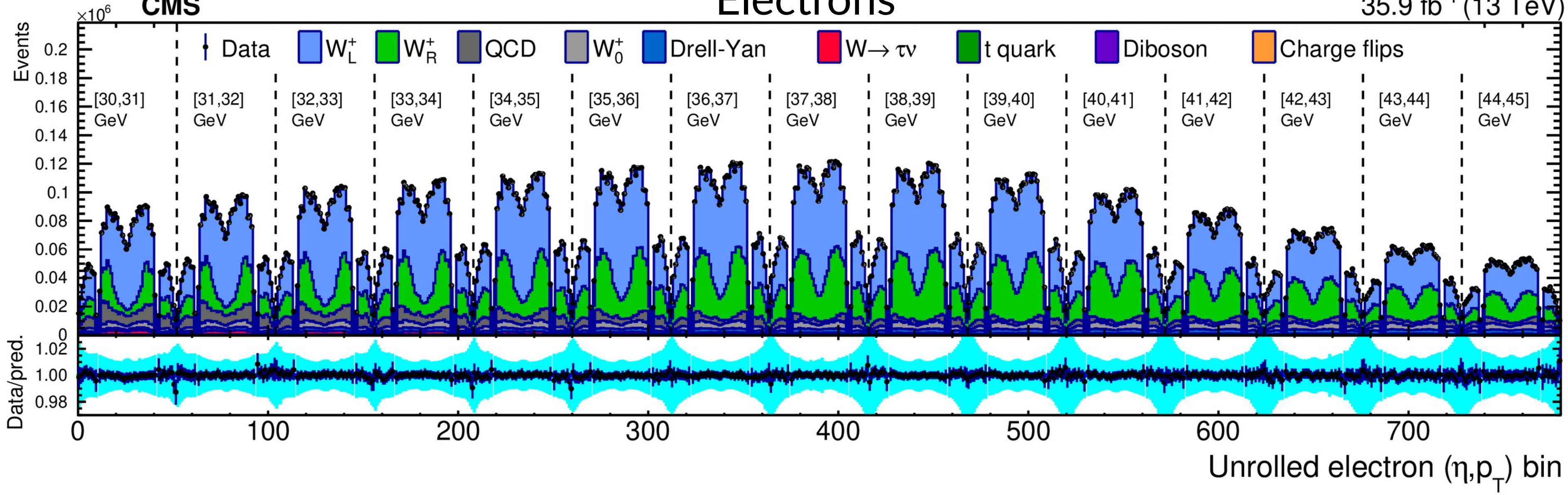
Example 2: W helicity in CMS

Reco distributions for one charge

Muons



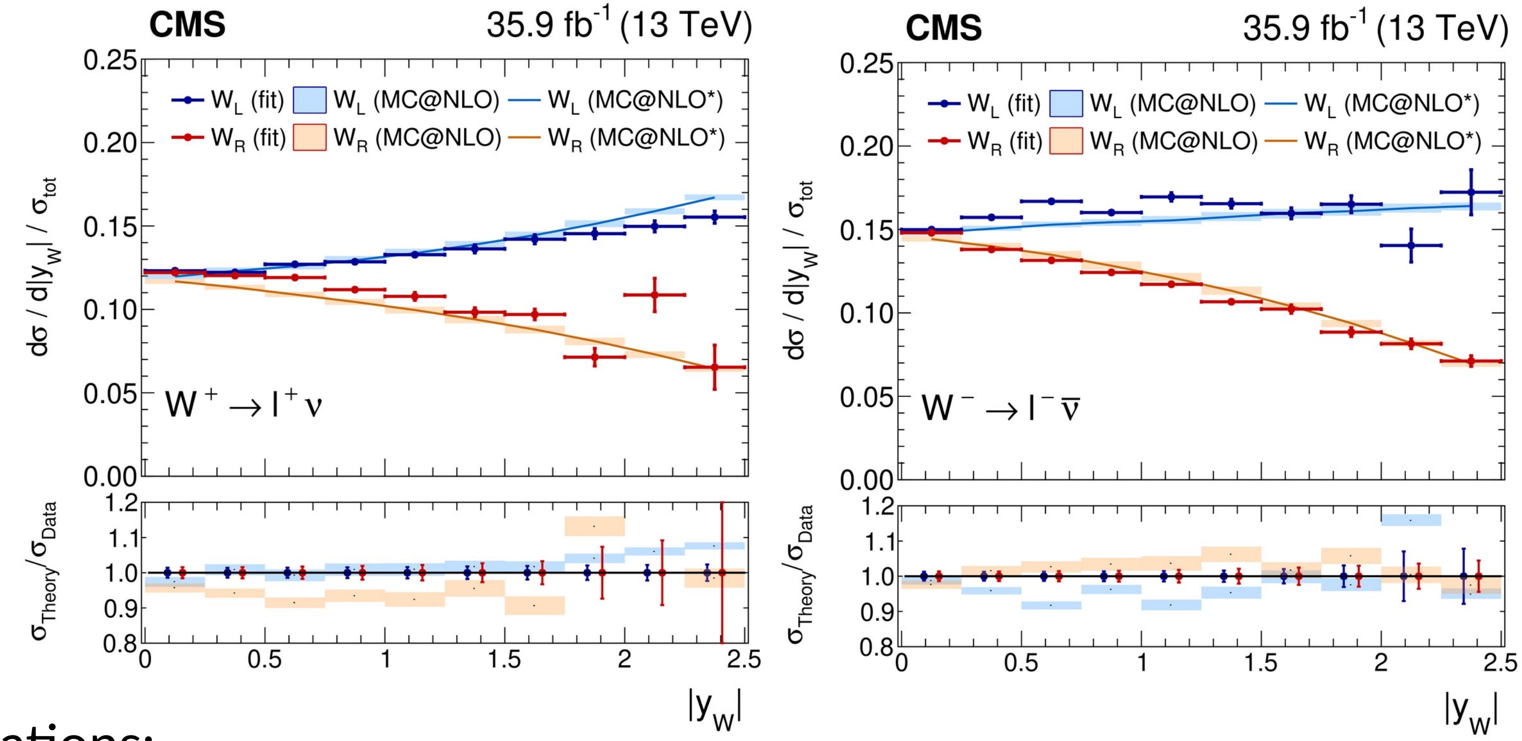
Electrons



[PRD 102 (2020) 092012]

Example 2: W helicity in CMS

Unfolded differential cross sections with full covariance matrix



Specifications:

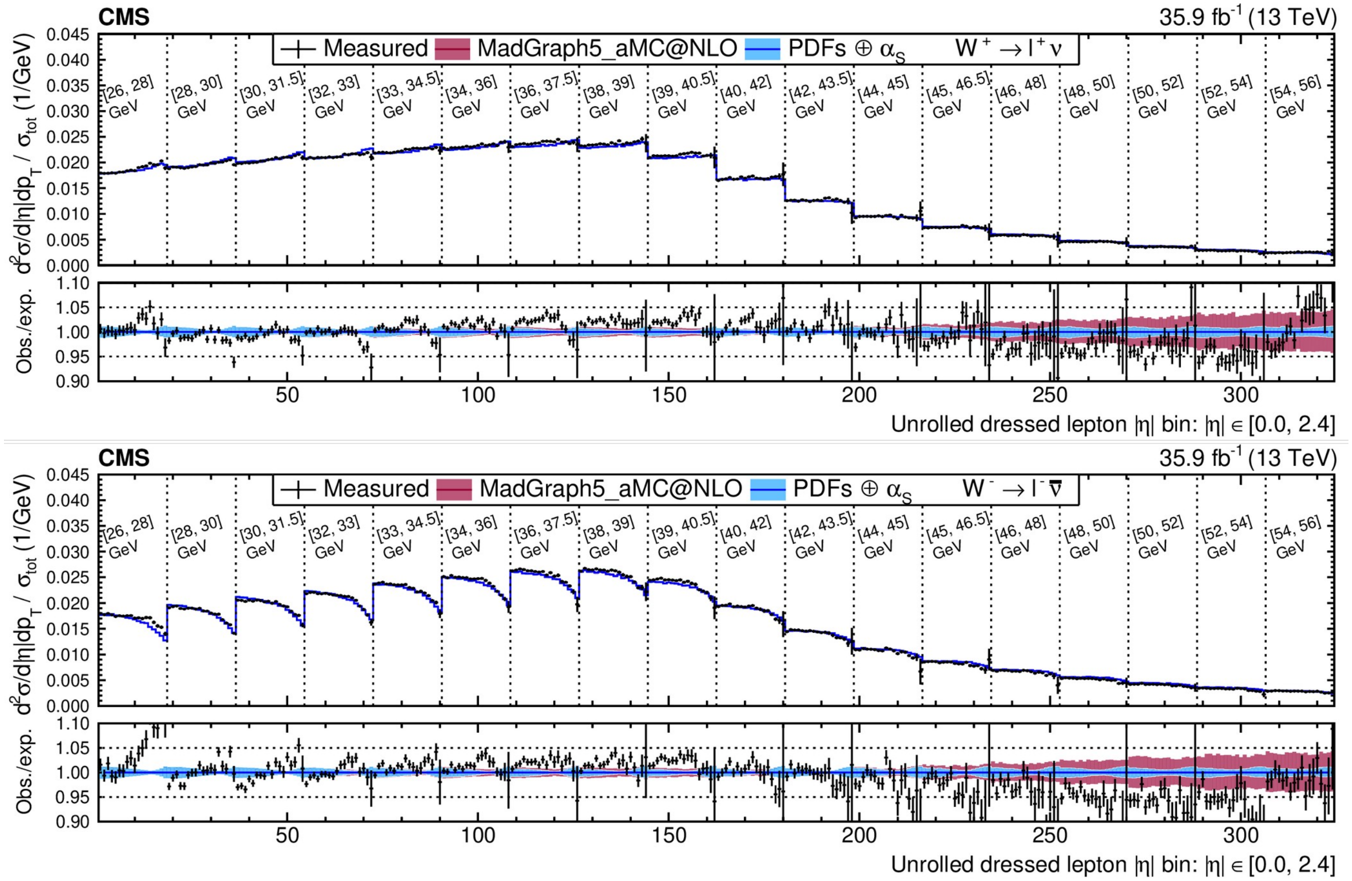
- 3320 reco bins, 40 gen bins, 78 processes in total
- 1354 systematic uncertainties
  - Theory (PDFs, QCD scale, …) Experiment (efficiencies, lepton energy/momentum scale, backgrounds, …)
- O(100,000) histograms
- It’s “challenging”

[PRD 102 (2020) 092012]

Example 2: W helicity in CMS

Also measured:

- Double differential cross section in |η| and p_T for W+ and W- simultaneously
- 2*18*18 = 648 gen bins
- 2448 reco bins
- 1051 nuisance parameters



The tool: minimization with tensorflow

Roofit via minuit insufficient

- Limited stability and efficiency (e.g. can not be parallelized)

Tensorflow library with automatic gradient computation via back propagation for minimization:

- Second derivative, trust region based minimizer to reliably find global minimum [arXiv:1506.07222]
- Fast, numerically accurate, stable
- Parallelized vector processing units and/or multiple threads
- Sparse tensor implementation to minimize memory consumption (if response matrix is close-to-diagonal, e.g. leptonic observables)
- See also [talk at PyHEP 2020 from J. Bendavid]

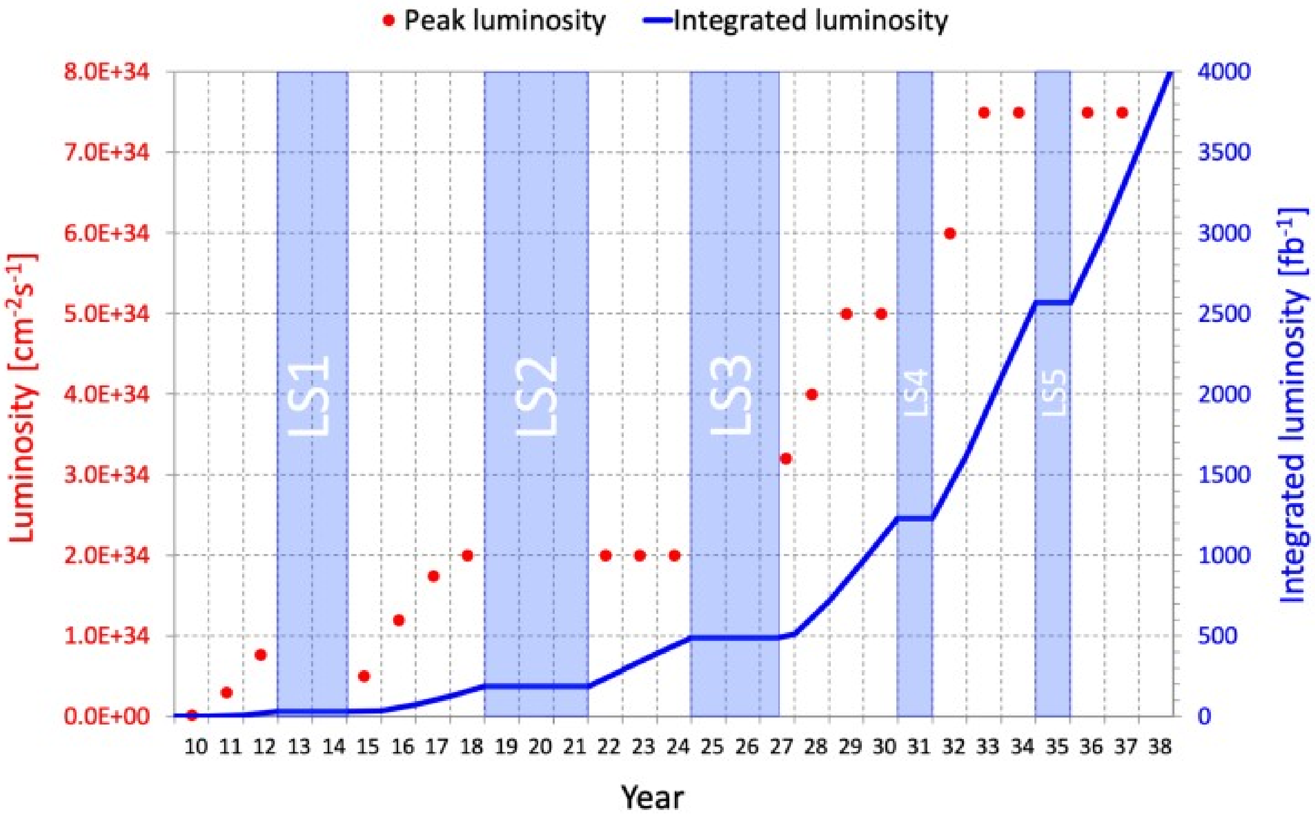
Future upgrade to newer tools foreseen, such as tensorflow 2, JAX, ...

- More efficient computation of hessian matrix

The future of profile likelihood unfolding

Previous analysis was "only" on 2016 data, a small fraction of current data and what will come with Run 3/ HL LHC

- More data will allow finer & more gen and reco bins
- More processes can be measured simultaneously
- Combination of data taking periods and improved precision will require more nuisance parameters

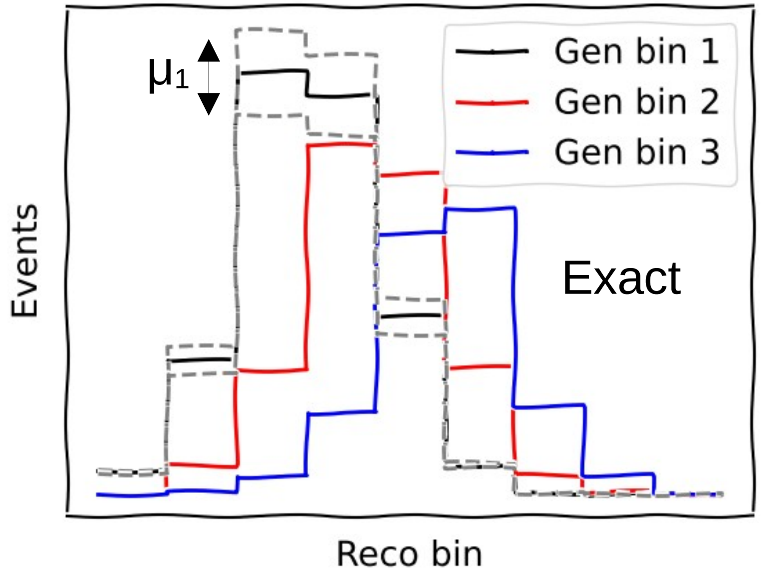


Linearized profile likelihood unfolding

3D tensor will grow to an unmanageable size

- Memory and computation
- #Histograms ~ #gen bins · #systematics

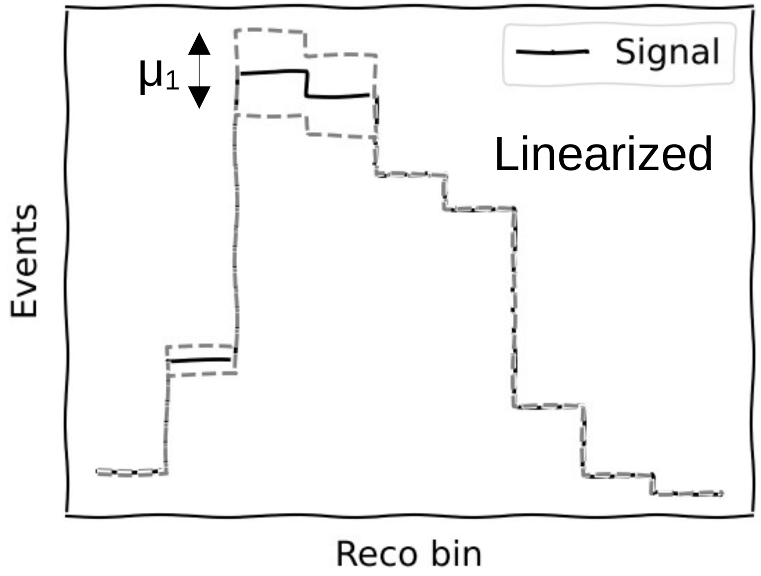
$$n_i^{\text{exp}} = \sum_p^{N^{\text{procs}}} \mu_p n_{i,p}^{\text{exp}} \prod_k^{N^{\text{syst}}} \kappa_{i,p,k}^{\Theta_k}$$



Linearize the dependence of the signal
(sum of gen bins) on nuisance parameters

$$n_i^{\text{exp}} = \sum_p^{N^{\text{bkg}}} n_{i,p}^{\text{exp}} \prod_k^{N^{\text{syst}}} \kappa_{i,p,k}^{\Theta_k} + n_{i,\text{sig}}^{\text{exp}} \prod_k^{N^{\text{syst}}} \kappa_{i,k,\text{sig}}^{\Theta_k} \prod_l^{N^{\text{gen}}} \kappa_{i,l,\text{sig}}^{\mu_l}$$

- Treating signal strength multiplier similar to nuisance parameters
- Signal tensor reduced to 2D
- #Histograms ~ #gen bins + #systematics



Linearized profile likelihood unfolding

$$n_i^{\text{exp}} = \sum_p^{N^{\text{bkg}}} n_{i,p}^{\text{exp}} \prod_k^{N^{\text{syst}}} \kappa_{i,p,k}^{\Theta_k} + n_{i,\text{sig}}^{\text{exp}} \prod_k^{N^{\text{syst}}} \kappa_{i,k,\text{sig}}^{\Theta_k} \prod_l^{N^{\text{gen}}} \kappa_{i,l,\text{sig}}^{\mu_l}$$

$$\boxed{\kappa_{i,k,\text{sig}} = 1 + \frac{v_{i,k,\text{sig}}^{\text{exp}}}{n_{i,\text{sig}}^{\text{exp}}}}$$

Size of systematic effect $v_{i,k,\text{sig}}^{\text{exp}}$ does not directly depend on individual gen bins anymore

- This assumption could potentially lead to a bias
- But signal strength modifier are unconstrained, starting value can be chosen freely
- Iterative procedure applied to mitigate bias

1) Initial fit

2) Re compute histograms with reweighting gen bin contribution via postfit signal strength modifiers

3) Repeat fit

Linearized profile likelihood unfolding

$$n_i^{\mathrm{exp}} = \sum_{p}^{N^{\mathrm{bkg}}} n_{i,p}^{\mathrm{exp}} \prod_{k}^{N^{\mathrm{syst}}} \kappa_{i,p,k}^{\Theta_k} + n_{i,\mathrm{sig}}^{\mathrm{exp}} \prod_{k}^{N^{\mathrm{syst}}} \kappa_{i,k,\mathrm{sig}}^{\Theta_k} \prod_{l}^{N^{\mathrm{gen}}} \kappa_{i,l,\mathrm{sig}}^{\mu_l}$$

$$\boxed{\kappa_{i,l,\mathrm{sig}} = 1 + \frac{v_{i,l,\mathrm{sig}}^{\mathrm{exp}}}{n_{i,\mathrm{sig}}^{\mathrm{exp}}}}$$

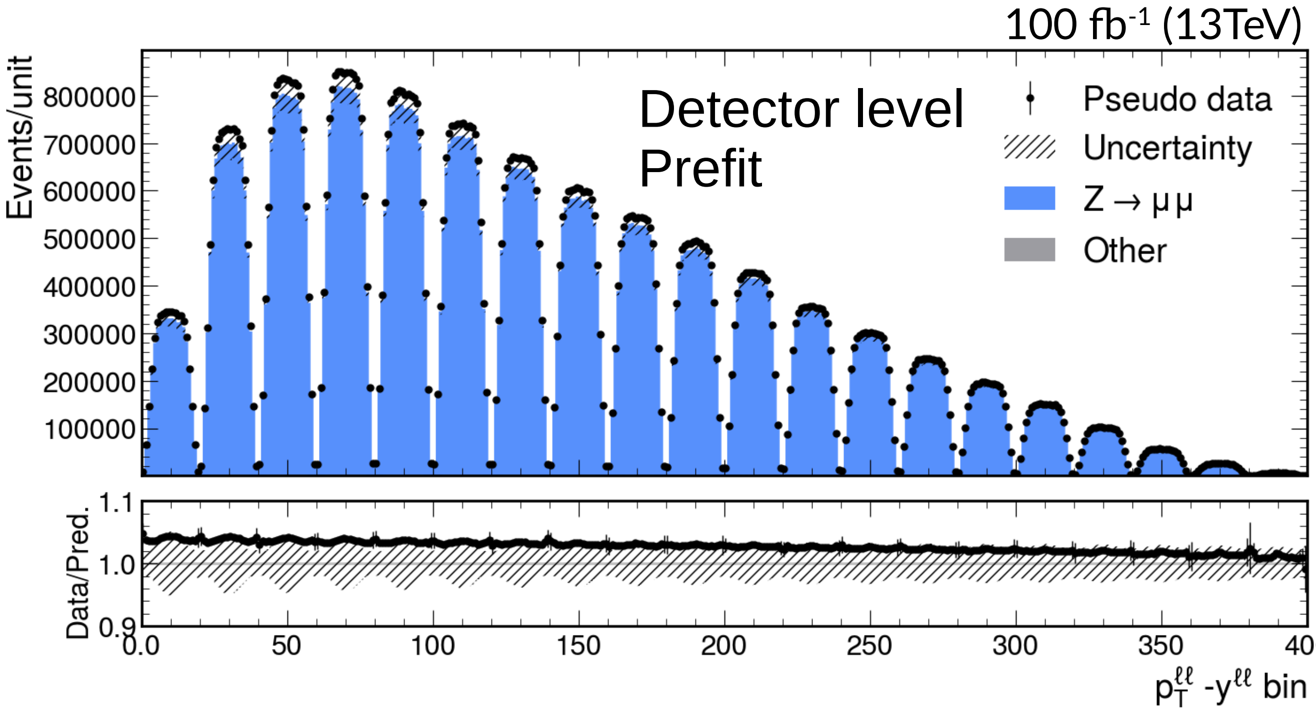
Size of gen bin variation $v_{i,l,\mathrm{sig}}^{\mathrm{exp}}$ is in principle arbitrary

- Small dependency observed – mainly for convergence of iterative procedure

Linearized profile likelihood unfolding

Validation: benchmark linearized likelihood unfolding vs. exact likelihood unfolding

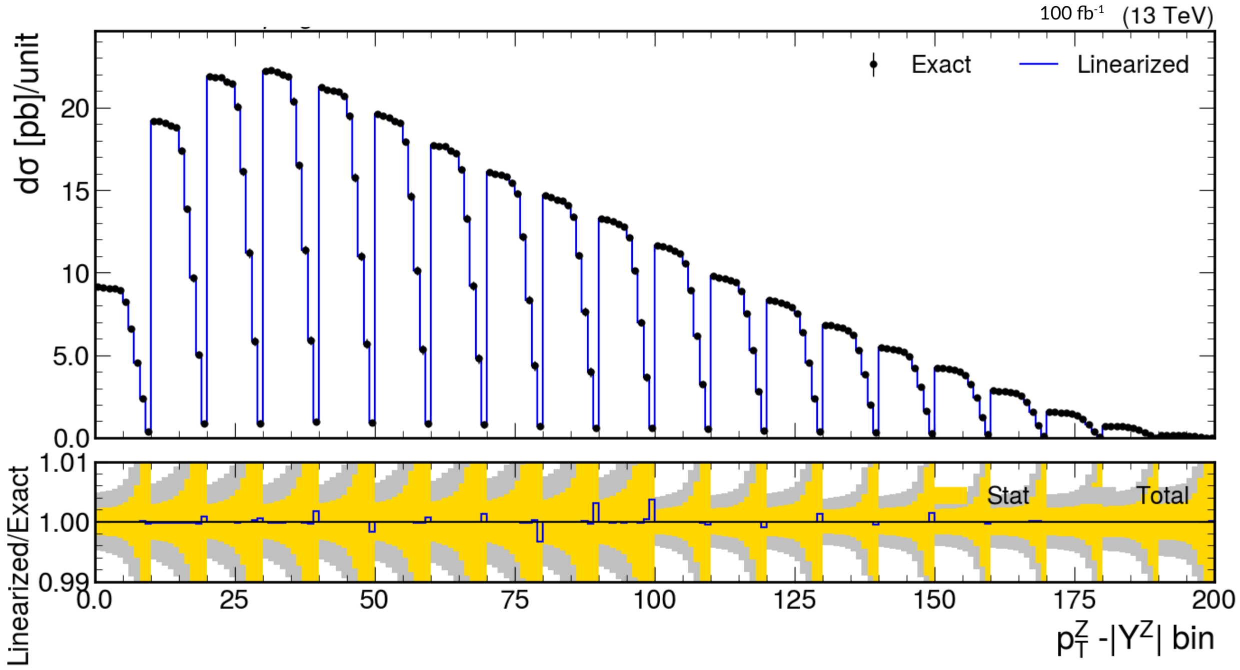
- Real world example: use MiNNLO MC with realistic detector simulation and unfold Z boson dilepton p_T, |Y|
  - 200 gen bins
  - 400 reco bins
  - 41 Explicit nuisance parameters (PDFs + α_S) + implicit MC stat. uncertainties
- Inject pseudo data by rewighting to HERAPDF2.0 PDF set (nominal is PDF4LHC21)
  - Do central values and uncertainties agree?



Linearized profile likelihood unfolding

Extracted differential cross section from initial fit

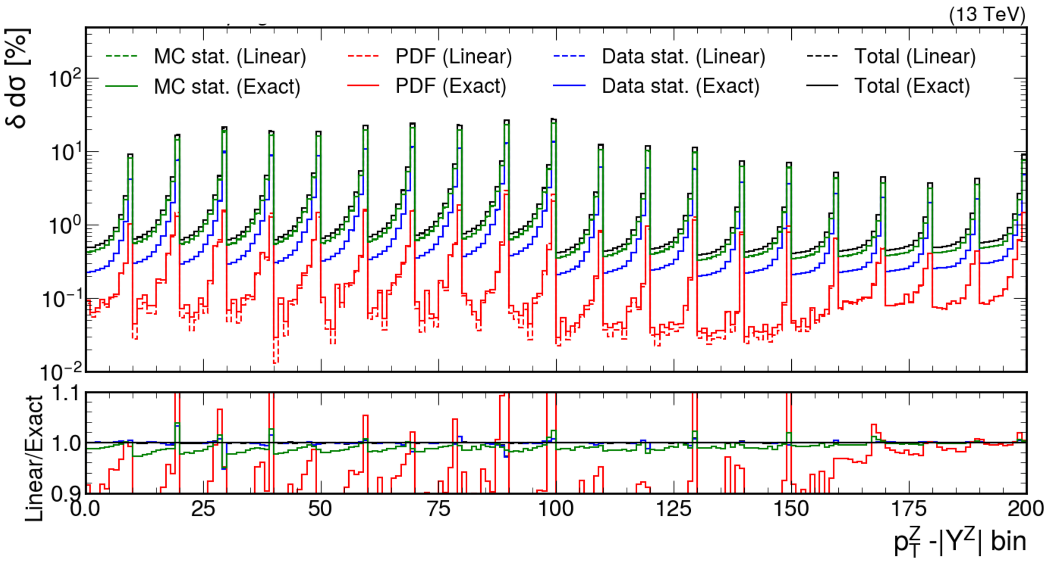
- Almost perfect agreement in central values between exact and linearized unfolding
- Deviations much smaller than stat. uncertainty



Linearized profile likelihood unfolding

Relative uncertainties from initial fit

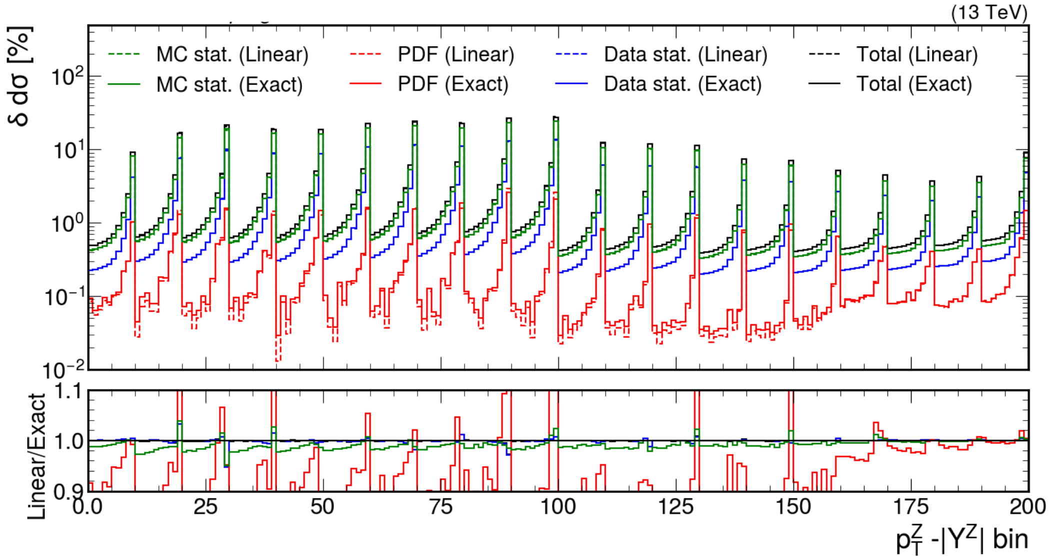
- Good agreement in total and data stat. uncertainty
- Larger relative disagreement for some individual sources of uncertainties e.g. MC stat. (green) and PDF (red)



Linearized profile likelihood unfolding

Relative uncertainties

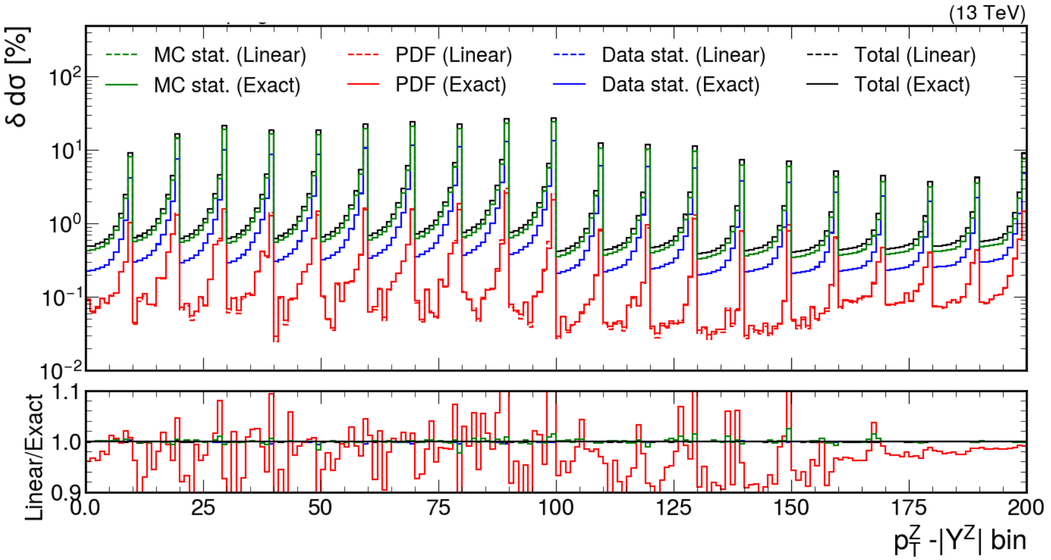
- Good agreement in total and data stat. uncertainty
- Larger relative disagreement for some individual sources of uncertainties e.g. MC stat. (green) and PDF (red)



Agreement improves through iterative procedure of linearized unfolding

- Re compute histograms with reweighting gen bin contribution via postfit signal strength modifiers – and repeat fit
- Size of gen bin variation $v_{i,l,\mathrm{sig}}^{\mathrm{exp}}$ chosen as 1% of gen bin contribution
  - Better choice possible, e.g. based on uncertainty
  - Studies ongoing

⇩ 1 iteration



Summary

Unfolded distributions provide input for global PDF/EFT/… fits

Binned profile likelihood unfolding is established as a reliable method

- Problem requires expensive numerical minimization

Modern libraries with automatic differentiation via back propagation allow robust and fast minimization

- Unfolding with up to 1000 gen bins well possible

However, complexity may grow soon to unmanageable level

- Linearization procedures can provide remedy
- Validation shows agreement with exact likelihood unfolding can be restored via iterative procedure (1 iteration sufficient in realistic toy study)

Backup