The inclusion of theory errors in PDF fitting.

The NNPDF4.0MHOU PDFs set

A. Barontini

University of Milan and INFN Milan

FH Particle Physics Monday Seminars, DESY

29/01/2024

Based on arXiv:1906.10698 and arXiv:2401.10319

Outline.

PHYSICS

- Why do we need PDFs?
- What are theory errors?
- How can we estimate them?
- Why is it relevant to include them in a PDF fit?

METHODOLOGY AND VALIDATION

- How does a NNPDF fit work?
- How can we include MHOU in a NNPDF fit?
- Can we validate our estimation?

RESULTS

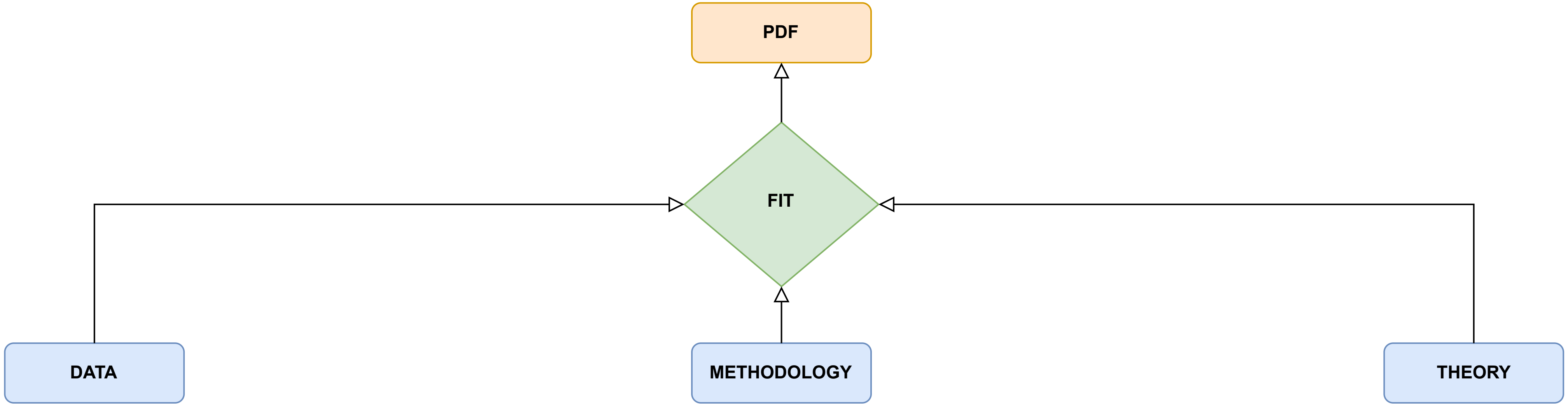
- Does the fit quality improve upon inclusion of theory errors?
- What is the impact on the PDFs?
- What about N3LO?

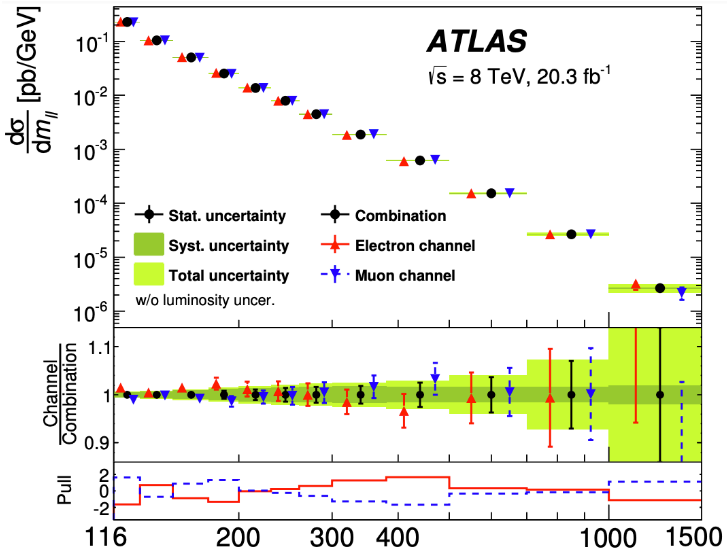
PHYSICS

- Why do we need PDFs?
- What are theory errors?
- How can we estimate them?
- Why is it relevant to include them in a PDF fit?

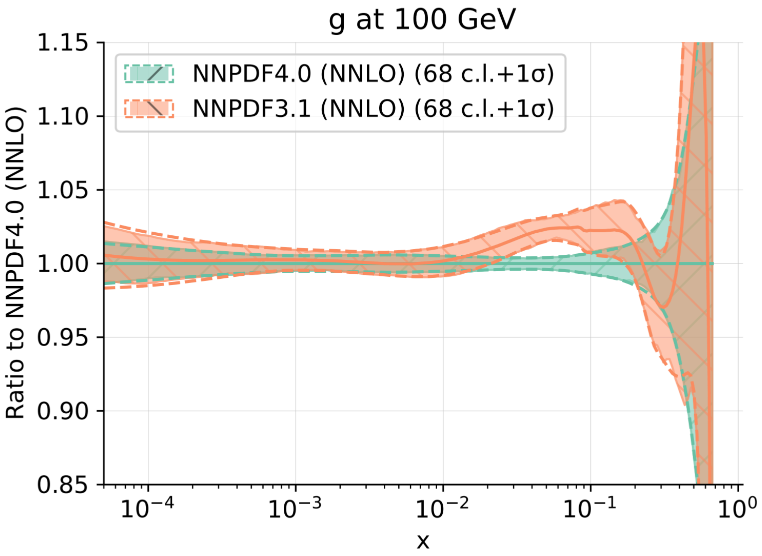
“We are not strangers, only the introduction is missing”
(Jesus Apolinaris)

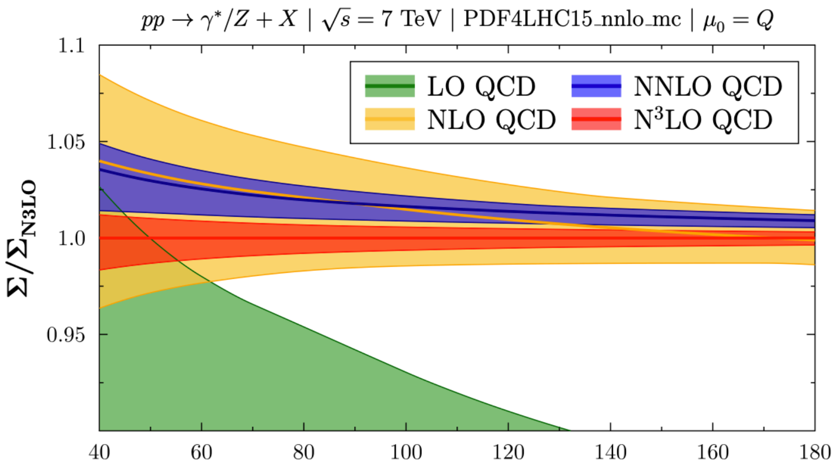
Motivation.





JHEP08(2016)009





JHEP12(2022)066

Describing a collision.

Thanks to **Factorization theorem**

PDFs

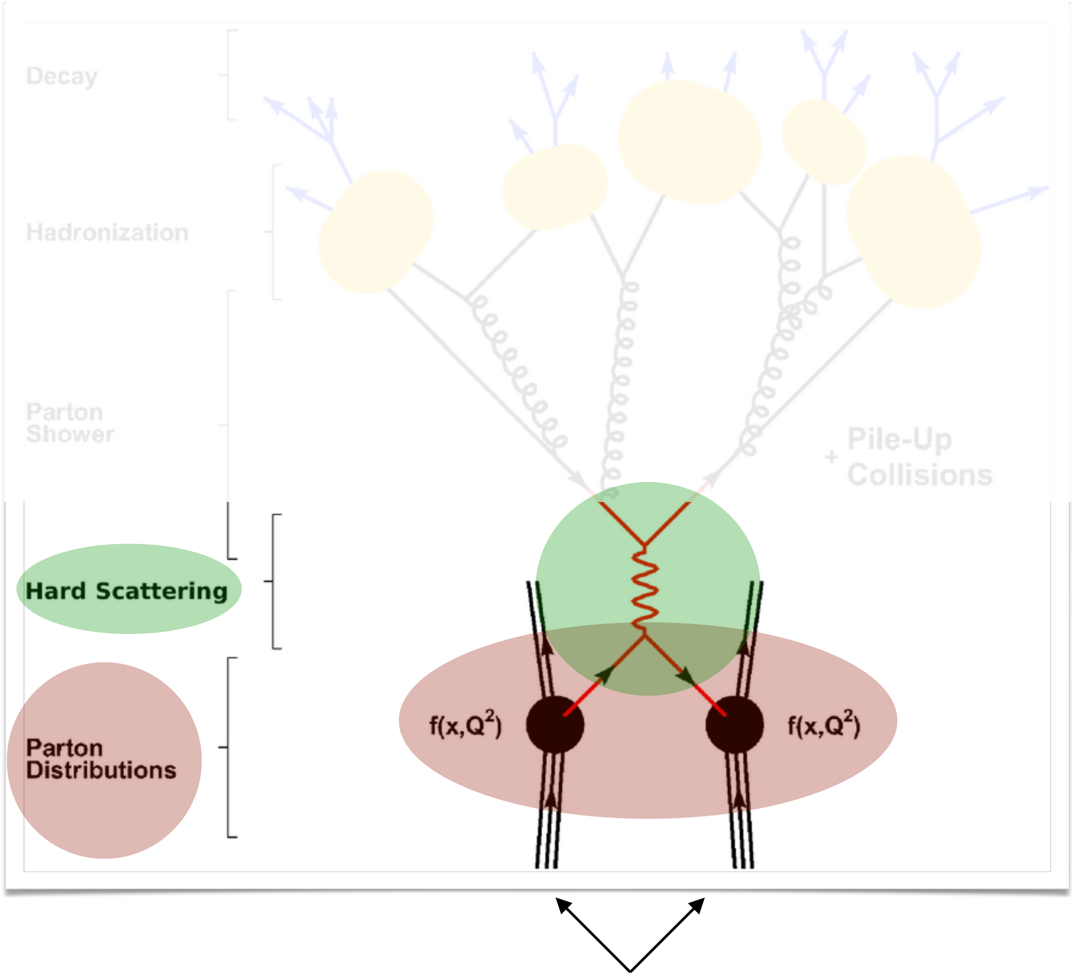
$$\sigma(x,Q^2) = \hat{\sigma}_{ij} \otimes f_i \otimes f_j = \int dz_1 dz_2 \hat{\sigma}(z_1,z_2,Q^2) f_i\left(\frac{x}{z_1},Q^2\right) f_j\left(\frac{x}{z_2},Q^2\right)$$

Partonic (hard) cross sections

- $\sigma(x,Q^2)$ is our **observable**
- Q^2 is the energy scale of the process
- $\hat{\sigma}(z_1,z_2,Q^2)$ can be computed in **perturbation theory**

  NLO, NNLO, ...
- $f_{i/j}(x,Q^2)$ **cannot** be computed in perturbation theory (and they are **universal**)

Non perturbative objects



Initial state = hadrons (protons, neutrons ,...)

PDF extraction.

Let's look at the **Factorization theorem** from another prospective

unknown

$$\sigma(x, Q^2) = \hat{\sigma}_{ij} \otimes f_i \otimes f_j = \int dz_1 dz_2 \hat{\sigma}(z_1, z_2, Q^2) f_i\left(\frac{x}{z_1}, Q^2\right) f_j\left(\frac{x}{z_2}, Q^2\right)$$

Measured in experiments

computed in perturbation theory

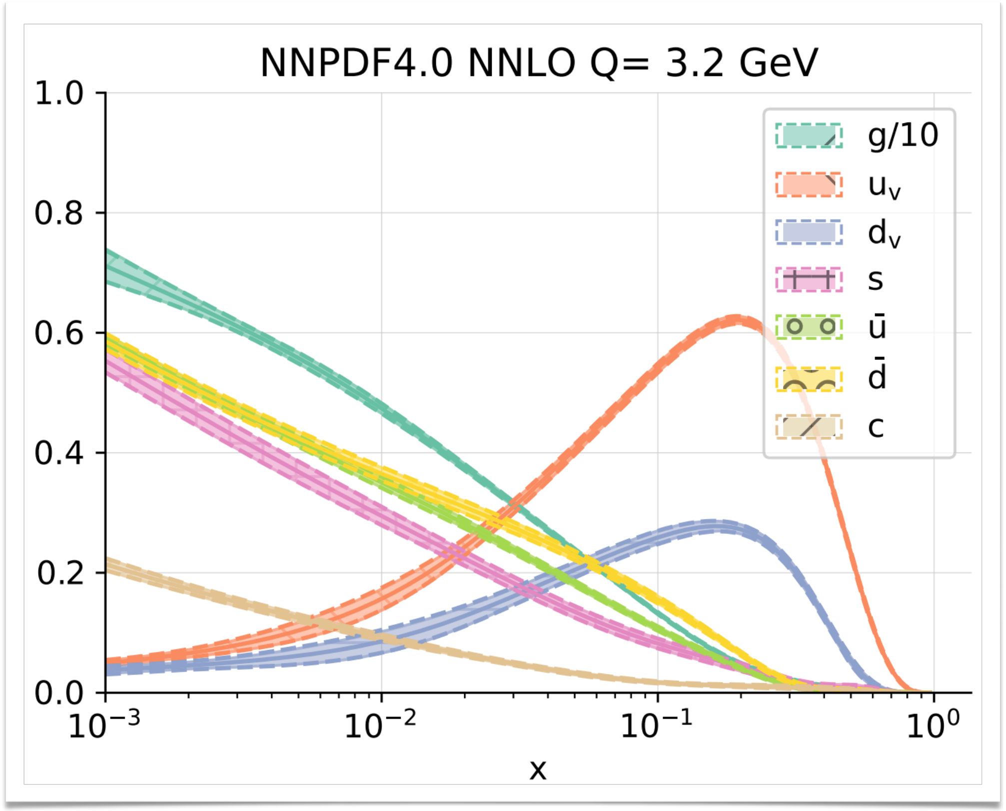
Inverse problem

Also, **DGLAP equations** allow us to compute the PDFs at all scale Q^2, once known at a certain scale Q_0^2

$$f_i(Q^2) = E_{ij}(Q^2 \leftarrow Q_0^2) f_j(Q_0^2)$$

PDFs are then just a set of **unknown functions**

$$f_i : [0,1] -> \mathbb{R}$$



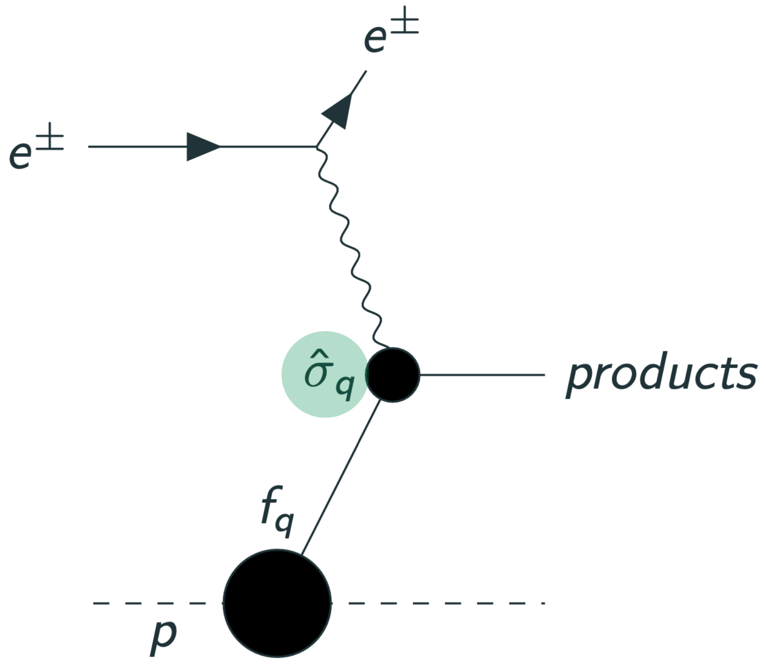
Theory errors.

$$F(Q) = \hat{\sigma}(Q^2) \otimes E_{ij}(Q^2 \leftarrow Q_0^2) \otimes f_j(Q_0^2)$$

→ ***Partonic cross sections*** are computed in perturbation theory

→ ***Anomalous dimensions*** inside DGLAP operator are computed in perturbation theory

Deep Inelastic Scattering (DIS)



$$\hat{\sigma}^{NLO} = \hat{\sigma}^{(0)} + \alpha_s \hat{\sigma}^{(1)} + \mathcal{O}(\alpha_s^2)$$

$$\gamma^{NLO} = \alpha_s \gamma^{(0)} + \alpha_s^2 \gamma^{(1)} + \mathcal{O}(\alpha_s^3)$$

MHOU

(Missing Higher Order Uncertainties)

How can we estimate them?

Theory errors: estimation.

Scale Variations

$$\overline{F}^{NLO}(\mu_f = \kappa_f Q, \mu_r = \kappa_r Q) - F^{NLO}(\mu_f = Q, \mu_r = Q) = \mathcal{O}(NNLO)$$

Factorization scale

Estimates MHOU of anomalous dimensions

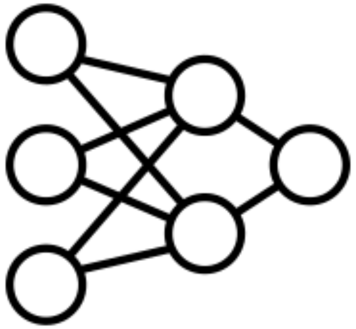
$$E^{NLO}(Q \leftarrow Q_0) \rightarrow \overline{E}^{NLO}(Q \leftarrow Q_0, \kappa_f)$$

Renormalization scale

Estimates MHOU of partonic cross sections

$$\hat{\sigma}^{NLO}(Q) \rightarrow \overline{\hat{\sigma}}(Q, \kappa_r)$$

$\kappa_f, \kappa_r \in (0.5, 2.0)$ is the most common choice



METHODOLOGY AND VALIDATION

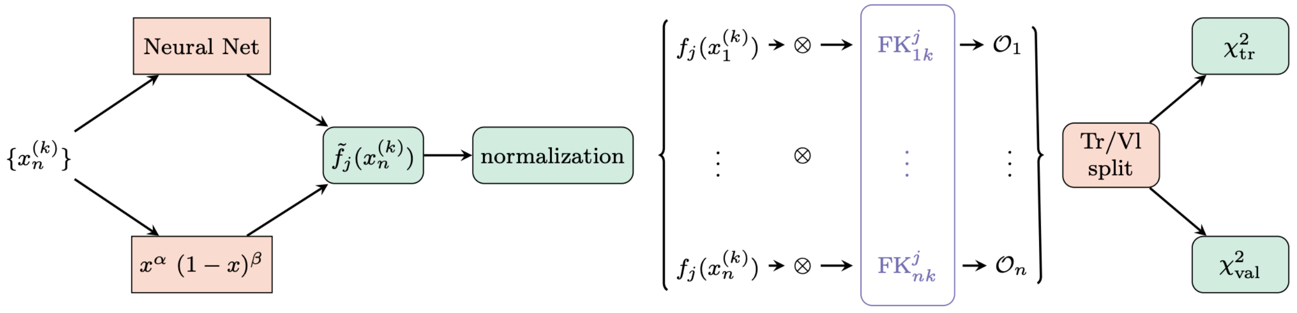
- How does a NNPDF fit work?
- How can we include MHOU in a NNPDF fit?
- Can we validate our estimation?

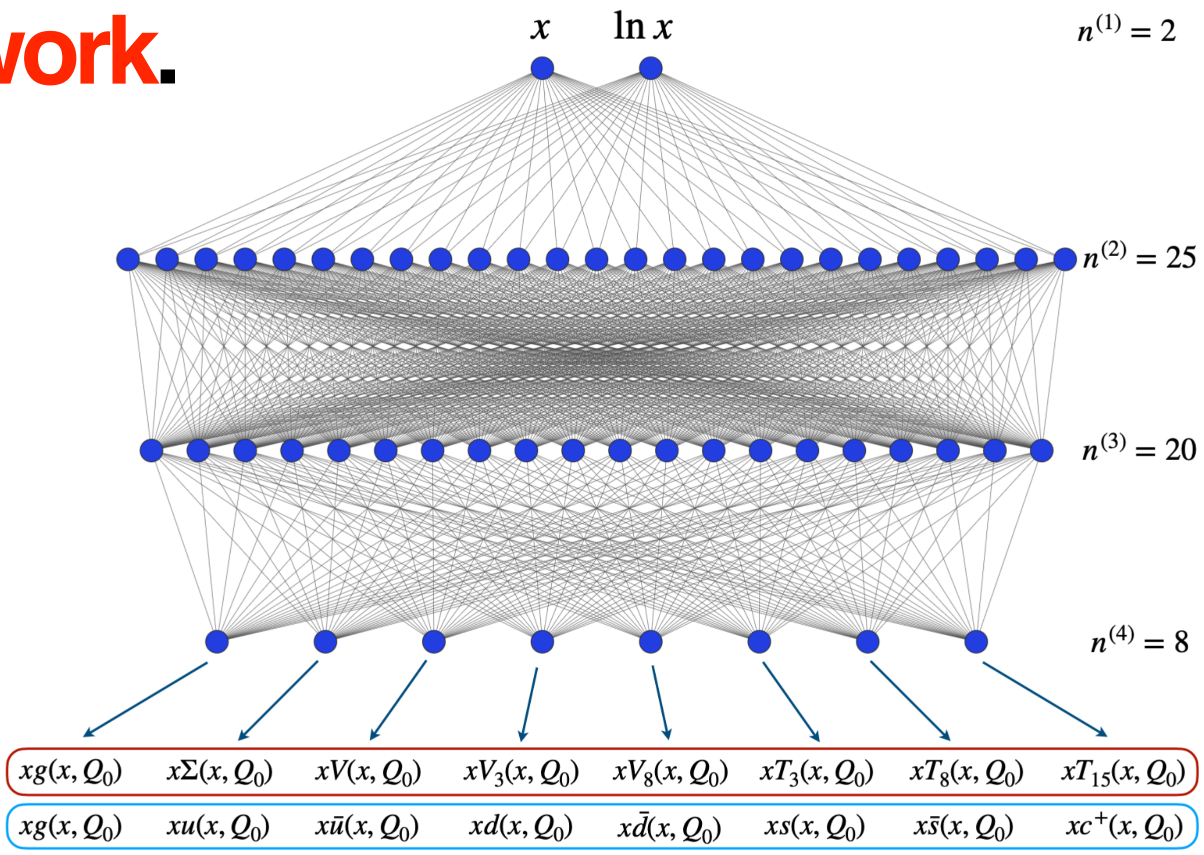
“Truth has nothing to do with the conclusion, and everything to do with the methodology”
(Stefan Molyneux)

Parametrization: the Neural Network.

$$f(x) = A_k x^{-\alpha_k}(1-x)^{\beta_k}\mathbf{NN}(x)$$

Architecture: 2-25-20-8
Activation functions: hyperbolic; linear for the last layer

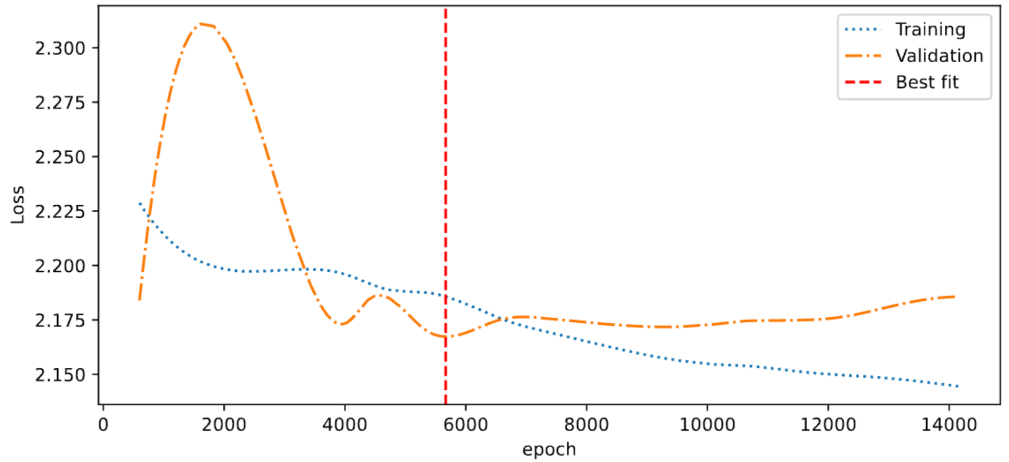




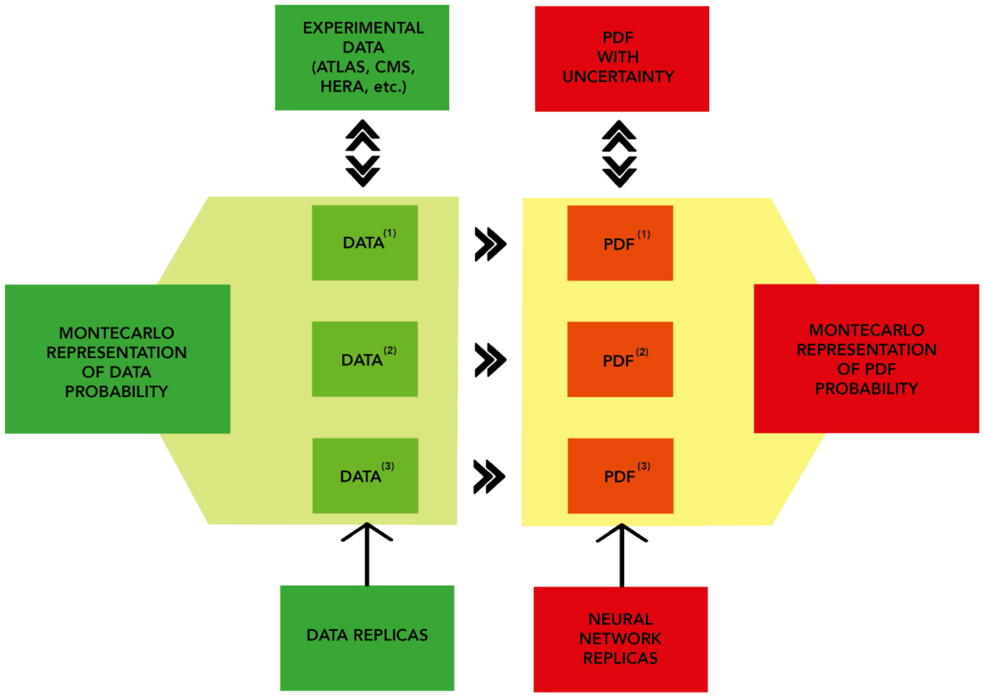
Neural Network: universal interpolator

Training

1. Divide data **D** into **training** set and **validation** set
2. Minimize training χ^2
3. Stop if validation χ^2 no longer improves
4. Take best validation χ^2



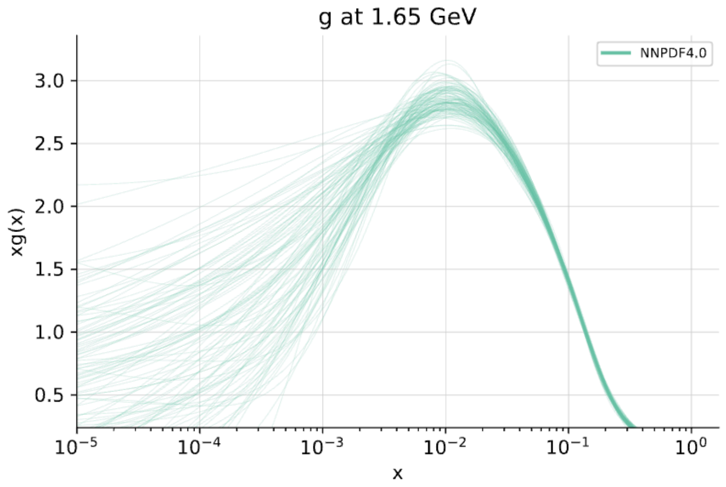
Propagating uncertainties: data to PDF.

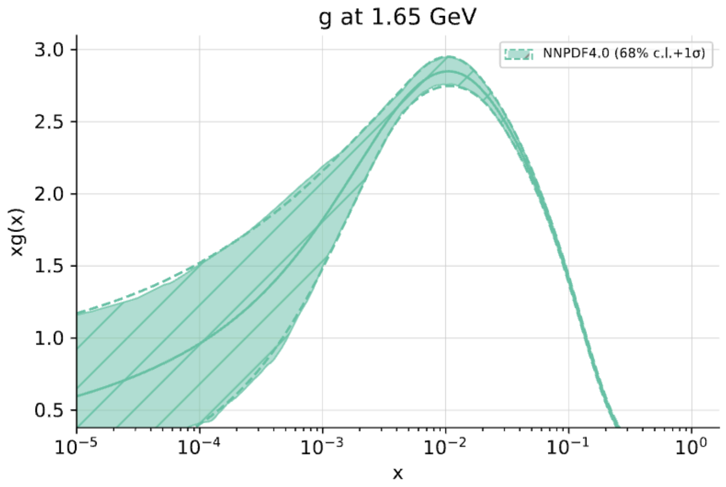


NNPDF adopts a **Monte Carlo** approach

1. Start with the original dataset **D** and its **covariance matrix C**
2. Generate N_{rep} **pseudodata** D_i according to C
3. Fit a **Neural Network** $\mathbf{NN}_i$ to each of the pseudodata replica
4. Deliver the full set of replicas

PDFs uncertainties are given by the distribution of the Monte Carlo set





NB: Another possibility is the Hessian approach. The two methods can be converted one in the other (hep-ph:1505.06736)

MHOU in a PDF fit: the *theory covmat*.

How to use it

→ Experimental and theoretical uncertainties enter in a symmetric way in the figure of merit used for PDF determination.

→ The theory covariance matrix S describes theoretical uncertainties and correlations.

→ Include it both in figure of merit and in pseudodata generation.

FIT WITHOUT THEORY ERRORS		**FIT WITH THEORY ERRORS**
$\chi^2 \propto (D_i - T_i)C_{ij}^{-1}(D_j - T_j)$	$\longrightarrow$	$\chi^2 \propto (D_i - T_i)(C+S)_{ij}^{-1}(D_j - T_j)$
Pseudodata replica $\propto C$	$\longrightarrow$	*Pseudodata replica* $\propto C + S$

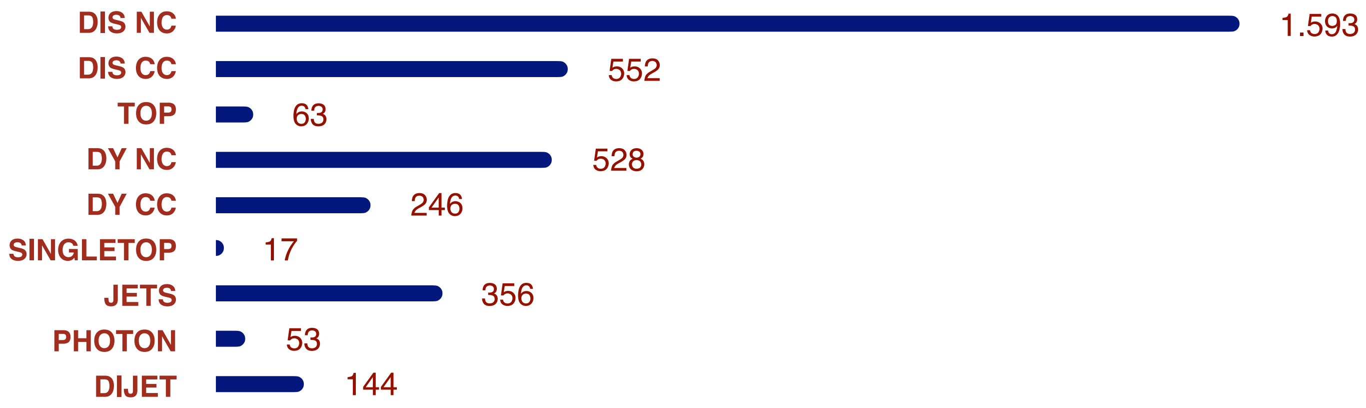
MHOU in a PDF fit: the *theory covmat*.

How to construct it

$$S_{ij} = n_m \sum_{V_m} \left(\overline{F}(\kappa_f, \kappa_{r_a}) - F \right)_{i_a} \left(\overline{F}(\kappa_f, \kappa_{r_b}) - F \right)_{j_b}$$

→ *Factorization scale* **correlates** all the points

→ *Renormalization scale* **correlates** points belonging to the same process



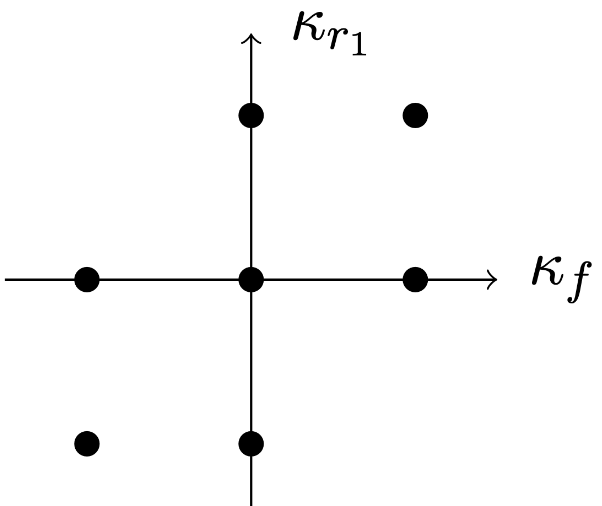
More on the construction: point prescriptions.

Depending on how many (κ_f, κ_r) points among the 9 possible points, one has a different **point prescription**

$$\Delta_{i_a}^{(\pm,0);(\pm,0)} = (\overline{F}(\kappa_f, \kappa_{r_a}) - F)_{i_a} \quad \rightarrow \quad + \to \kappa_{f,r} = 2.0 \quad - \to \kappa_{f,r} = 0.5 \quad 0 \to \kappa_{f,r} = 1.0$$

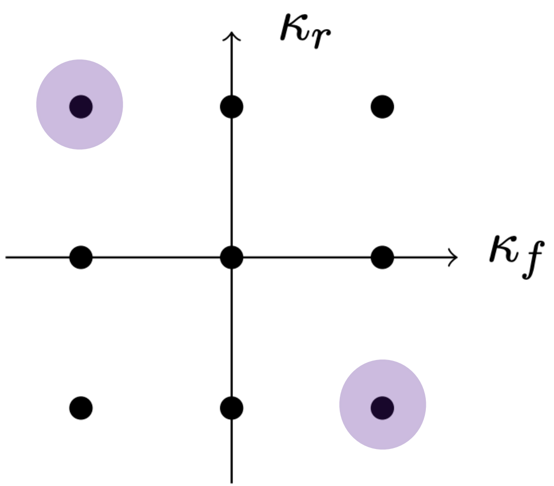
7 points

$$S_{i,j} = \frac{1}{3}\Big[\Delta_i^{+0}\Delta_j^{+0} + \Delta_i^{-0}\Delta_j^{-0} + \Delta_i^{0+}\Delta_j^{0+} + \Delta_i^{0-}\Delta_j^{0-} + \Delta_i^{++}\Delta_j^{++} + \Delta_i^{--}\Delta_j^{--}\Big]$$



9 points

$$S_{i_1,j_2} = \frac{1}{4}\Big[\Delta_i^{+0}\Delta_j^{+0} + \Delta_i^{-0}\Delta_j^{-0} + \Delta_i^{0+}\Delta_j^{0+} + \Delta_i^{0-}\Delta_j^{0-} + \Delta_i^{++}\Delta_j^{++} + \Delta_i^{+-}\Delta_j^{+-} + \Delta_i^{-+}\Delta_j^{-+} + \Delta_i^{--}\Delta_j^{--}\Big]$$



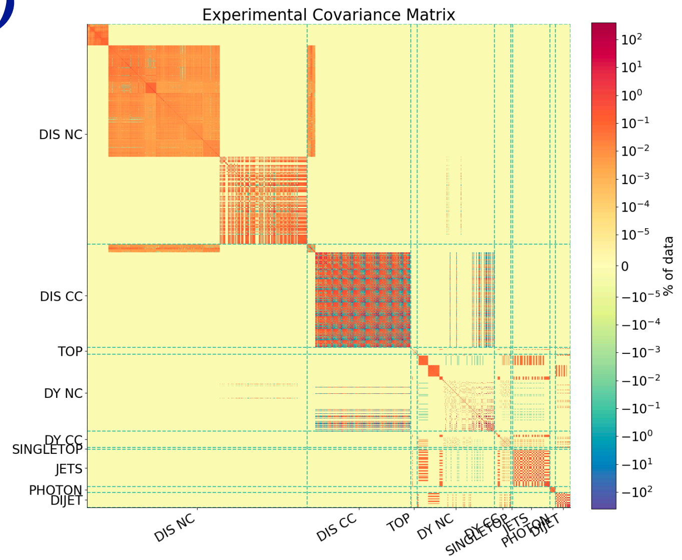
How do they look like?

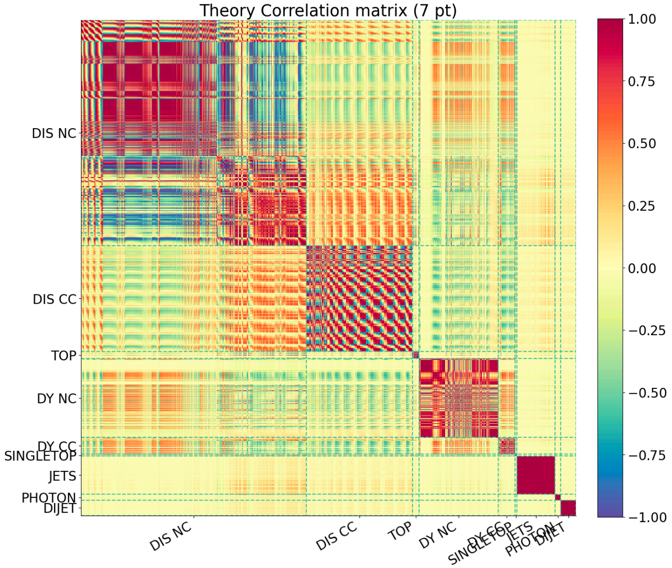
C

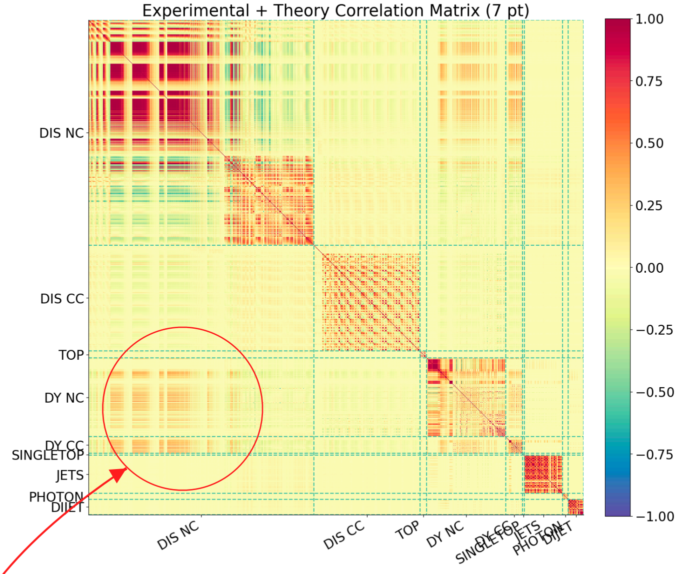
S

$C + S$

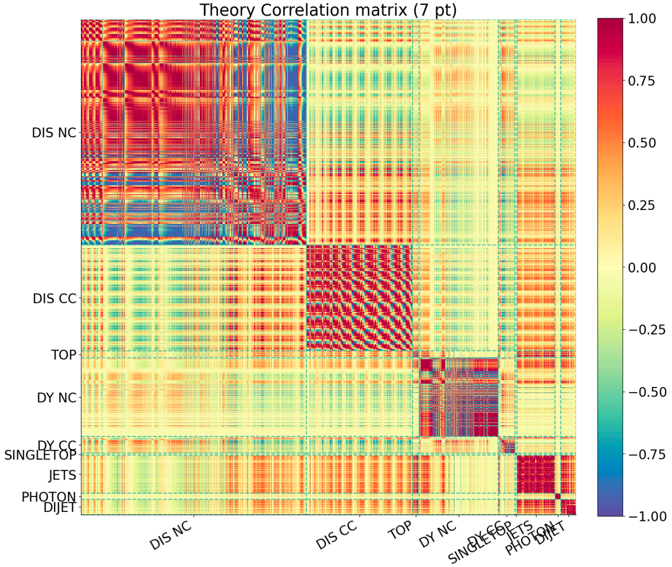
→ NLO

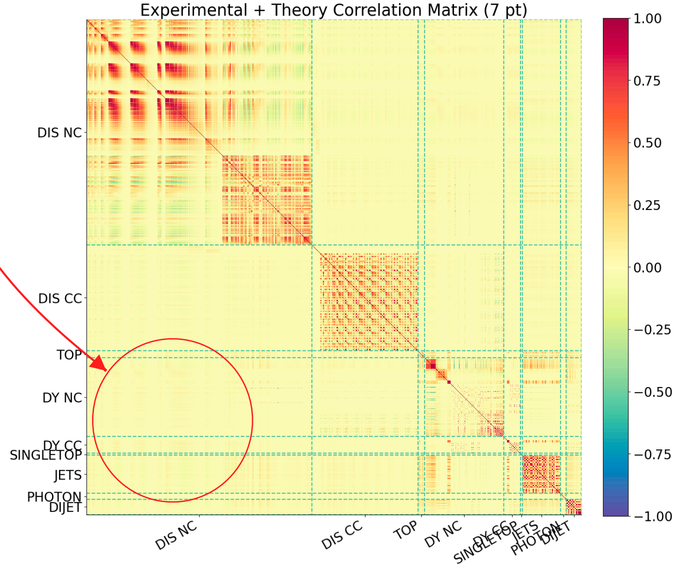




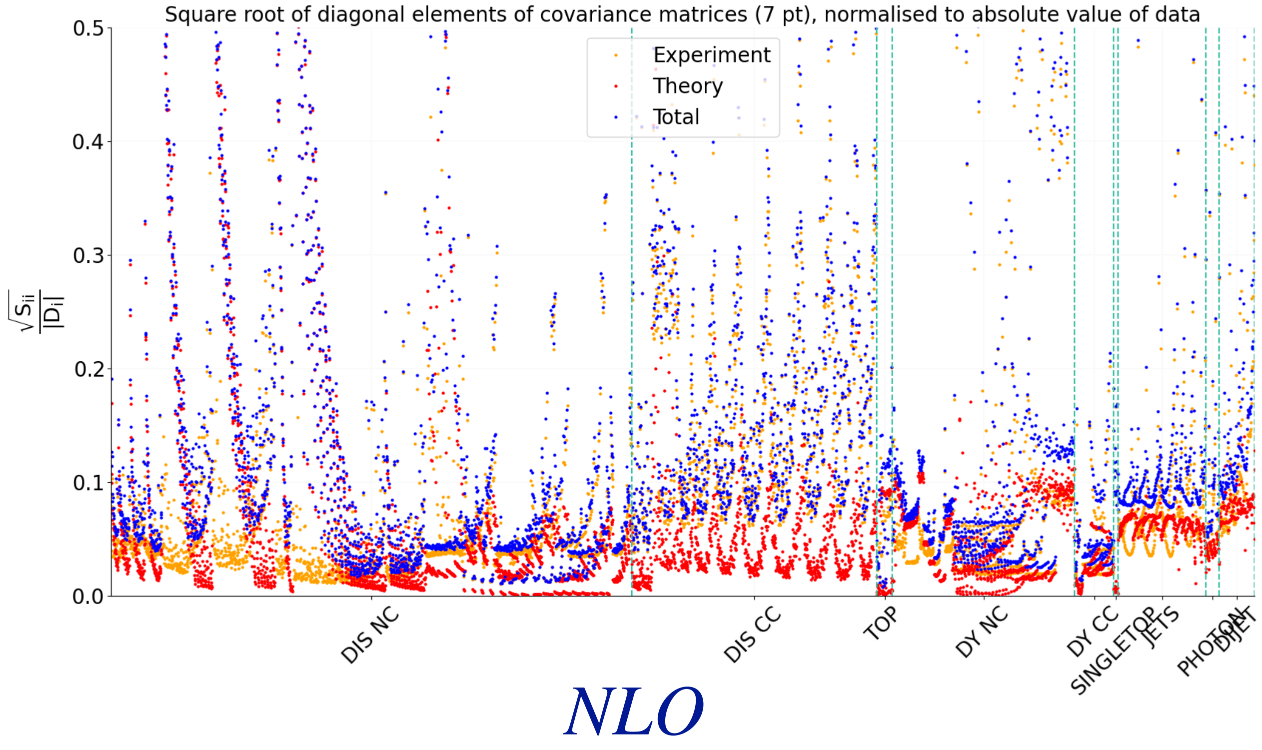


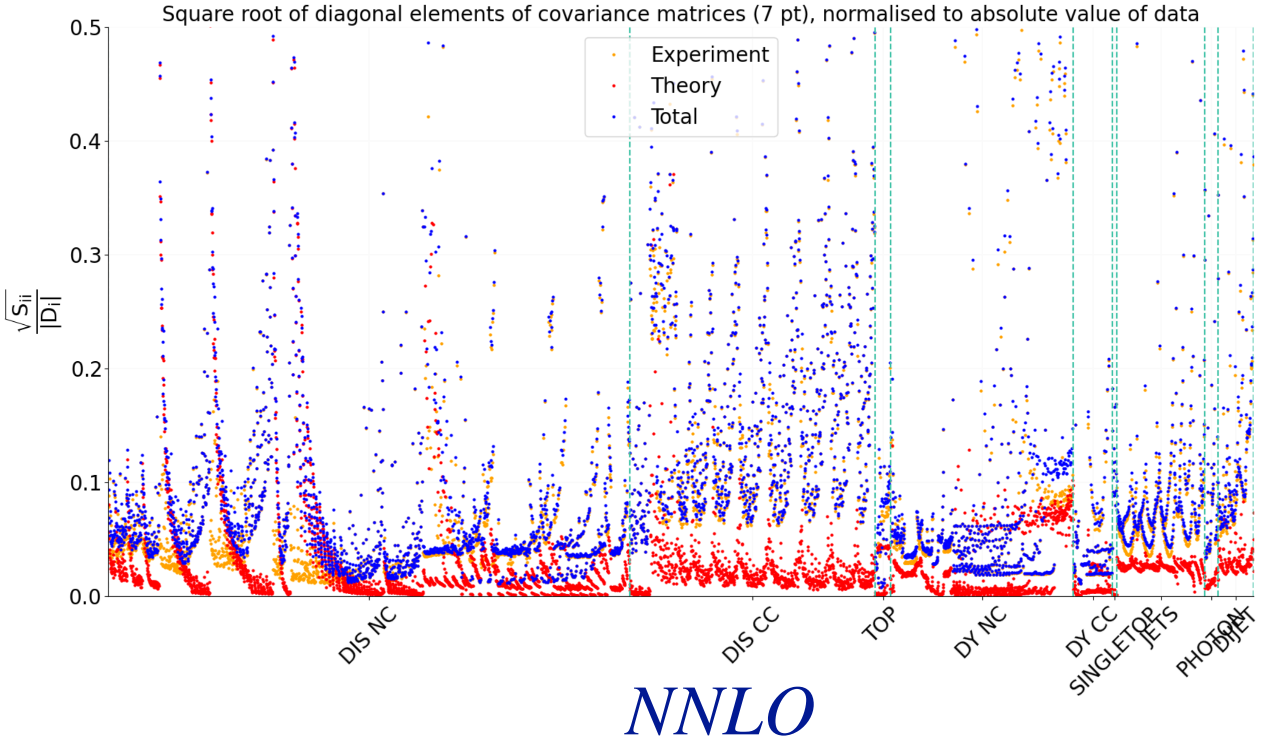
→ $NNLO$





Diagonal elements.



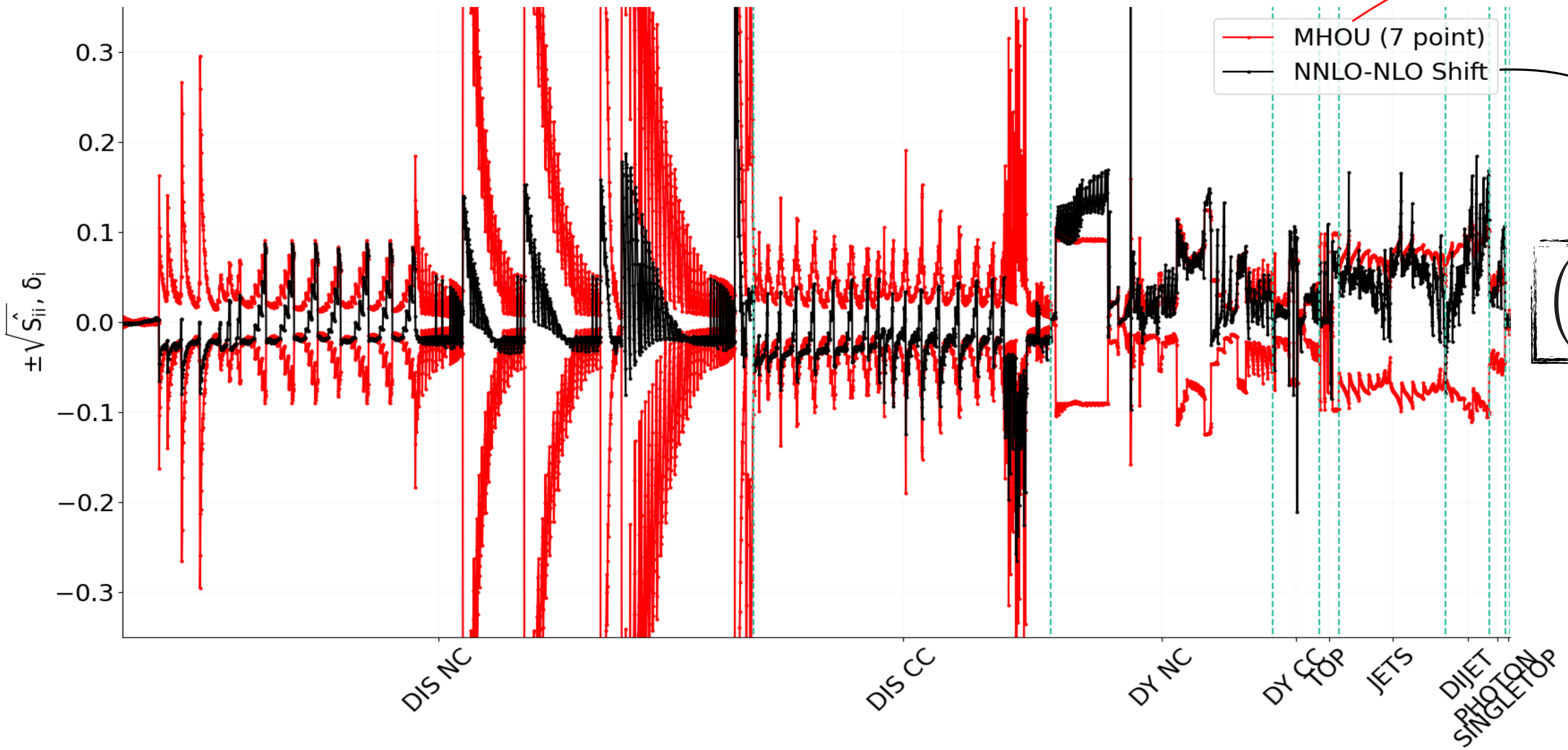


At **NNLO** theory errors are clearly subdominant, while at **NLO** they are of the same size of experimental errors

Validation: is it reproducing the known NNLO?

Most of the predictions are currently known up to $\mathcal{O}(NNLO)$:
we can test the NLO MHOU !

$$\left(\frac{\sqrt{S_{ii}^{NLO}}}{F_i^{NLO}}\right) \times 100$$



$$\left(\frac{F_i^{NNLO} - F_i^{NLO}}{F_i^{NLO}}\right) \times 100$$

RESULTS

- Does the fit quality improve upon inclusion of theory errors?
- What is the impact on the PDFs at NLO and NNLO?
- What about N3LO?

“We're always, by the way, in fundamental physics, always trying to investigate those things in which we don't understand the conclusions. After we've checked them enough, we're okay”
(Richard P. Feynman)

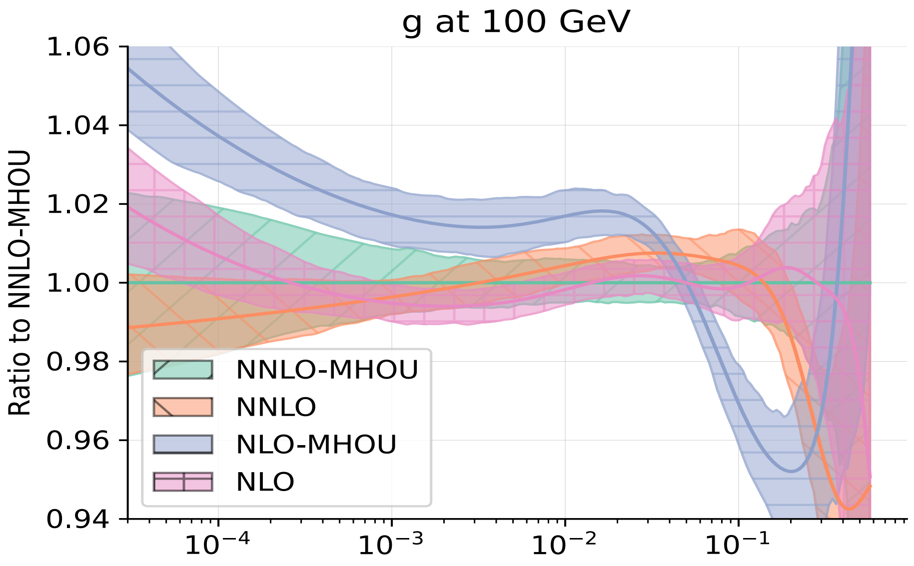
Fit quality.

Dataset χ^2	N_{dat}	NLO $C + S^{(nucl)}$	NLO $C + S^{(nucl)} + S^{(7pt)}$	NNLO $C + S^{(nucl)}$	NNLO $C + S^{(nucl)} + S^{(7pt)}$
DIS NC	2100	1.30	1.22	1.23	1.20
DIS CC	989	0.92	0.87	0.90	0.90
DY NC	736	2.01	1.71	1.20	1.15
DY CC	157	1.48	1.42	1.48	1.37
Top pairs	64	2.08	1.24	1.21	1.43
Single-inclusive jets	356	0.84	0.82	0.96	0.81
Dijets	144	1.52	1.84	2.04	1.71
Prompt photons	53	0.59	0.49	0.75	0.67
Single top	17	0.36	0.35	0.36	0.38
Total	4616	1.34	1.23	1.17	1.13

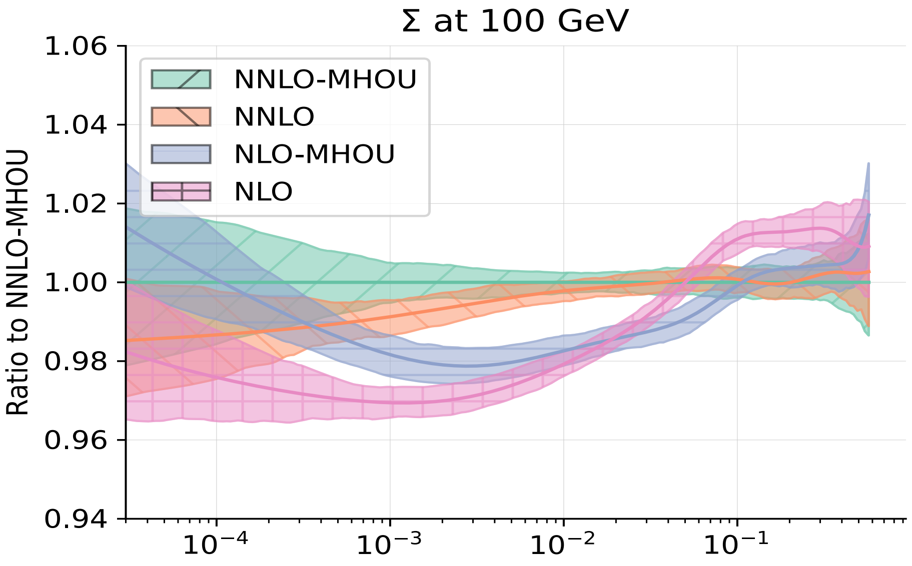
- The total χ^2 **decreases** upon inclusion of MHOU for both NLO and NNLO
- For most of the **process groups** the NLO theory covariance matrix correctly accounts for the missing NNLO terms

PDF comparison.

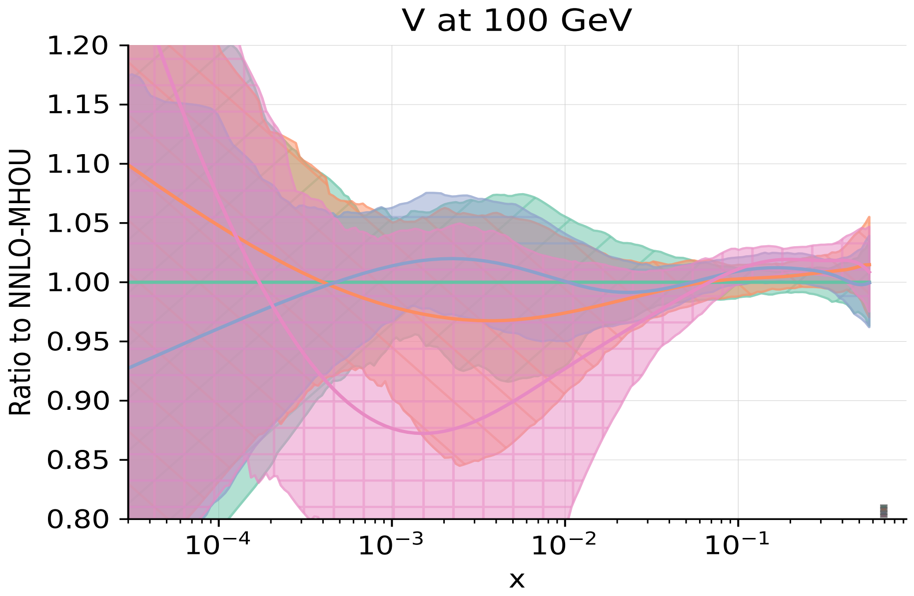
g



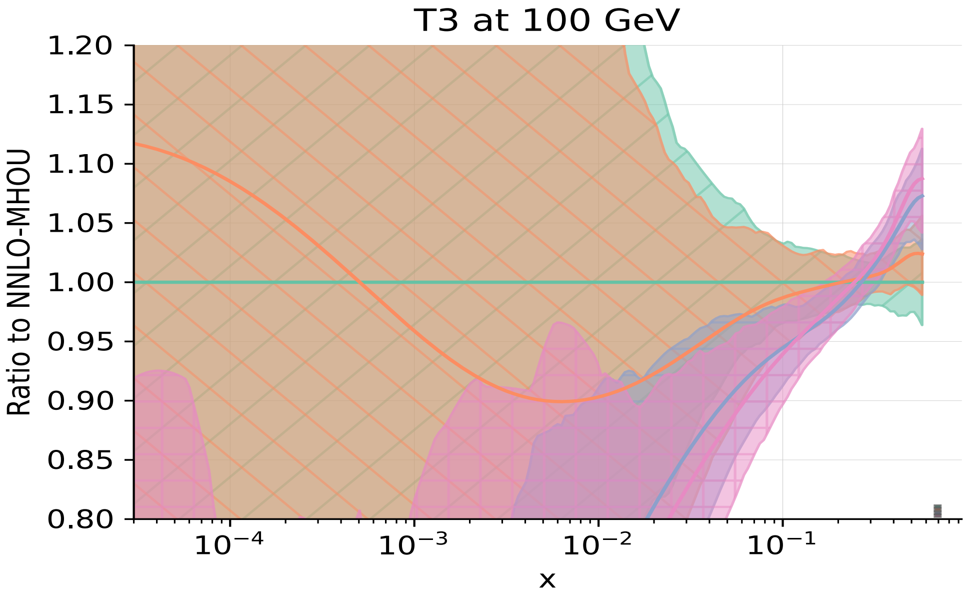
Σ



V

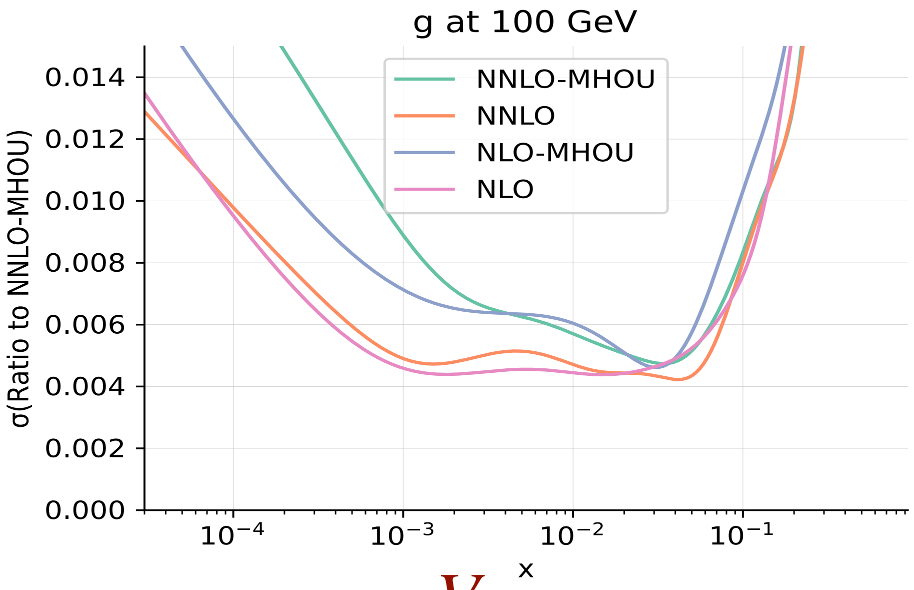


$T3$

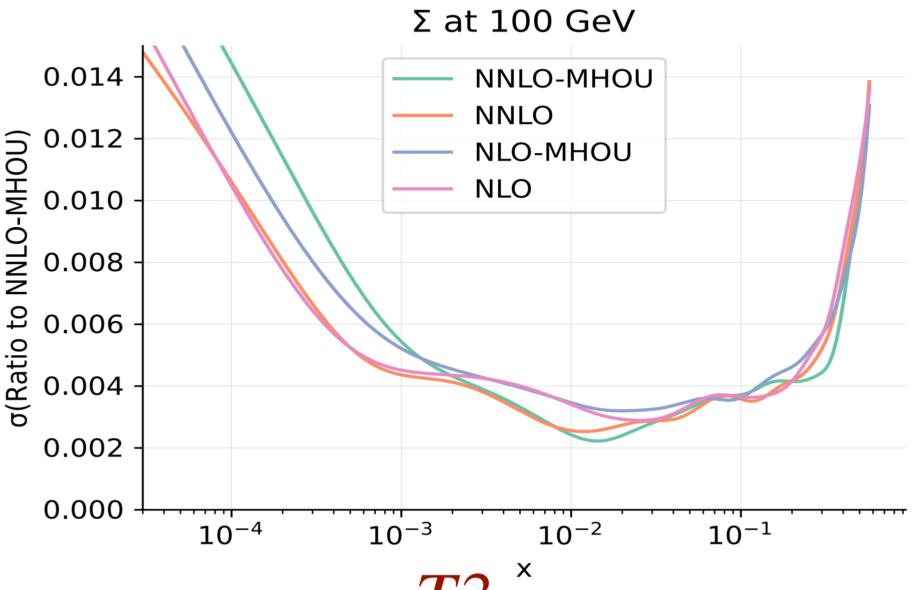


PDF uncertainties.

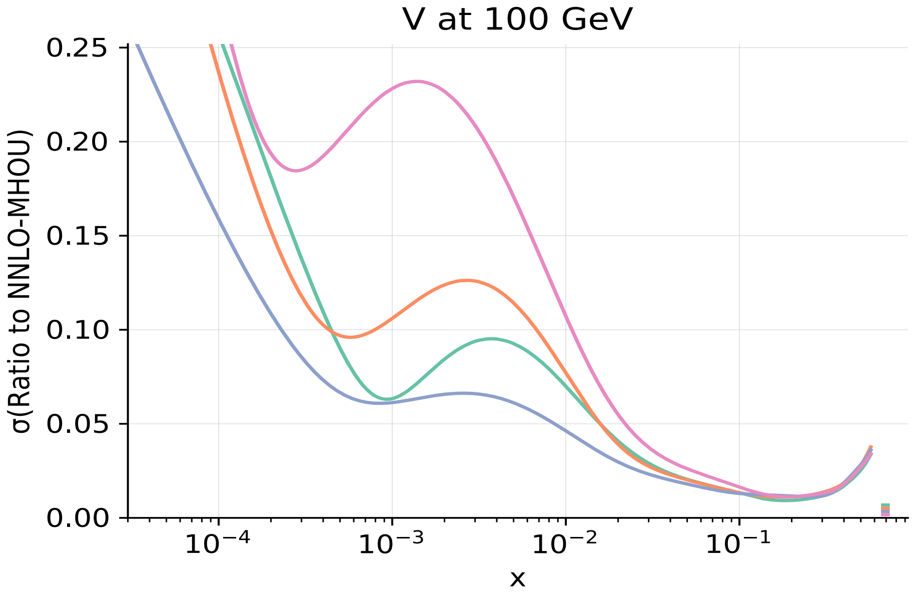
g



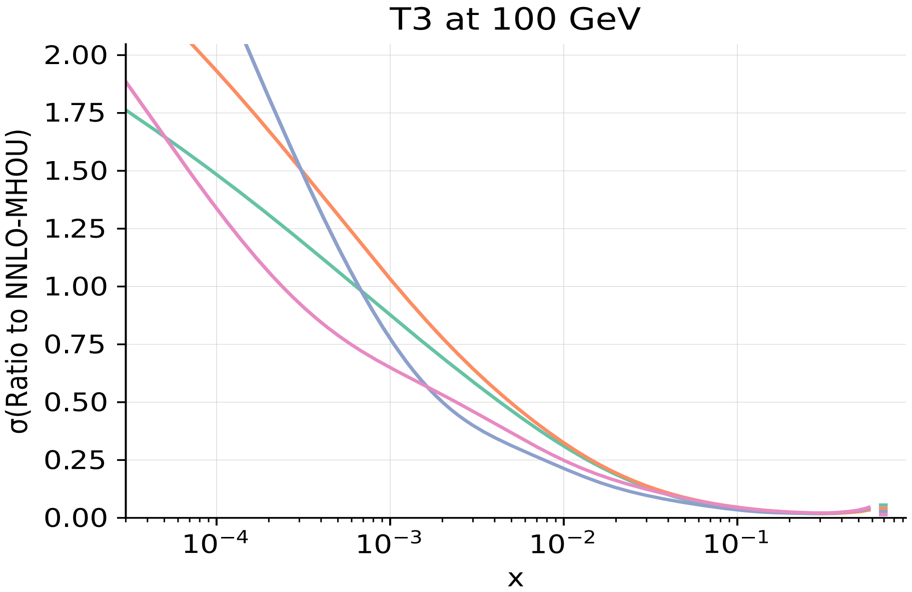
Σ



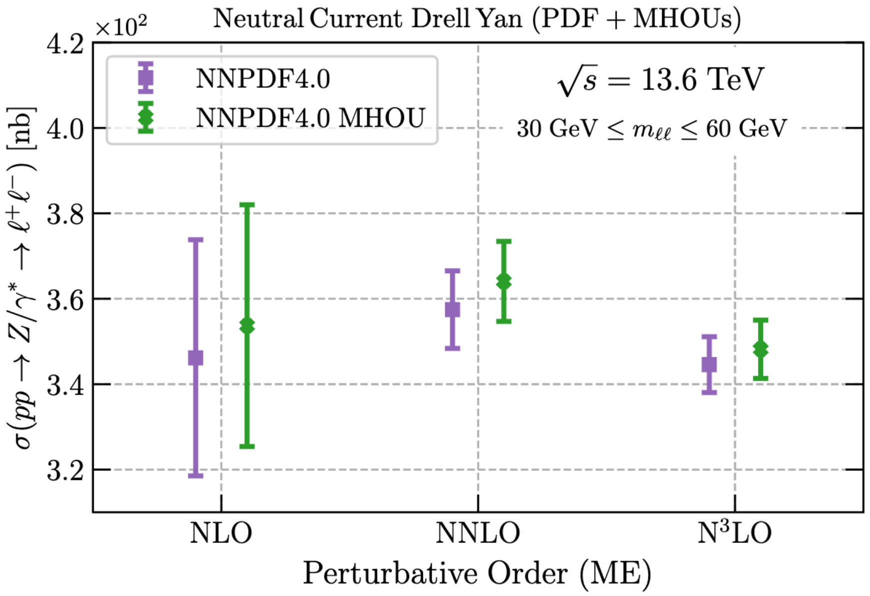
V

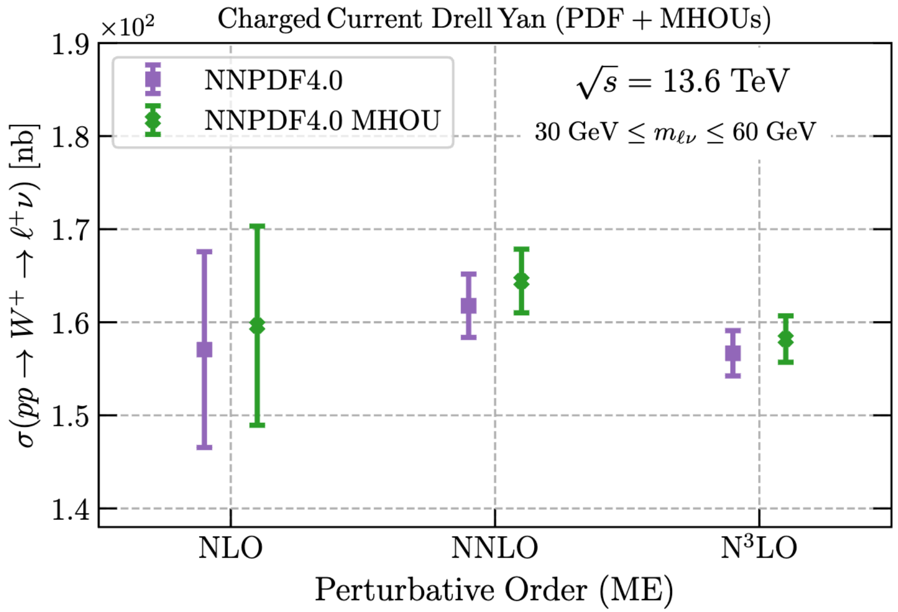


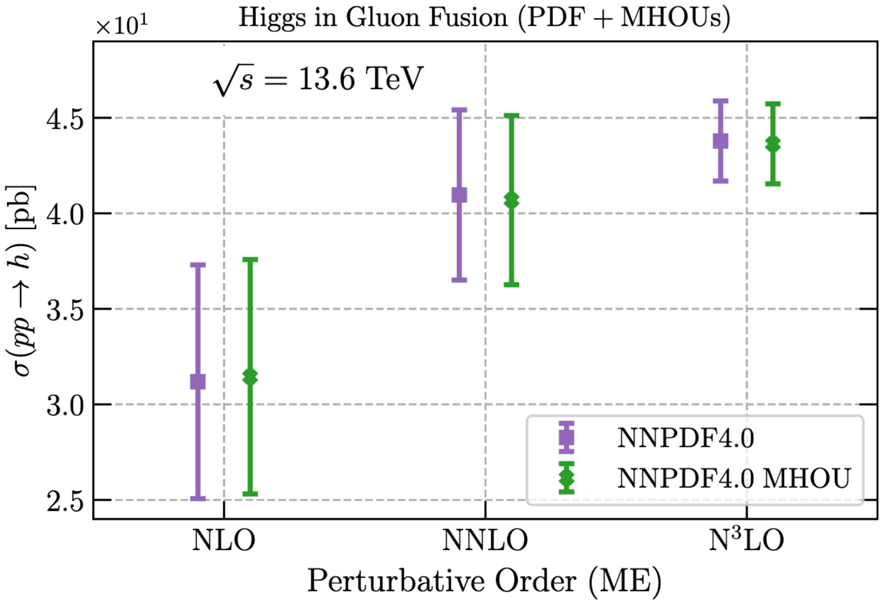
$T3$

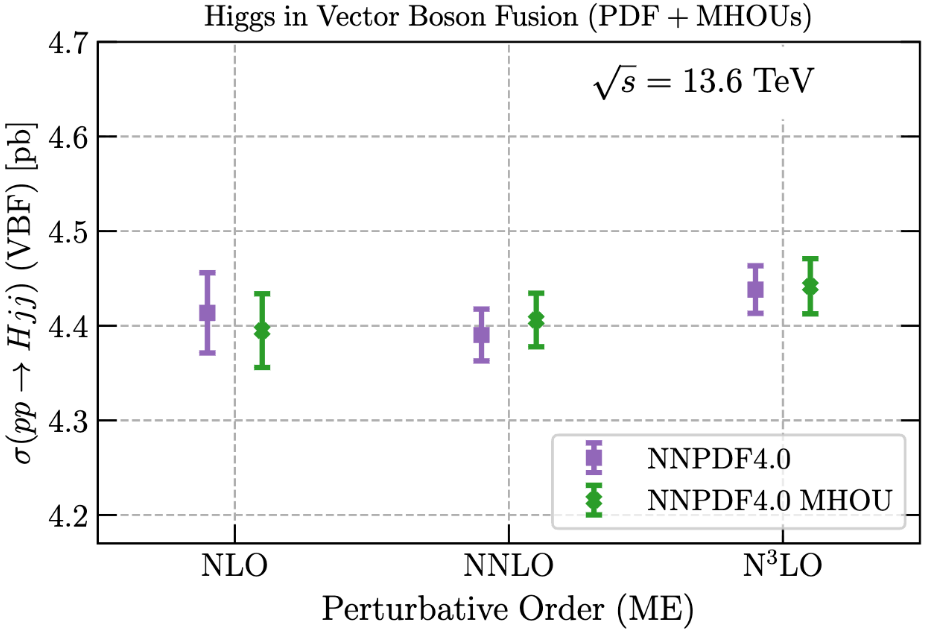


Perturbative convergence.









Conclusions and outlooks.

- Thanks to *scale variations* it is possible to estimate MHOU while, thanks to the theory covmat formalism, it is possible to include such estimation in a PDF fit.

- Including MHOU in a PDF fit is necessary to have faithful uncertainties and central values.

- The perturbative convergence from NLO to N3LO improves once theory errors are accounted for.

Thanks for your attention!

BACKUP

Asymptotic freedom.

In **QCD** we are usually expand quantities in terms of the **strong coupling**

$$\alpha_s(Q^2)$$

(Notable counterexample is lattice QCD)

$$\hat{\sigma}^{NLO}(z_1, z_2, Q^2) = \hat{\sigma}^{(0)}(z_1, z_2, Q^2) + \alpha_s(Q^2)\hat{\sigma}^{(1)}(z_1, z_2, Q^2) + \mathcal{O}(\alpha_s^2)$$

($NLO =$ Next-to-leading order)

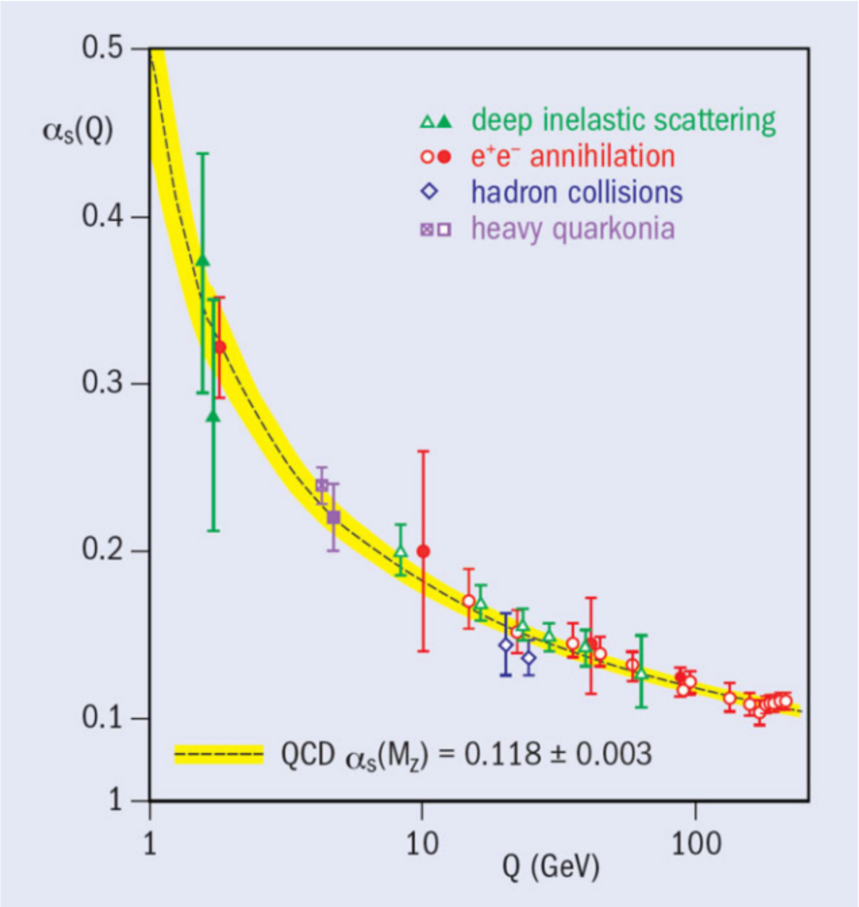
But $\alpha_s(Q^2)$ is a decreasing function of the energy scale

perturbative QCD
(pQCD)
from ~ 1 GeV

Non perturbative QCD
below ~ 1 GeV

Partonic cross sections

PDFs
(Mass of the proton ~ 0.938 GeV)



Physik Journal 3 12 31

How can we extract them?

Inverse problems.

Number of datapoints is **finite** while function space is **infinite-dimensional**

Fitting PDFs is always an **under-determined** problem

ASSUMPTIONS

Fixed parametrization

→ Reduce the number of parameters

→ Assumptions = choice of the parameters to be fitted

Neural Network

NNPDF choice

→ Applies a **regularization**

→ Assumptions = encoded in the network (and not only…)

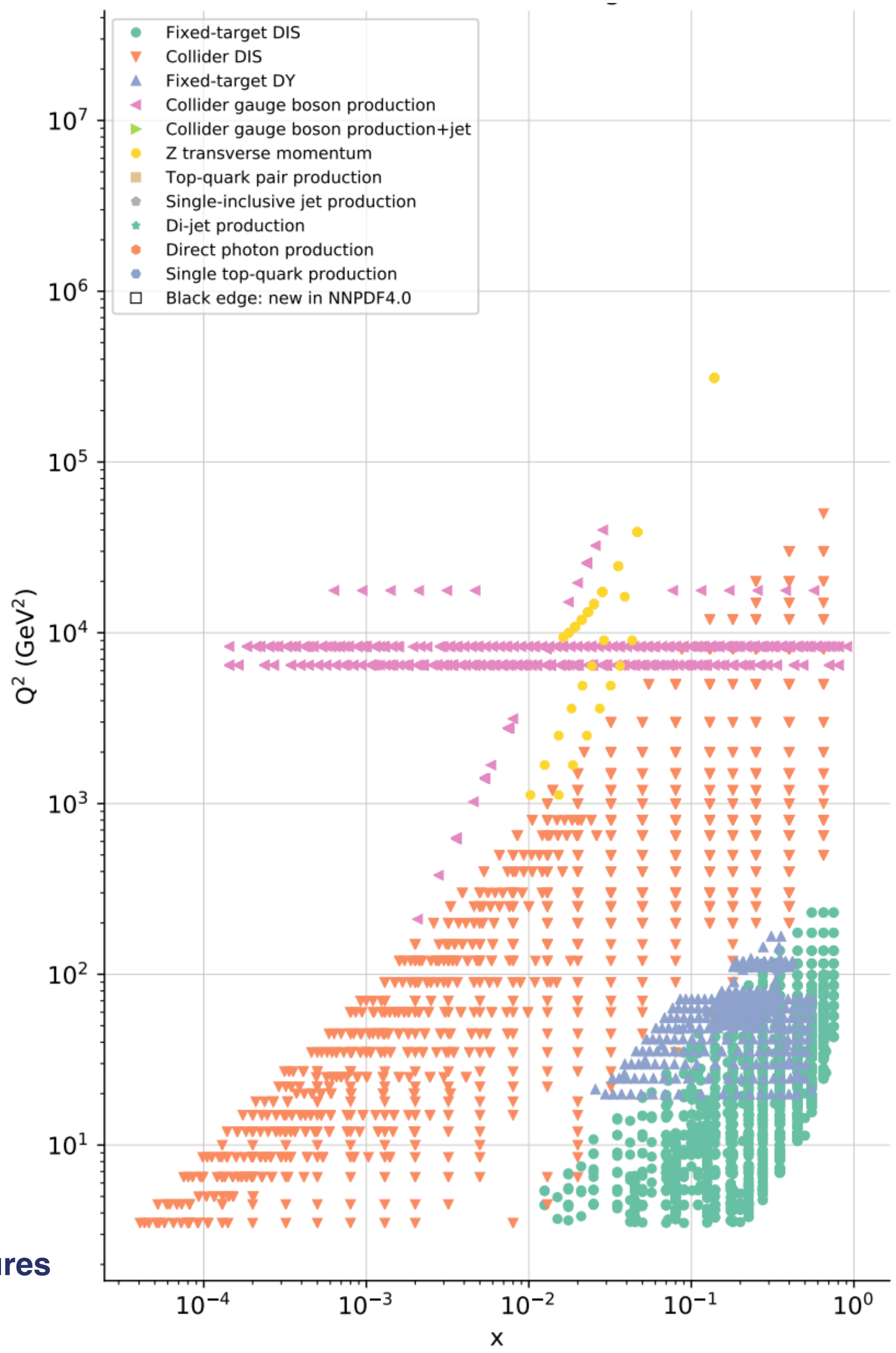
Which is better?

→ Needs theoretical insight on PDFs shape

→ Can be biased by human prejudice

→ Needs theoretical insight on more **abstract features**

→ Human prejudice effect can be minimized



~4600 datapoints in NNPDF4.0

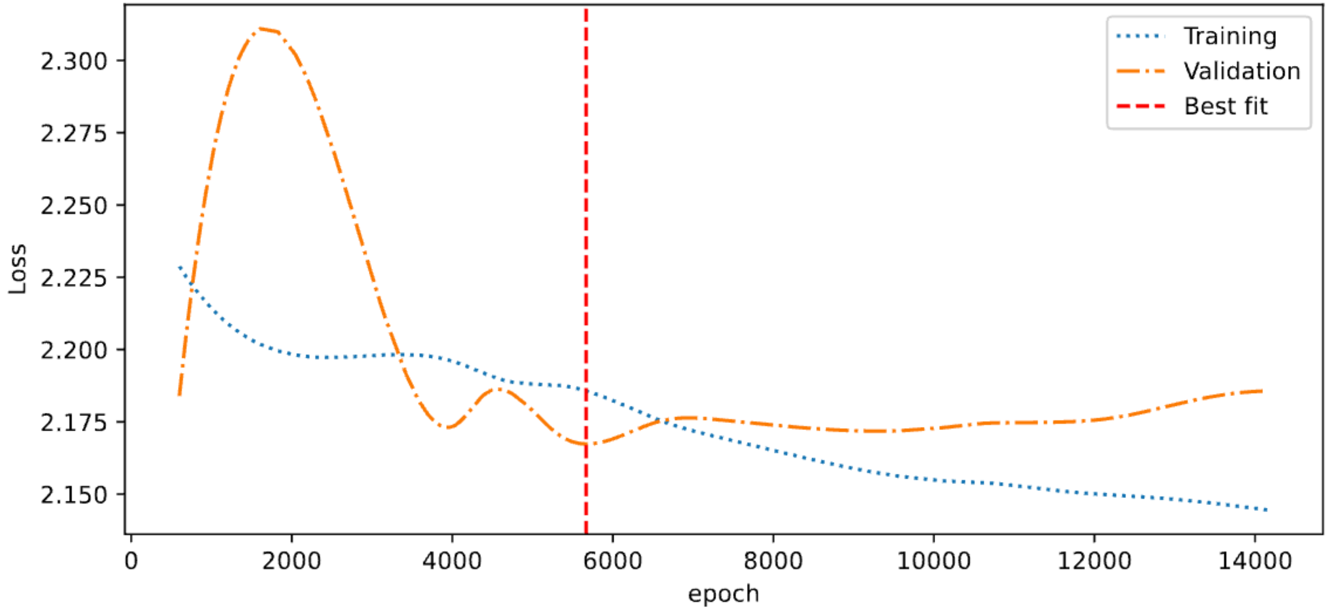
Training the neural network.

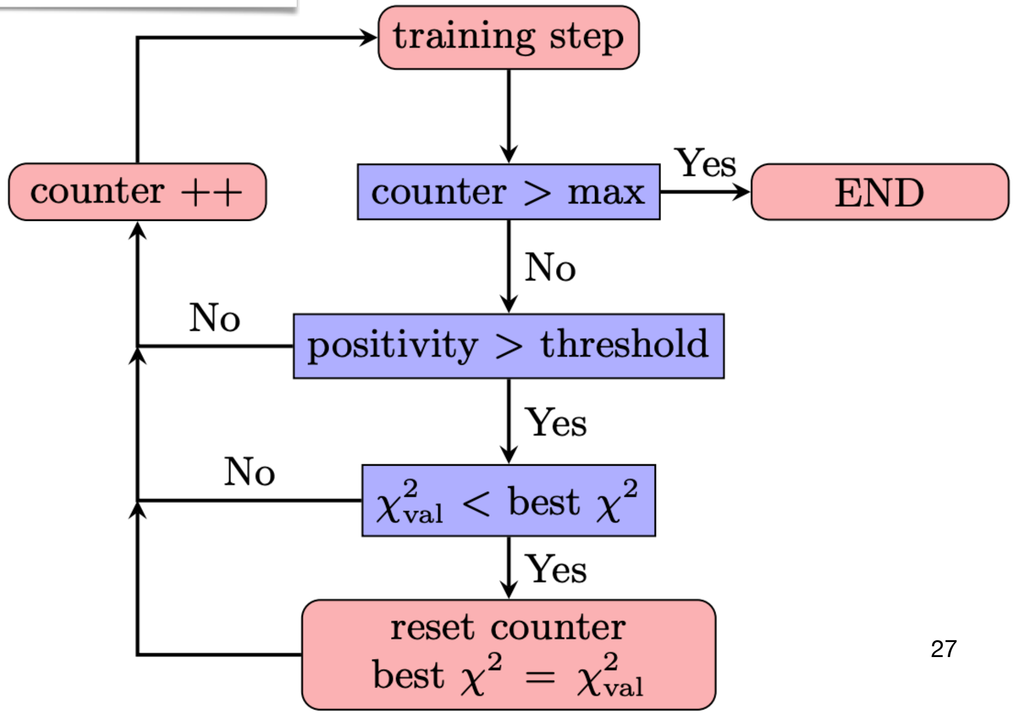
Avoid fitting the noise (**overfitting)**

Cross-validation

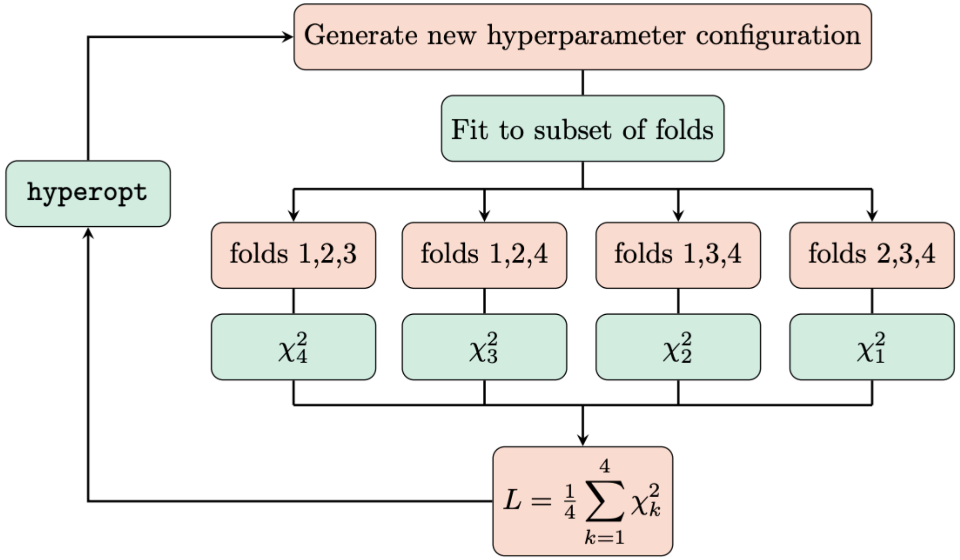
Stopping

1. Divide data **D** into **training** set and **validation** set
2. Minimize training χ^2
3. Stop if validation χ^2 no longer improves
4. Take best validation χ^2





Automated model selection



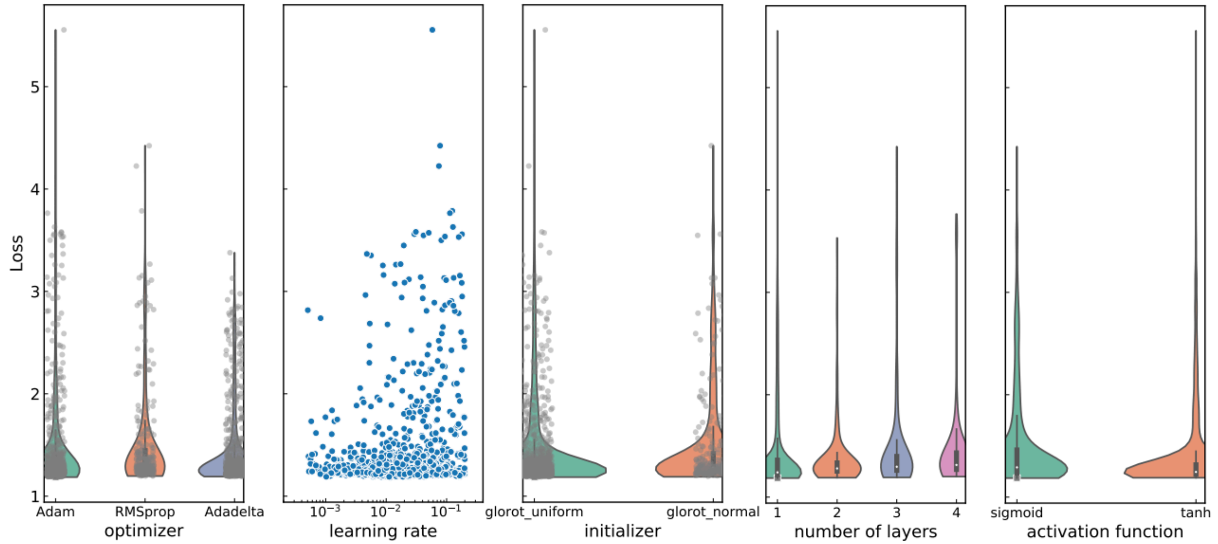
Minimize sources of **bias** in the PDFs:

- Functional form $\longrightarrow$ Neural Network
- Model parameters $\longrightarrow$ **Hyperoptimization**

Idea is to scan over a large enough hyperparameter space and select the best set

Best $\longrightarrow$ best χ^2 on a **test dataset** (never seen by the NN)

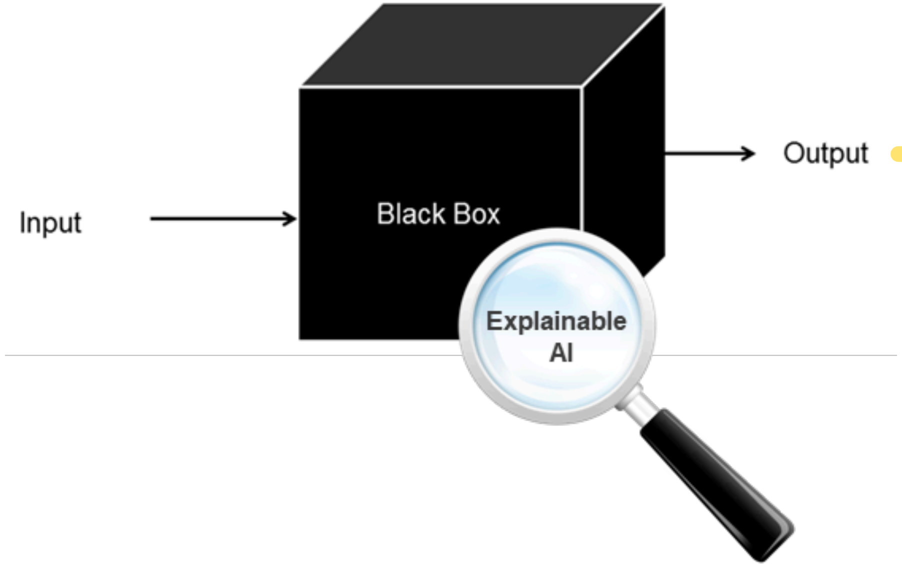
NB: Still requires some human input (more on this later)



Can we trust our results?

Downside of Neural Networks:
we lack a **full analytical insight** on the process

NN is often considered to be a **black box**



Tests a priori (WIP)

- Test internal features of the NN
- “Analytical” approach

Tests a posteriori

- Test properties of the results
- Empirical approach

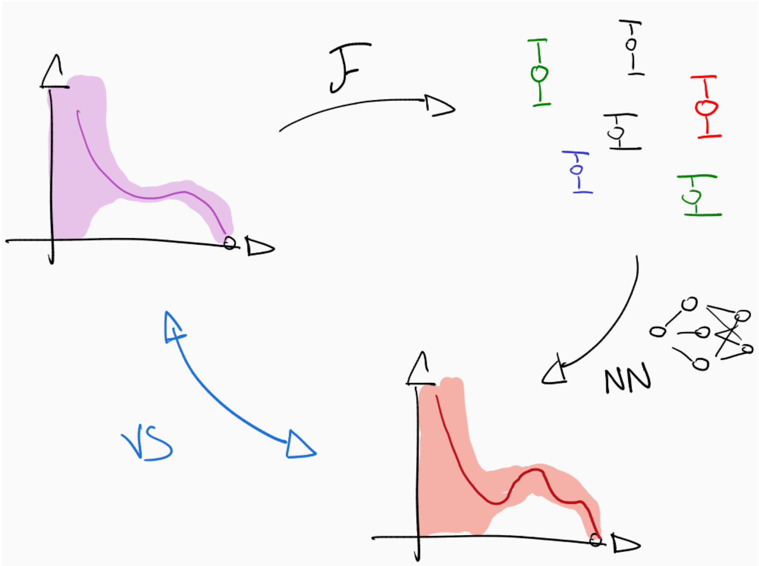
Focus on these!

Closure and future tests

Closure test

Test the algorithm in a controlled environment where the **“truth” is known**

1. Choose a PDF as underlying truth
2. Generate central fake data (**LEVEL 0**)
3. Generate smeared fake data with the experimental covariance matrix (**LEVEL 1**)
4. Generate and fit pseudodata replica (**LEVEL 2**)
5. Compare the results with known distribution

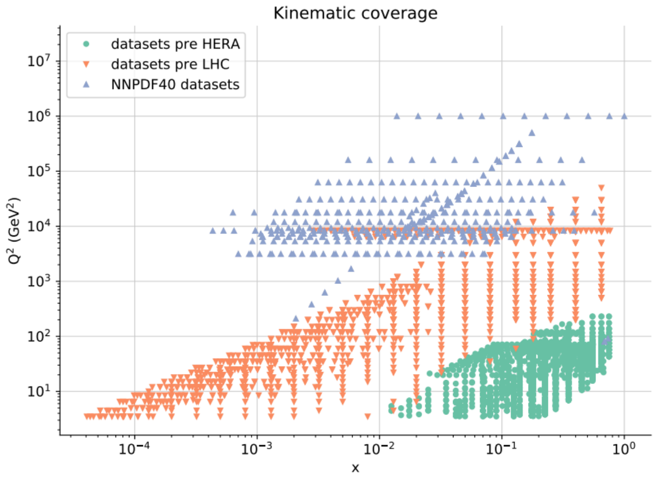


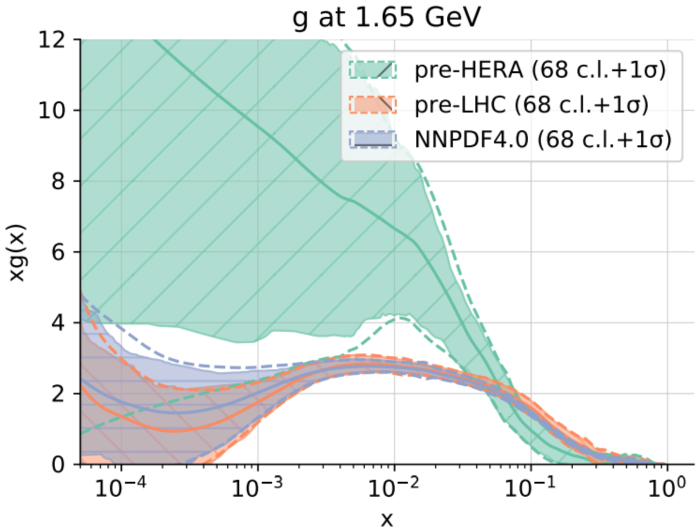
Future test

What about data you have not seen yet?

Traveling in time is not possible but I know **history!**

Divide the dataset **chronologically** and perform a fit for each set: **yesterday’s extrapolation region is today’s data region**



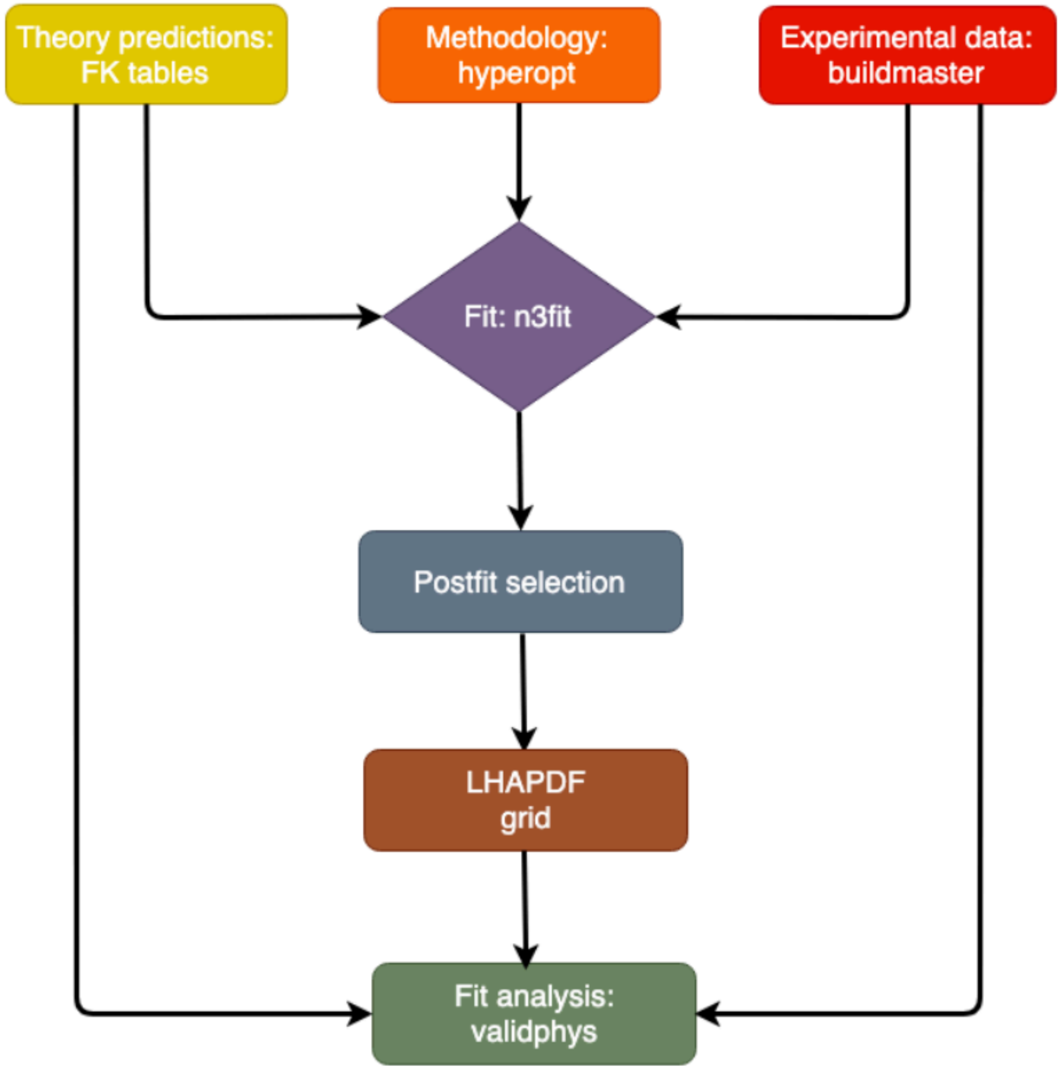


The NNPDF code is open-source

The full NNPDF code has been made **public** along with **user friendly documentation**

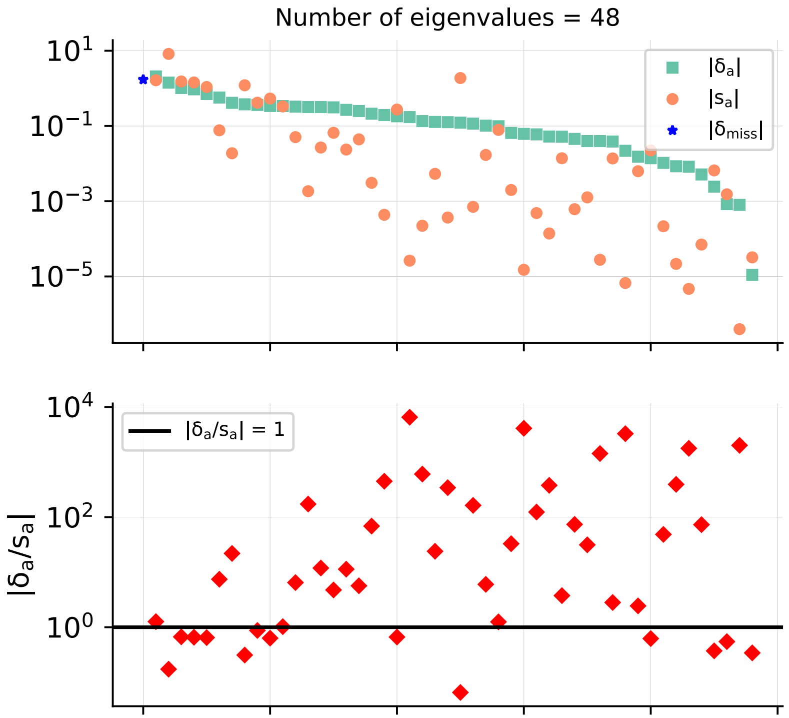
https://github.com/NNPDF/nnpdf

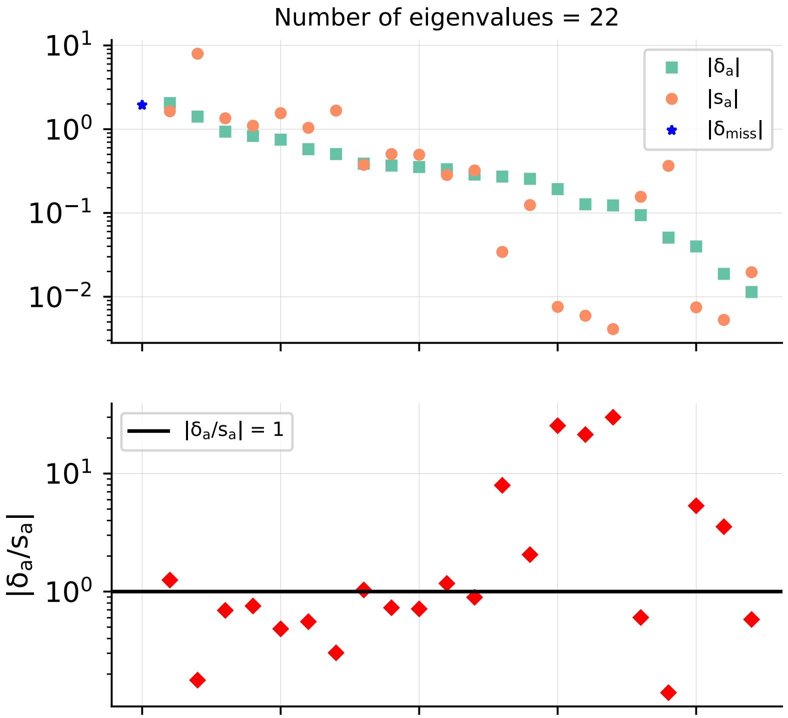
https://docs.nnpdf.science/



Validation: comparing point prescriptions.

$$\delta_i = \left(\frac{F_i^{NNLO} - F_i^{NLO}}{F_i^{NLO}}\right) \rightarrow \delta^\alpha = \sum_{i=1}^{N_D} \delta_i e_i^\alpha \rightarrow \delta_i^S = \sum_{\alpha=1}^{N_{sub}} \delta^\alpha e_i^\alpha \rightarrow \theta = \arccos\left(\frac{|\delta^S|}{|\delta|}\right) \rightarrow \delta_i^{miss} = \delta_i - \delta_i^S$$





Where e^α are the eigenvectors of the theory covariance matrix with eigenvalue $\lambda^\alpha = (s^\alpha)^2$ such that $s^\alpha > 0$

Good agreement for the **largest eigenvalues** with both prescriptions.

9 pts prescription **underestimates** the size of the shift for smaller eigenvalues

However 9 pts prescription performs better in terms of the **angles** θ

Prescription	N_{sub}	θ [°] DIS NC	DIS CC	TOP	DY NC	DY CC	SINGLETOP	JETS	PHOTON	DIJET	TOTAL
7-point	22	39	18	24	23	38	14	15	12	12	32
9-point	48	37	15	20	23	34	12	13	7	12	28