Global fits of PDFs: insights from the NNPDF framework

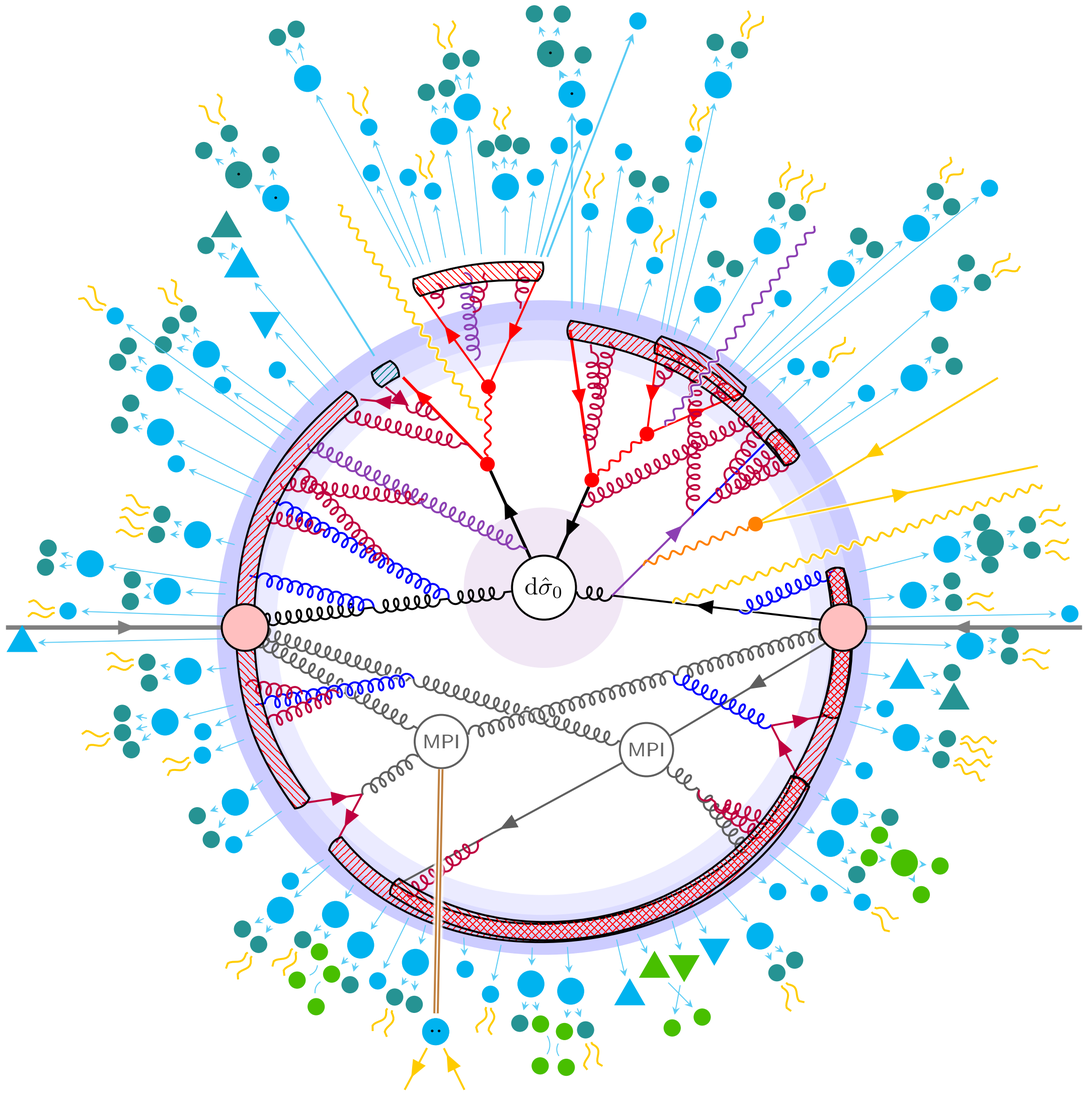
Digital Twins For Nuclear and Particle Physics - NPTWINS 2026



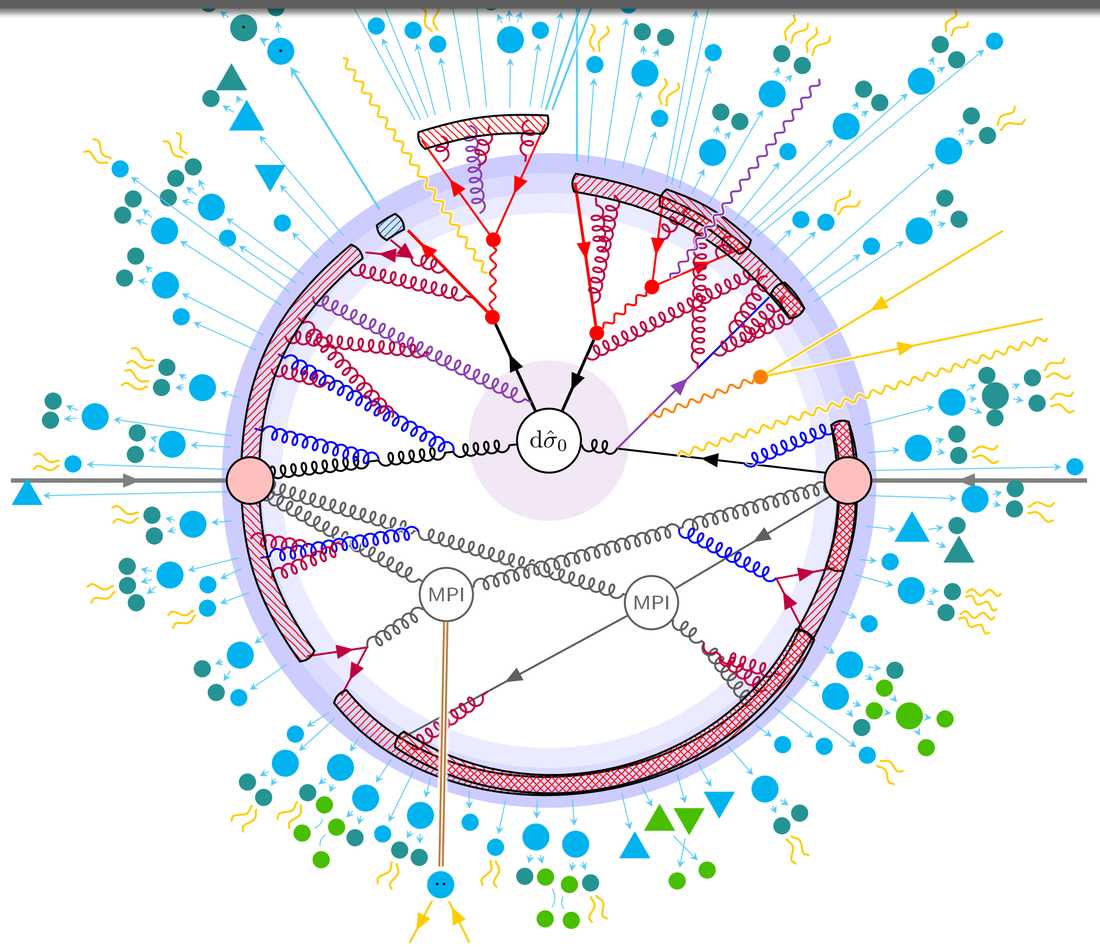


UNIVERSITÀ
DI TORINO

Tommaso Giani, 11/05/2026

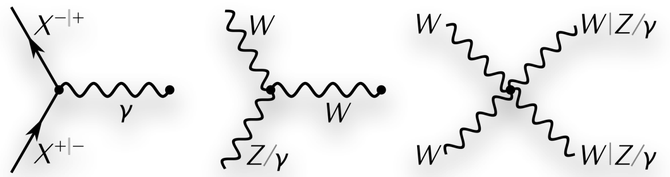

$\mathrm{d}\hat{\sigma}_0$
MPI
MPI

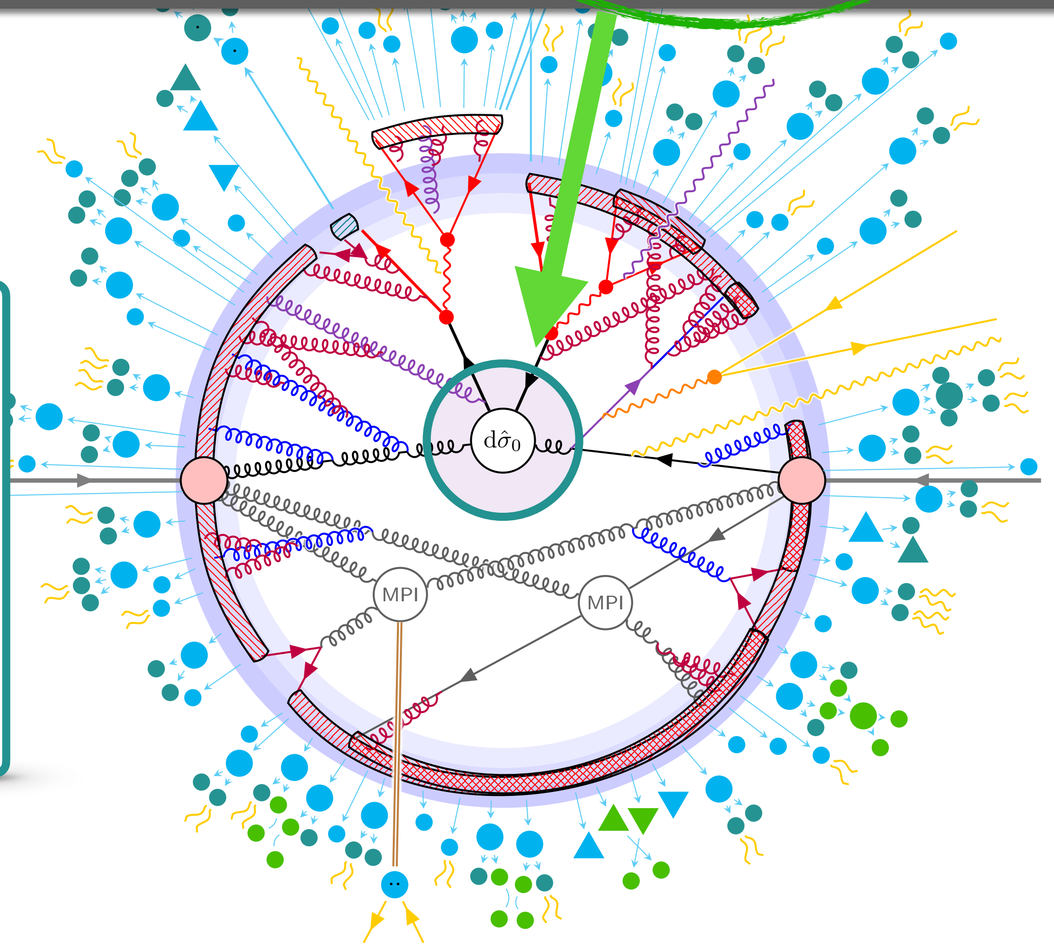
$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$



$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right)\left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

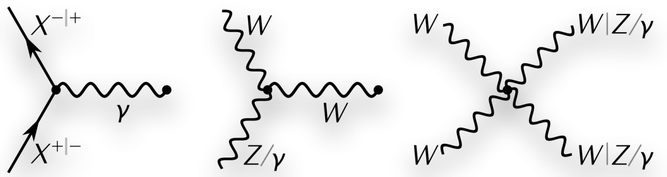
Hard scattering $\hat{\sigma}_0$: encodes **short-range interactions**; computed from first principles.

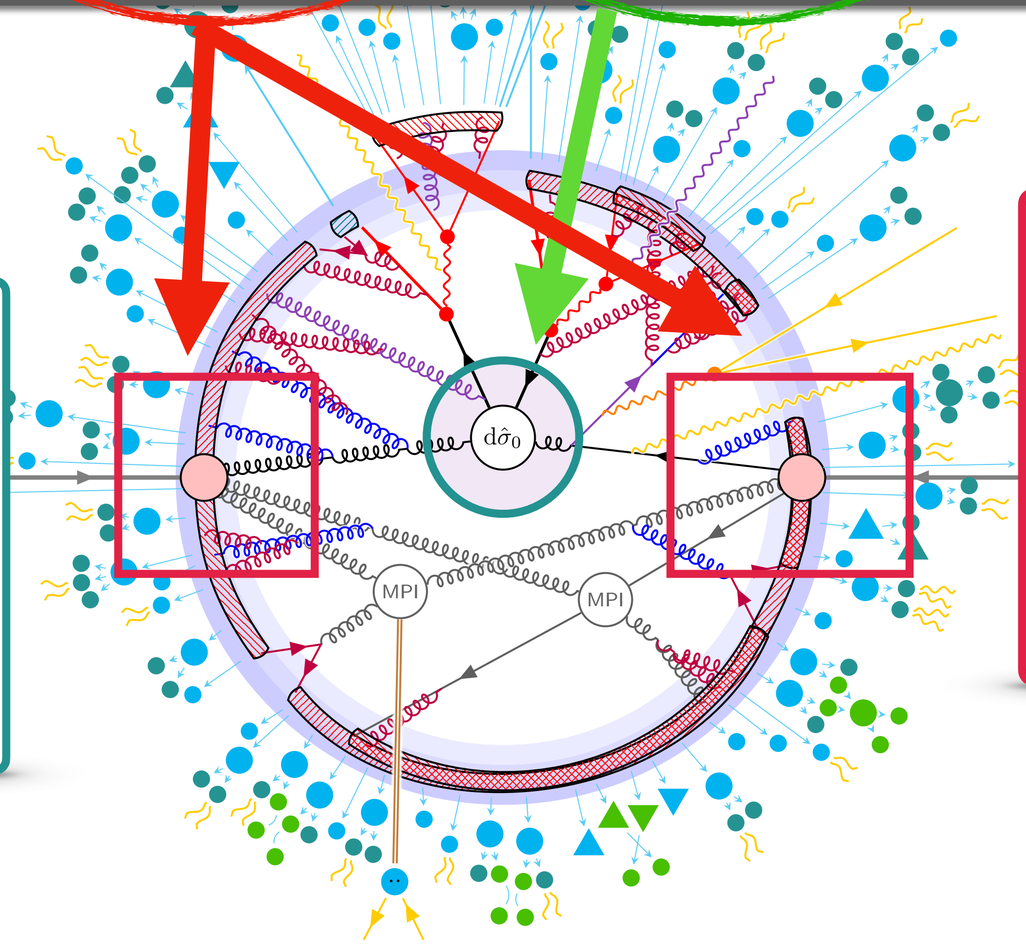




$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right)\left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Hard scattering $\hat{\sigma}_0$: encodes **short-range interactions**; computed from first principles.



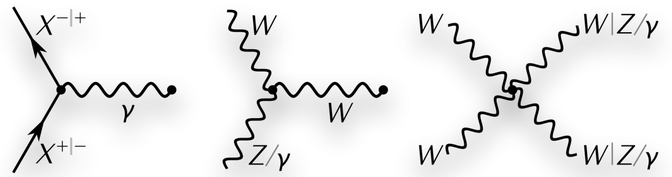


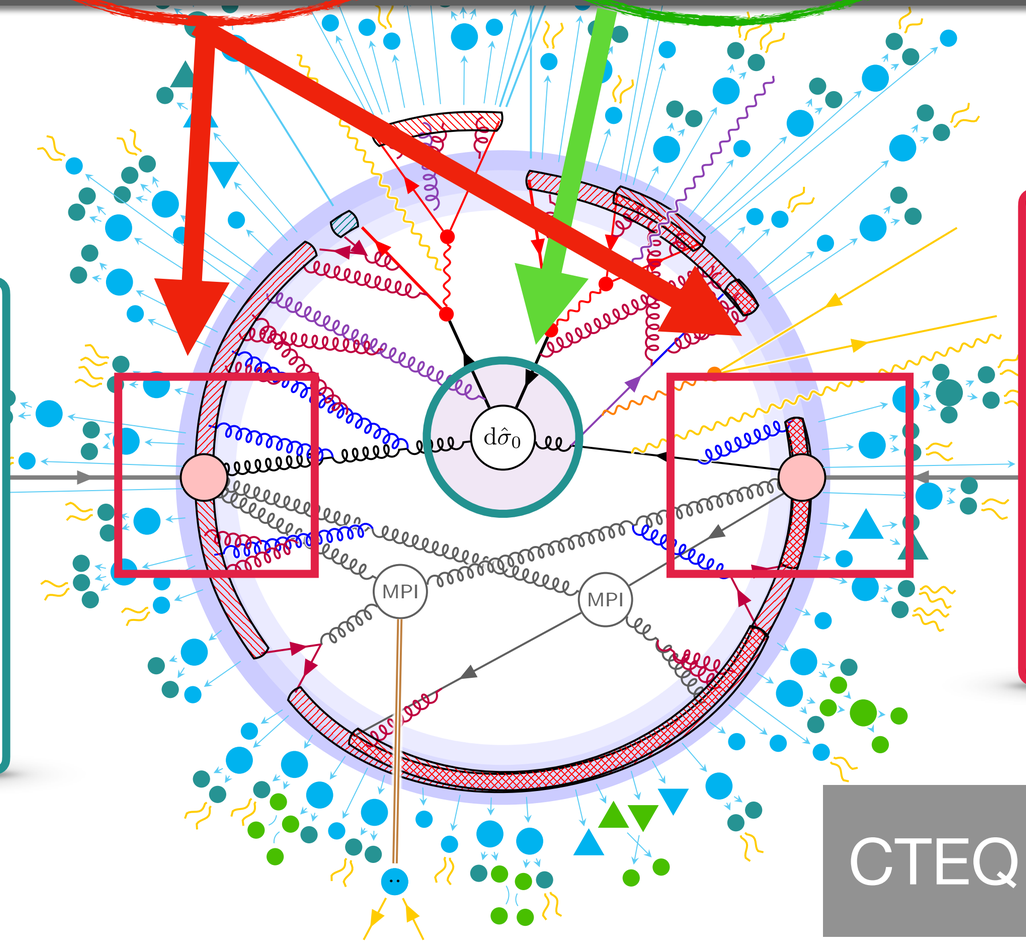
Parton Distribution Functions (PDFs): encodes **long-range non-perturbative interactions**; cannot be computed from first principle and **have to be determined from experimental Data**.

PDFs are Universal

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right)\left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Hard scattering $\hat{\sigma}_0$: encodes **short-range interactions**; computed from first principles.





Parton Distribution Functions (PDFs): encodes **long-range non-perturbative interactions**; cannot be computed from first principle and **have to be determined from experimental Data**.

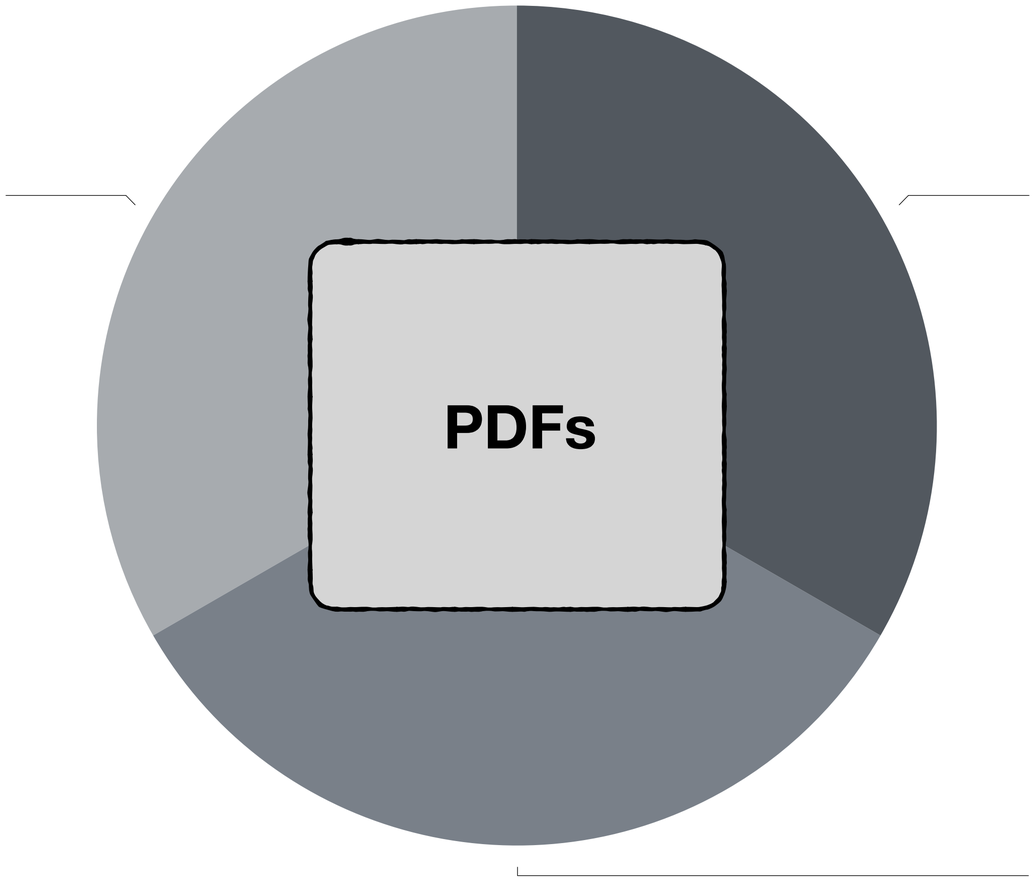
PDFs are Universal

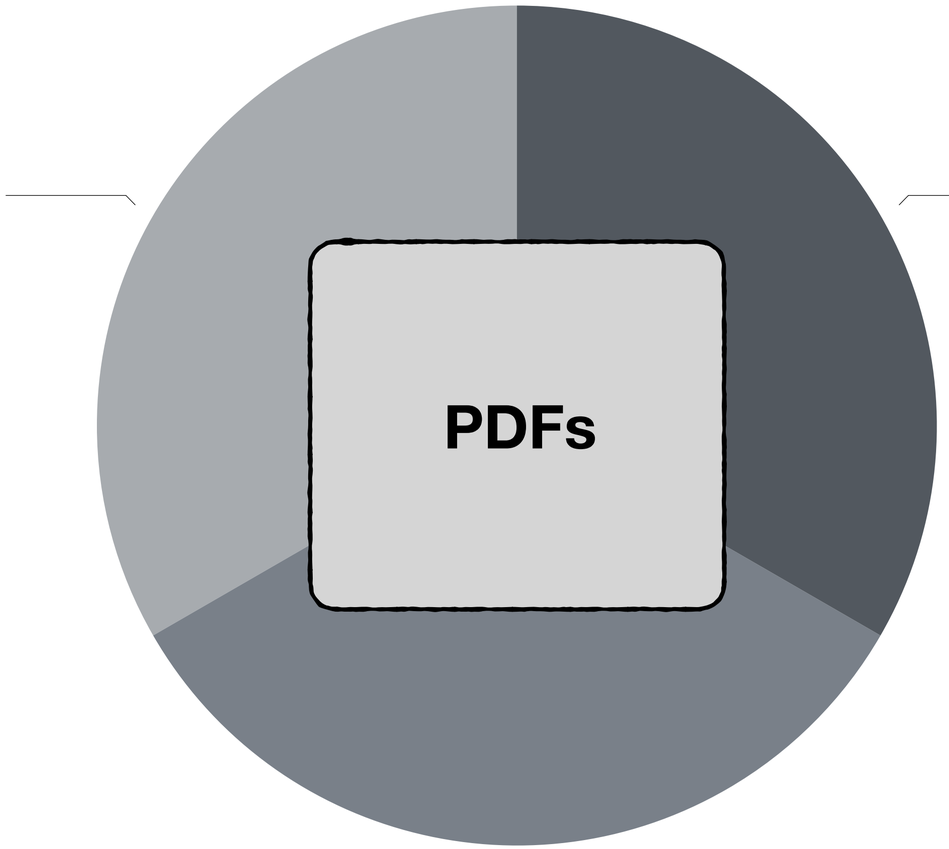
CTEQ MSHT NNPDF

1. **The NNPDF framework**

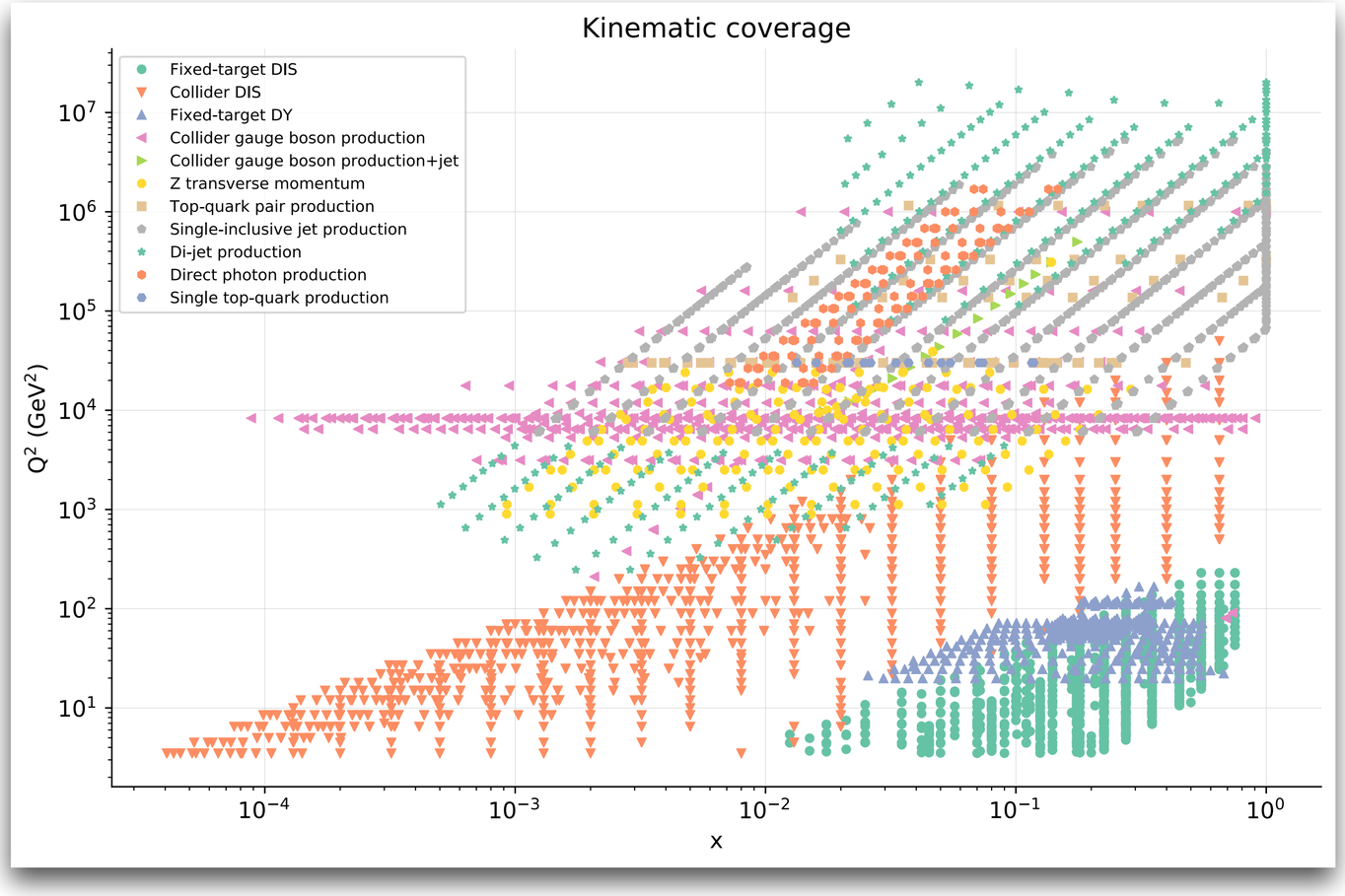
2. **Building the NNPDF methodology:**
   ensemble regression hyperoptimisation algorithm

The NNPDF framework


PDFs

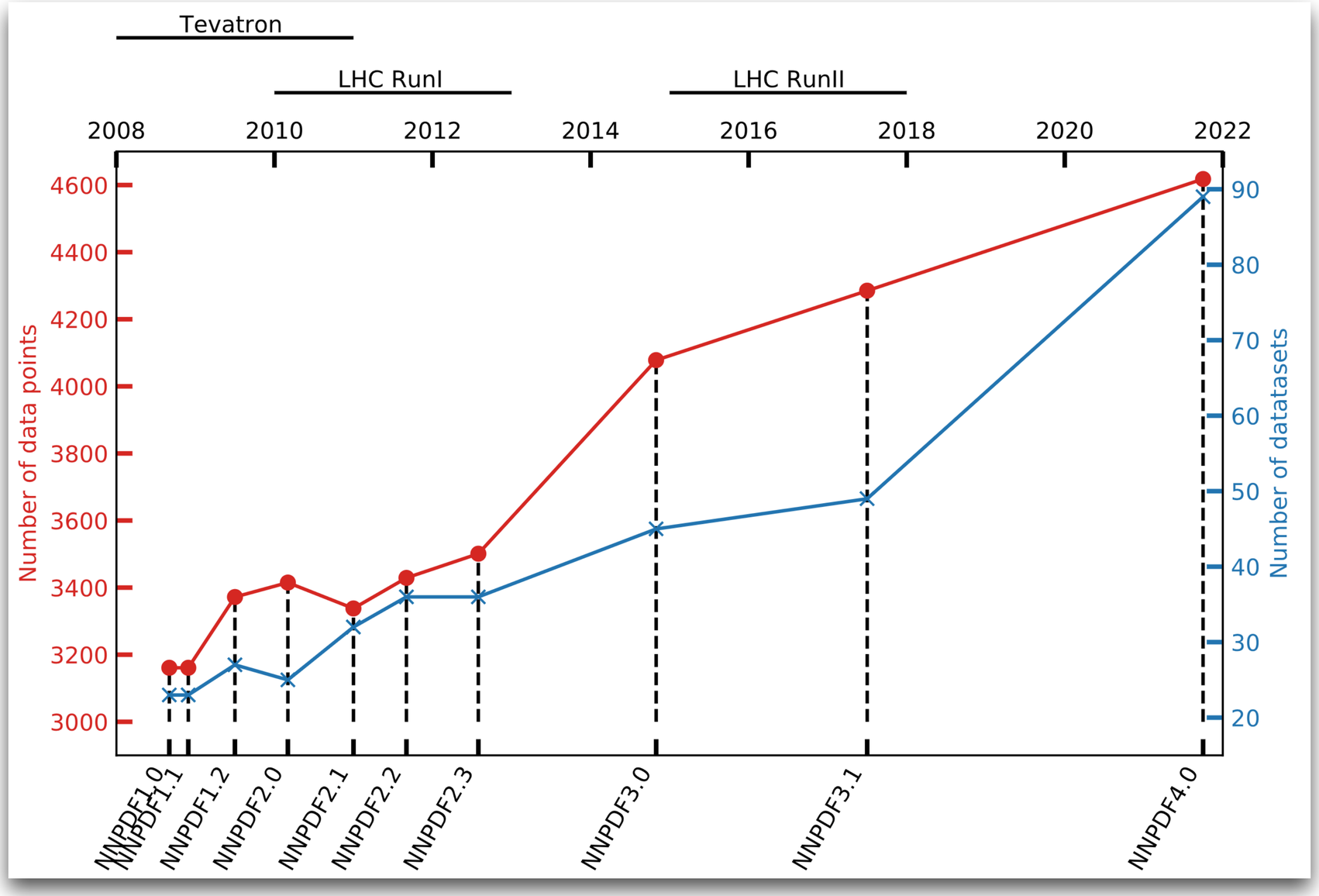

PDFs

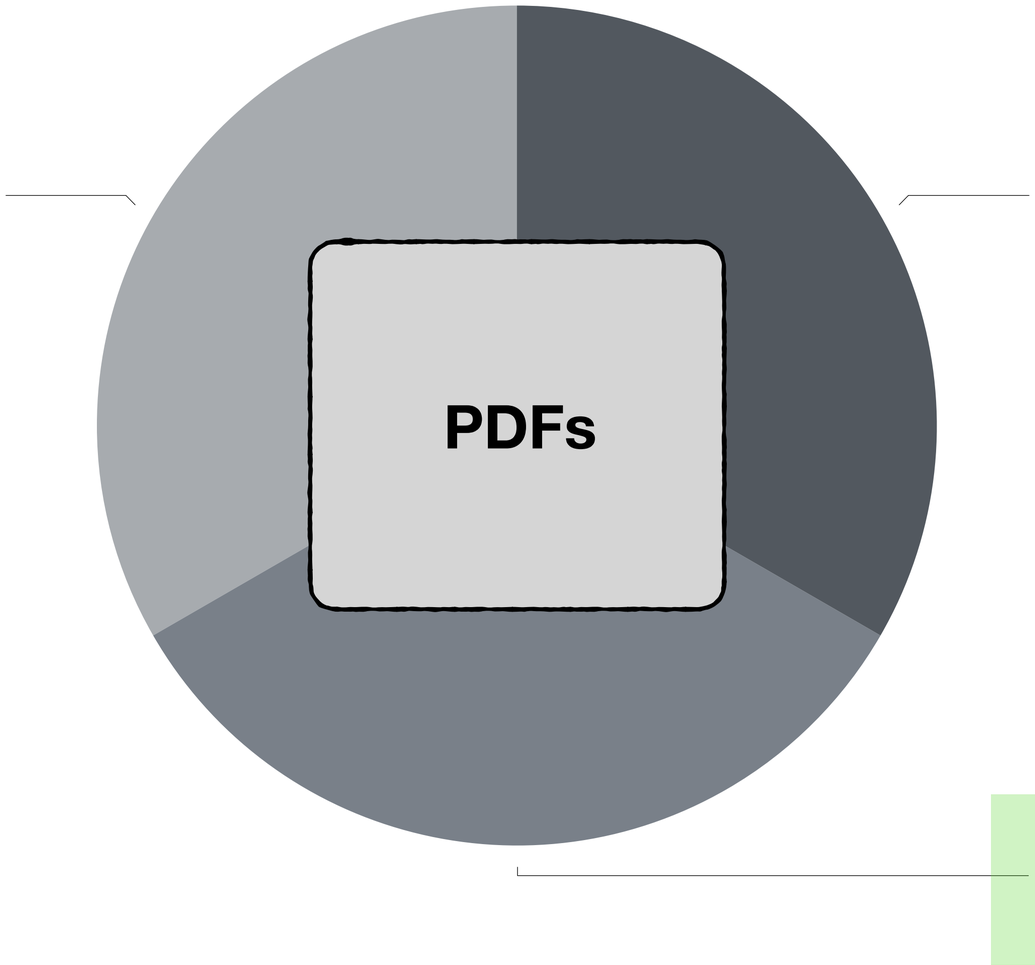
Experimental data



- $\mathcal{O}(5000)$ **datapoints** that span a wide range of kinematic regions
- Precision of the data reach the **percent level accuracy**
- Significant amount of the datasets ($\mathcal{O}(500)$ **datapoints**) were introduced in the NNPDF4.0 release (LHC Run II data)

From NNPDF1.0 to 4.0 …




PDFs

Theory

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Theory

Fixed order accuracy

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Theory

PDF evolution

Fixed order accuracy

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Theory

PDF evolution

Fixed order accuracy

Treatment of missing higher order uncertainties

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Theory

PDF evolution

Fixed order accuracy

Treatment of missing higher order uncertainties

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Higher twist effects

Theory

PDF evolution

Fixed order accuracy

Treatment of missing higher order uncertainties

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Heavy flavour scheme

Higher twist effects

Theory

PDF evolution

Fixed order accuracy

Treatment of missing higher order uncertainties

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Heavy flavour scheme

Higher twist effects

Treatment of charm PDF

Theory

PDF evolution

Fixed order accuracy

Treatment of missing higher order uncertainties

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

Heavy flavour scheme

Higher twist effects

QED

Treatment of charm PDF

Theory

PDF evolution

Fixed order accuracy

Treatment of missing higher order uncertainties

$$\sigma = \sum_{i,j} \int dx_1 dx_2 \; f_i\left(x_1, \mu\right) f_j\left(x_2, \mu\right) C_{ij}\left(x_1 x_2 s, \alpha_s, \frac{Q}{\mu}\right) \left[1 + \mathcal{O}\left(\frac{\Lambda^p}{Q^p}\right)\right]$$

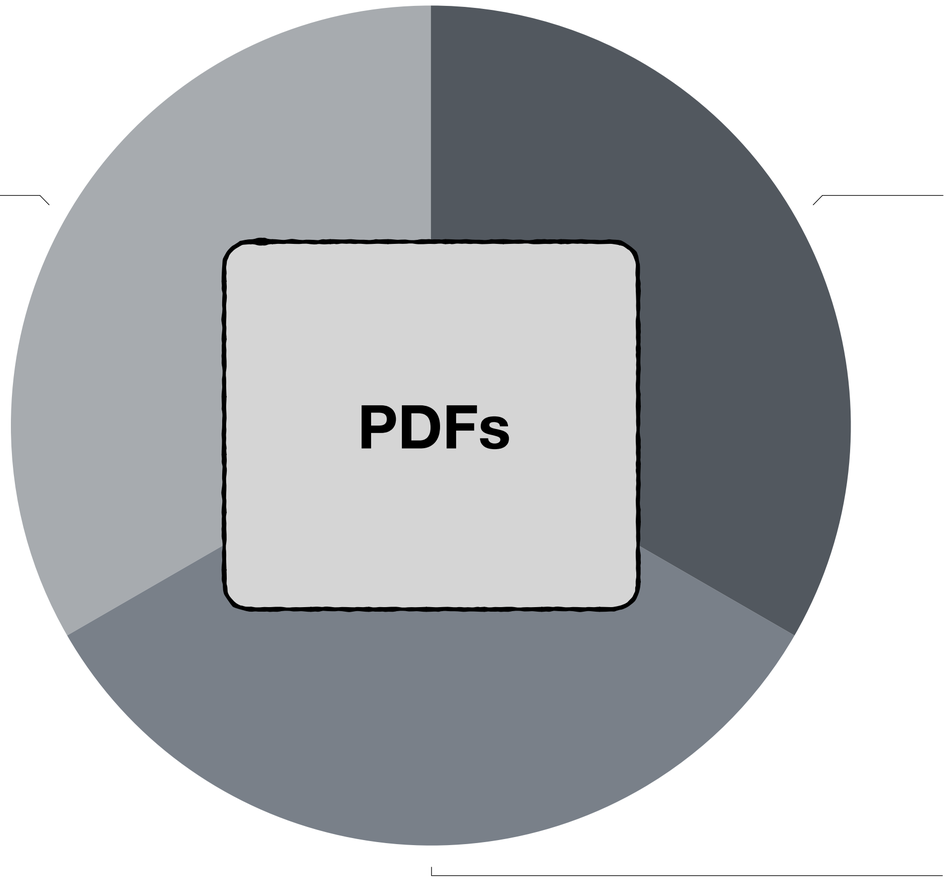
Heavy flavour scheme

Higher twist effects

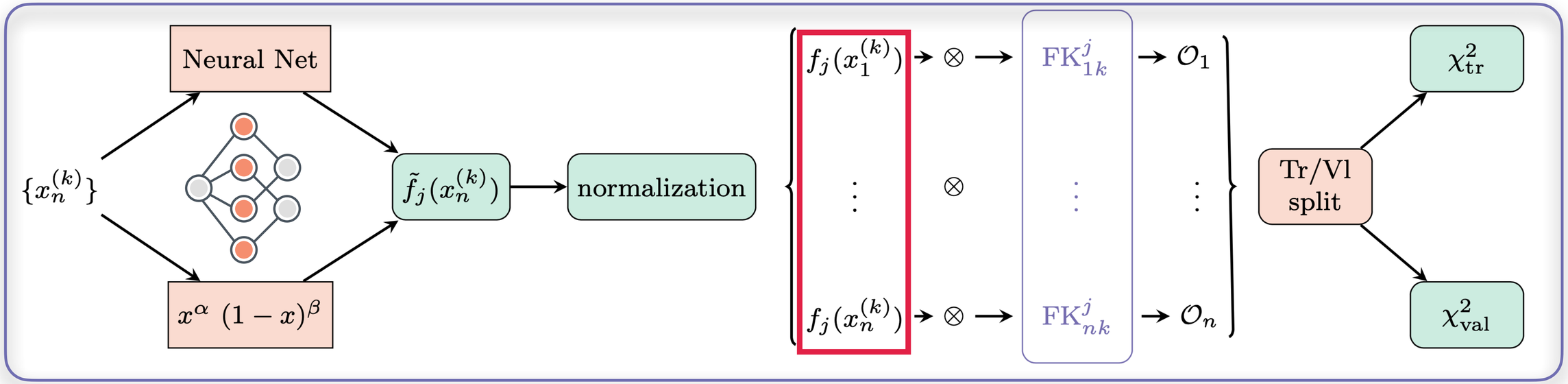
QED

Treatment of charm PDF

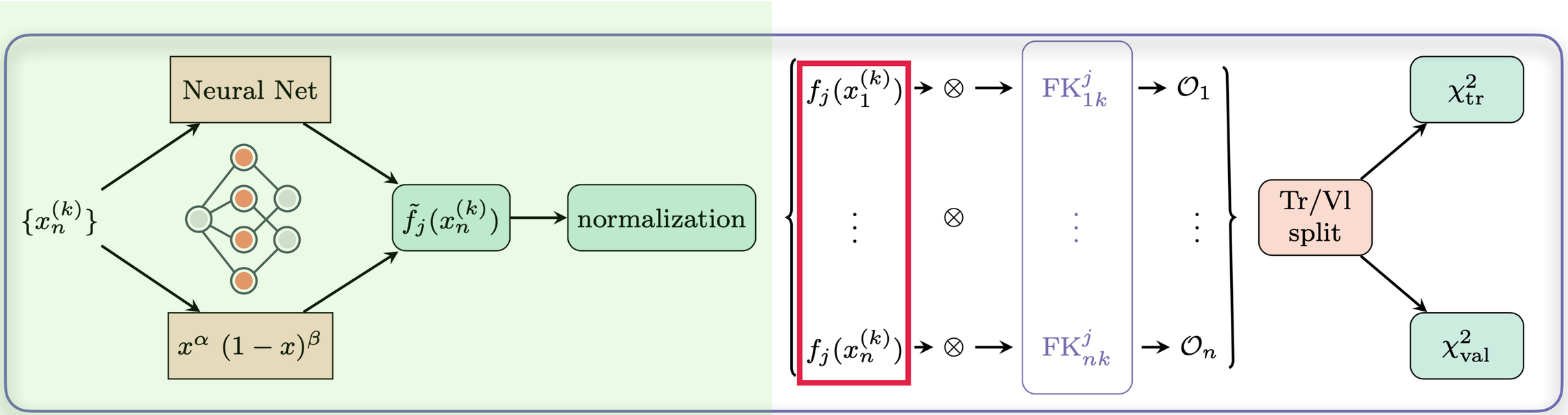
Interpolation grids


PDFs

Methodology

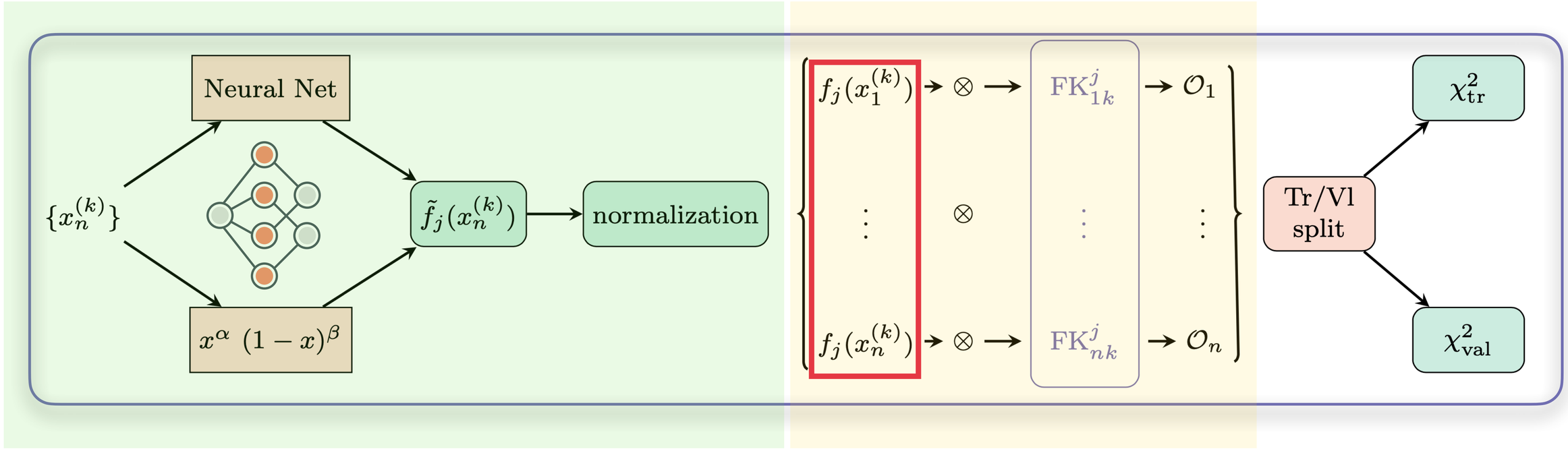


Methodology



Parameterisation

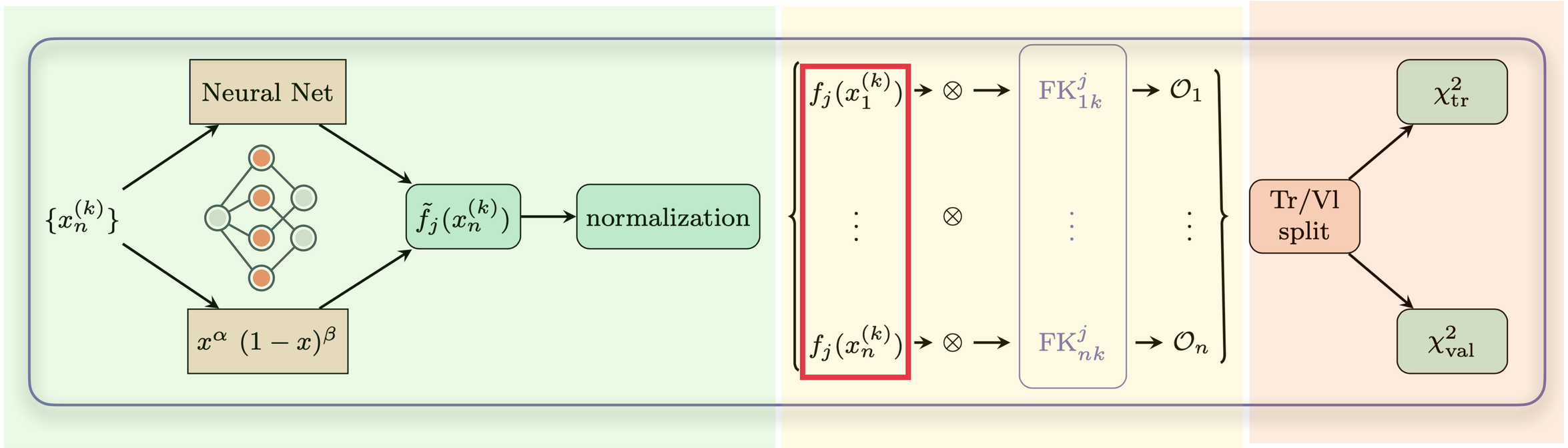
Methodology



Parameterisation

Perturbative input
(fast interpolation grids)

Methodology



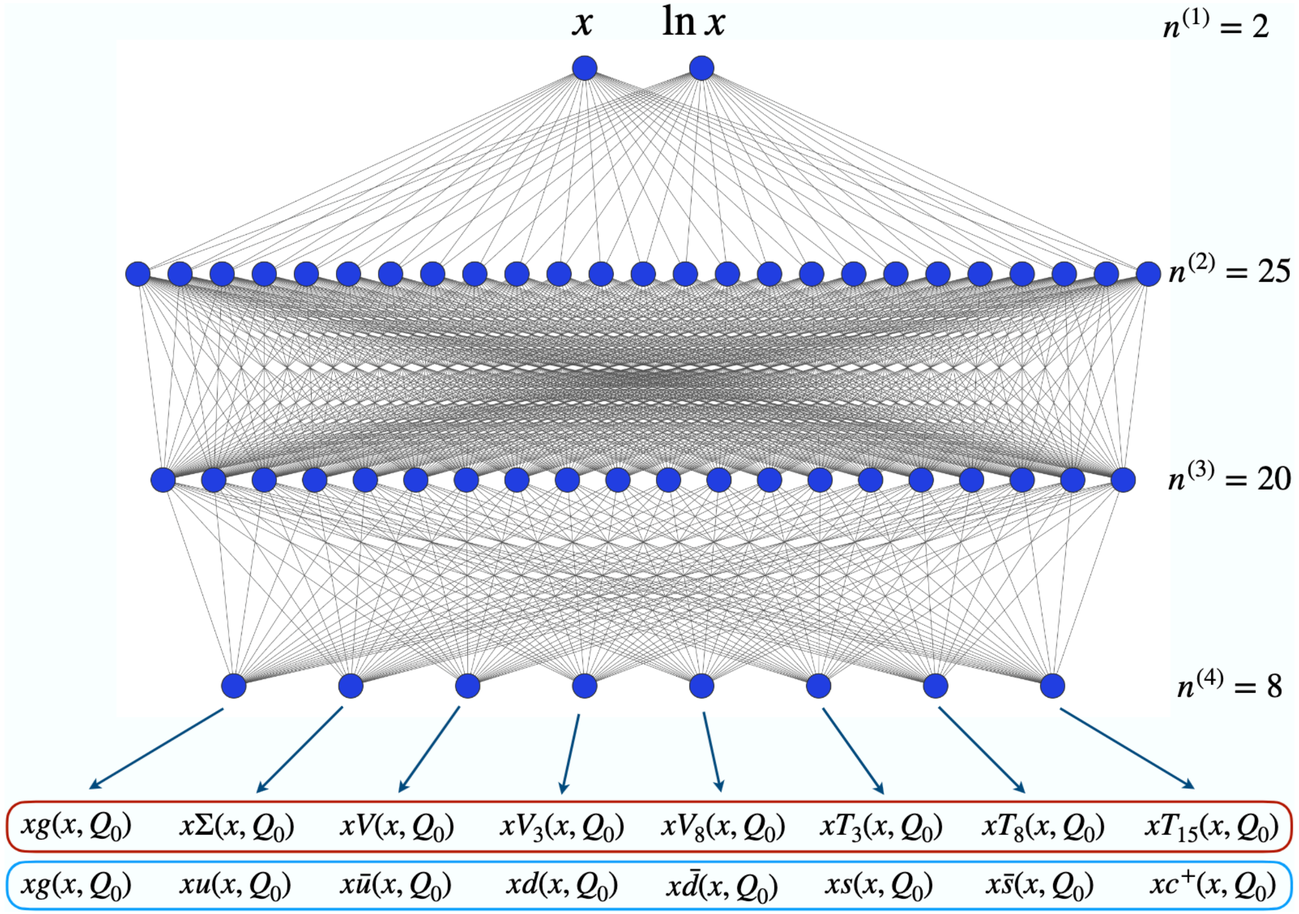
Parameterisation

Perturbative input
(fast interpolation grids)

Training

Parameterisation

DNNs to parametrise the PDFs using Stochastic Gradient Descent (SGD) as minimiser



$$xf(x,Q_0;\vec{w}) = A(1-x)^{\eta}x^{\delta}\,\mathrm{NN}\left(x;\vec{w}\right)$$

PDFs are usually extracted in the so-called **Evolution Basis**.

$$\int_0^1 dx\,x\left(\Sigma(x,Q)+g(x,Q)\right) = \int_0^1 dx V_3(x,Q) = 1$$

$$\int_0^1 dx V(x,Q) = \int_0^1 dx V_8(x,Q) = 3$$

Error propagation: Monte Carlo replicas

- Generate N data replicas using experimental distributions (full info about uncertainties and their correlations)
- Perform N independent fits:
  ensemble of N trained PDF replicas $\{\tilde{f}_k, k = 1, \ldots, N\}$

Samples from $P(f\,|\,data)$

$$E[\mathcal{O}] = \int Df\, P(f\,|\,data)\, \mathcal{O}(f)$$

$$V[\mathcal{O}] = \int Df\, P(f\,|\,data)\, \left[\mathcal{O}\left(f\right) - E[\mathcal{O}]\right]^2$$

$$E[\mathcal{O}] \approx \frac{1}{N}\sum_k \mathcal{O}\left(\tilde{f}_k\right)$$

$$V[\mathcal{O}] \approx \frac{1}{N}\sum_k \left[\mathcal{O}\left(\tilde{f}_k\right) - E[\mathcal{O}]\right]^2$$

Building the  methodology: ensemble regression hyperoptimisation algorithm

Hyperparameters θ

One of the main reasons to resort to Neural Network (NN) was **to reduce biases** in defining a functional form, however:

❌ The hyperparameters θ that define the NN have to be chosen

❌ Random and/or Manual selection of hyperparameters θ are tedious and not guaranteed

Parameter	NNPDF4.0
Architecture	2-25-20-8
Activation function	hyperbolic tangent
Initializer	`glorot_normal`
Optimizer	`Nadam`
Clipnorm	6.0×10^{-6}
Learning rate	2.6×10^{-3}
Maximum # epochs	17×10^{3}
Stopping patience	10% of max epochs
Initial positivity $\Lambda^{(\mathrm{pos})}$	185
Initial integrability $\Lambda^{(\mathrm{int})}$	10

Hyperparameters θ

One of the main reasons to resort to Neural Network (NN) was **to reduce biases** in defining a functional form, however:

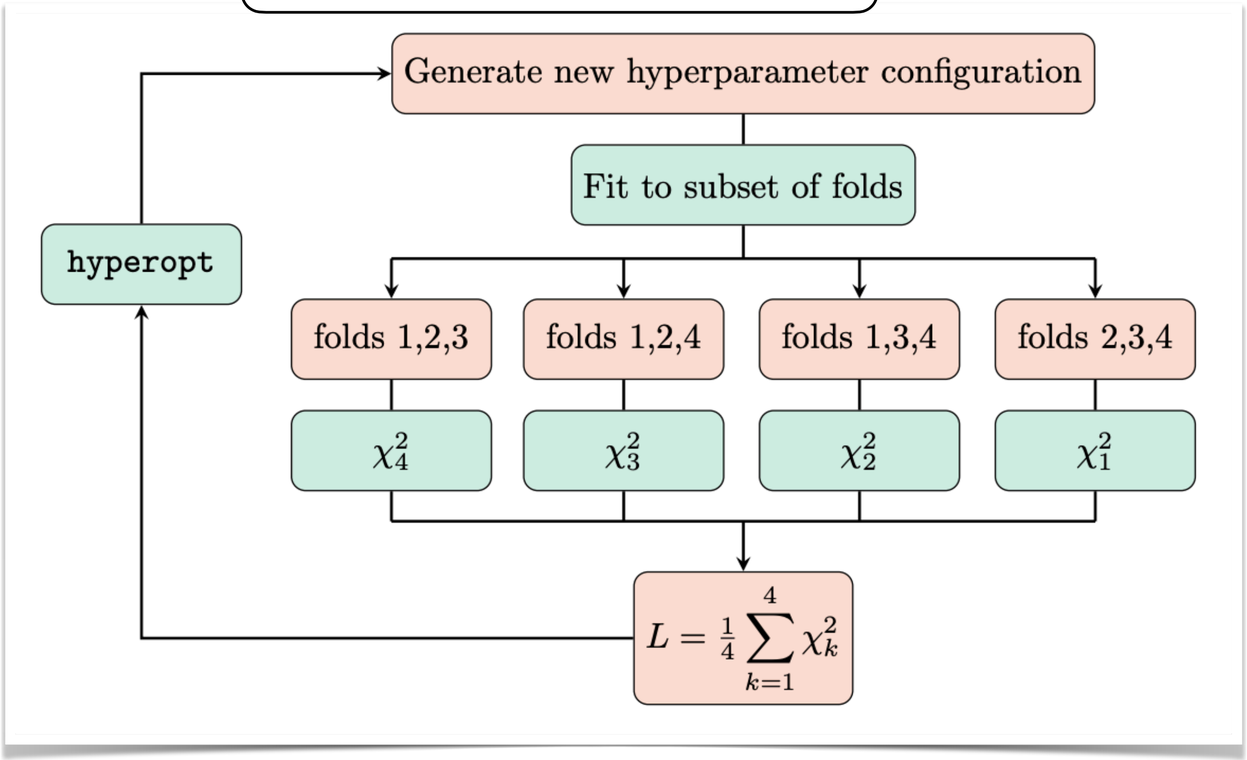
❌ The hyperparameters θ that define the NN have to be chosen

❌ Random and/or Manual selection of hyperparameters θ are tedious and not guaranteed

Parameter	NNPDF4.0
Architecture	2-25-20-8
Activation function	hyperbolic tangent
Initializer	`glorot_normal`
Optimizer	`Nadam`
Clipnorm	6.0×10^{-6}
Learning rate	2.6×10^{-3}
Maximum # epochs	17×10^{3}
Stopping patience	10% of max epochs
Initial positivity $\Lambda^{(\mathrm{pos})}$	185
Initial integrability $\Lambda^{(\mathrm{int})}$	10

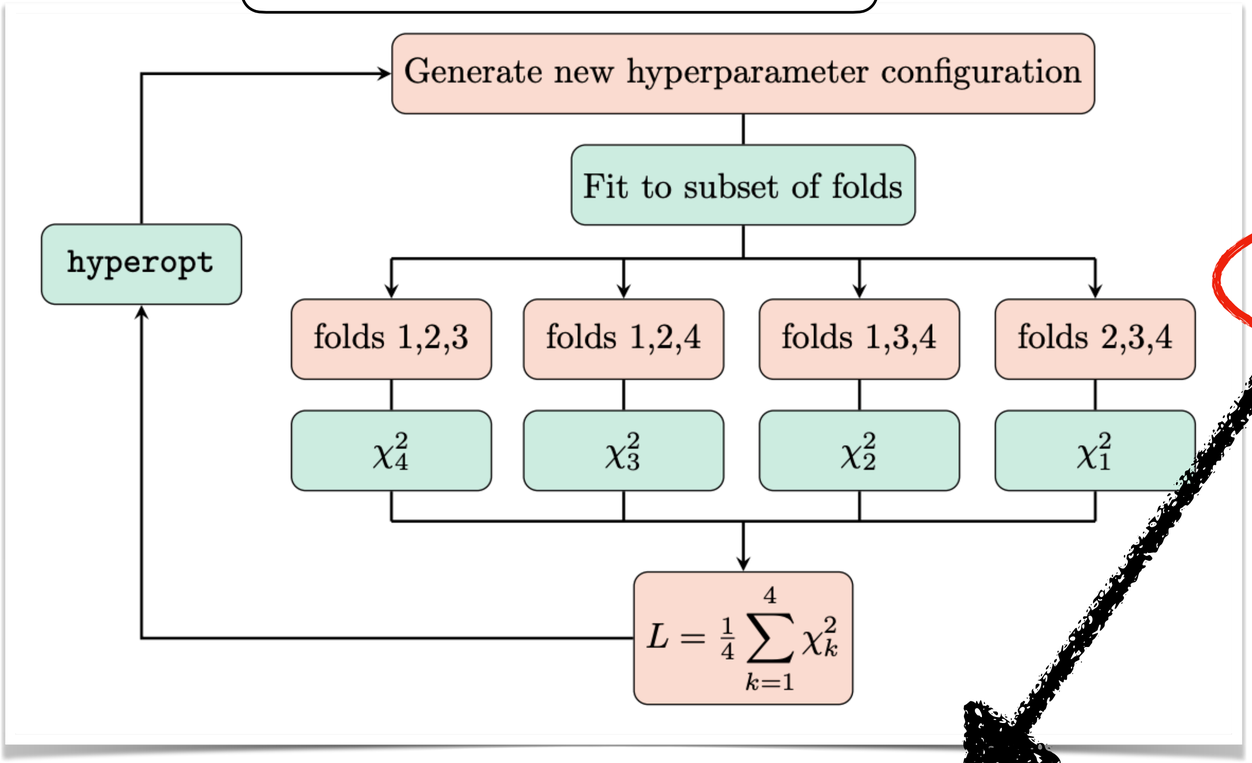
✅ Perform an **automated scan of the search space** by running fits with thousands of hyperparameter combinations using a **suitable metric !!**

NNPDF4.0 hyperopt



- The full dataset is partitioned in different folds
- Figure of merit is defined as the χ^2 of the excluded folds
- Minimise the figure of merit and select the best hyperparam configuration

NNPDF4.0 hyperopt



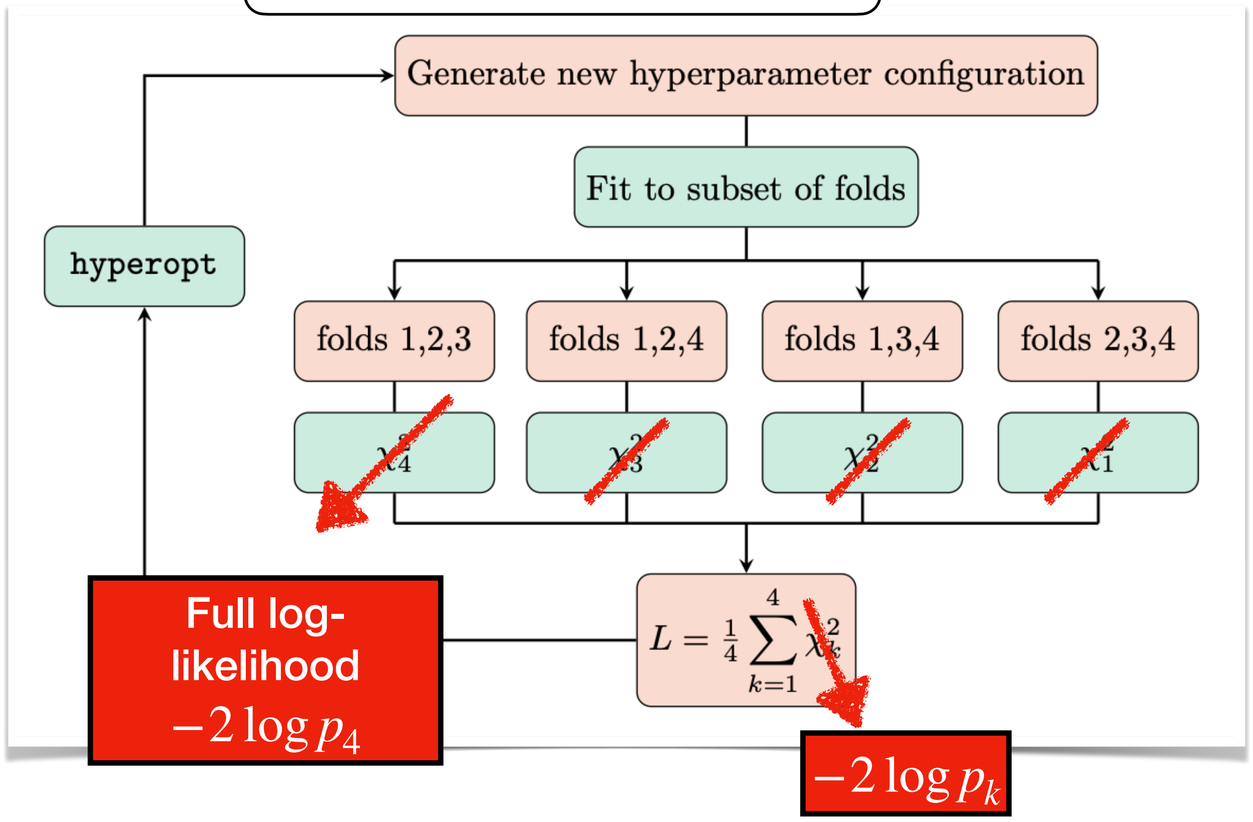
- The full dataset is partitioned in different folds
- Figure of merit is defined as the χ^2 of the excluded folds
- Minimise the figure of merit and select the best hyperparam configuration

The χ^2 does not take into account the full PDF error, but only the experimental one

The effect of hyperparameter variations is neglected in final error estimate and central values

4.1

NNPDF4.0 hyperopt



- The full dataset is partitioned in different folds
- Figure of merit is defined as the χ^2 of the excluded folds

Full log-likelihood

- Minimise the figure of merit and ~~select the best hyperparam configuration~~

Select an ensemble of statstically equivalent methodologies maximising the likelihood

Likelihood of the excluded fold k

$$p_k(\theta) = \frac{1}{N(\theta)} \exp\left[-\frac{1}{2}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)\left(C_y^{(k)} + C_{PDF}^{(k)}(\theta)\right)^{-1}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)^T\right]$$

Likelihood of the excluded fold

$$p_k(\theta) = \frac{1}{N(\theta)} \exp\left[-\frac{1}{2}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)\left(C_y^{(k)} + C_{PDF}^{(k)}(\theta)\right)^{-1}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)^T\right]$$

Central values of fold k

Likelihood of the excluded fold

$$p_k(\theta) = \frac{1}{N(\theta)} \exp\left[-\frac{1}{2}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)\left(C_y^{(k)} + C_{PDF}^{(k)}(\theta)\right)^{-1}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)^T\right]$$

Central values of fold k

Theory prediction for fold k, computed using a fit which does not include fold k itself

Likelihood of the excluded fold

$$p_k(\theta) = \frac{1}{N(\theta)} \exp\left[-\frac{1}{2}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)\left(C_y^{(k)} + C_{PDF}^{(k)}(\theta)\right)^{-1}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)^T\right]$$

Central values of fold k

Theory prediction for fold k, computed using a fit which does not include fold k itself

Experimental covmat for fold k

Likelihood of the excluded fold

$$p_k(\theta) = \frac{1}{N(\theta)} \exp\left[-\frac{1}{2}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)\left(C_y^{(k)} + C_{PDF}^{(k)}(\theta)\right)^{-1}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)^T\right]$$

Central values of fold k

Theory prediction for fold k, computed using a fit which does not include fold k itself

Experimental covmat for fold k

PDF error for fold k, computed using a fit which does not include fold k itself

Likelihood of the excluded fold

$$p_k(\theta) = \frac{1}{N(\theta)} \exp\left[-\frac{1}{2}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)\left(C_y^{(k)} + C_{PDF}^{(k)}(\theta)\right)^{-1}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)^T\right]$$

Central values of fold k

Theory prediction for fold k, computed using a fit which does not include fold k itself

Experimental covmat for fold k

PDF error for fold k, computed using a fit which does not include fold k itself

Normalisation of the likelihood $\sqrt{\det 2\pi\left(C_y^{(i)} + C_{\text{PDF}}^{(i)}(\theta)\right)}$

Running hyperopt: hyperparameter space

1) choose the hyperparameters space

Architecture	Layers	2,3,4,5
	Number of nodes	min: 10, max: 25
	Activation	sigmoid, tanh
Training	Optimizer	Nadam, Adam
	Learning rate	min: 10^{-3}, max: 10^{-2}, sampling: uniform
	Clipnorm	min: 10^{-7}, max: 10^{-5}, sampling: log
Stopping	Epochs	min: 15×10^3, max: 25×10^3
	Patience	min: 0.1, max: 0.2

- discrete variables (number of nodes and layers …)
- continuous variables (learning rate, clipnorm …)
- Conditional choices (optimizer specific parameters …)

Running hyperopt: hyperparameter space

1) choose the hyperparameters space

Architecture	Layers	2,3,4,5
	Number of nodes	min: 10, max: 25
	Activation	sigmoid, tanh
Training	Optimizer	Nadam, Adam
	Learning rate	min: 10^{-3}, max: 10^{-2}, sampling: uniform
	Clipnorm	min: 10^{-7}, max: 10^{-5}, sampling: log
Stopping	Epochs	min: 15×10^3, max: 25×10^3
	Patience	min: 0.1, max: 0.2

- discrete variables (number of nodes and layers …)
- continuous variables (learning rate, clipnorm …)
- Conditional choices (optimizer specific parameters …)

- Complex, tree-structure hyperparameter space
- single evaluation of loss requires full training of NNs across different folds and replicas

Running hyperopt: hyperparameter space

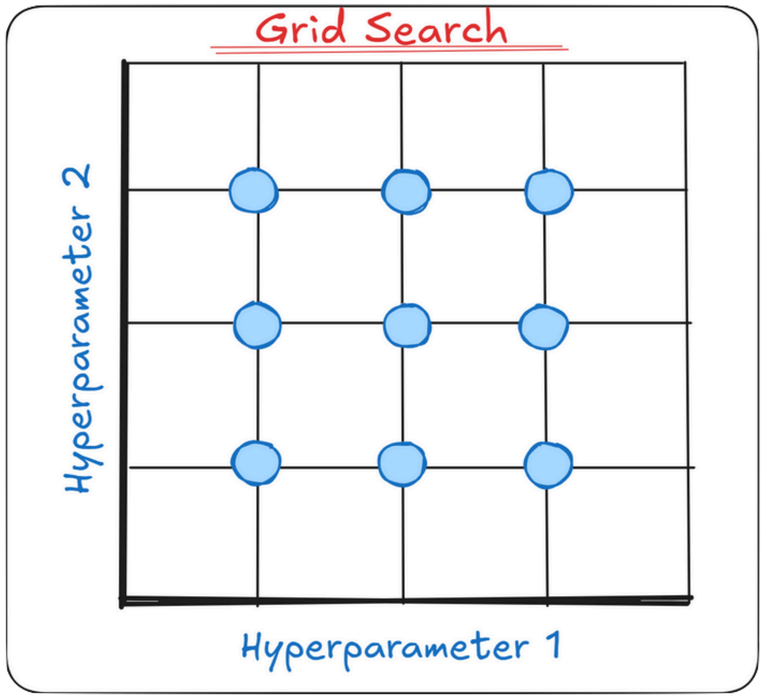
1) choose the hyperparameters space

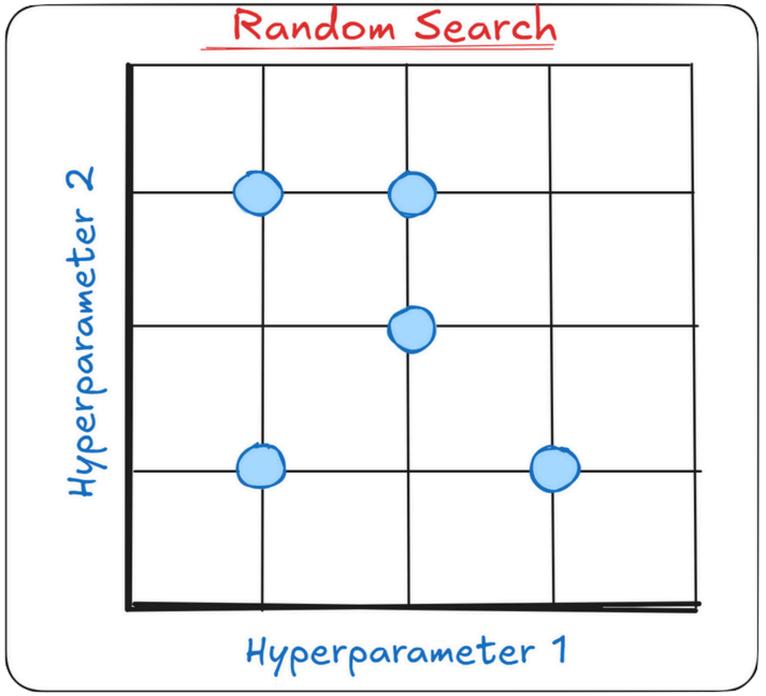
Architecture	Layers Number of nodes Activation	2,3,4,5 min: 10, max: 25 sigmoid, tanh
Training	Optimizer Learning rate Clipnorm	Nadam, Adam min: 10^{-3}, max: 10^{-2}, sampling: uniform min: 10^{-7}, max: 10^{-5}, sampling: log
Stopping	Epochs Patience	min: 15×10^3, max: 25×10^3 min: 0.1, max: 0.2

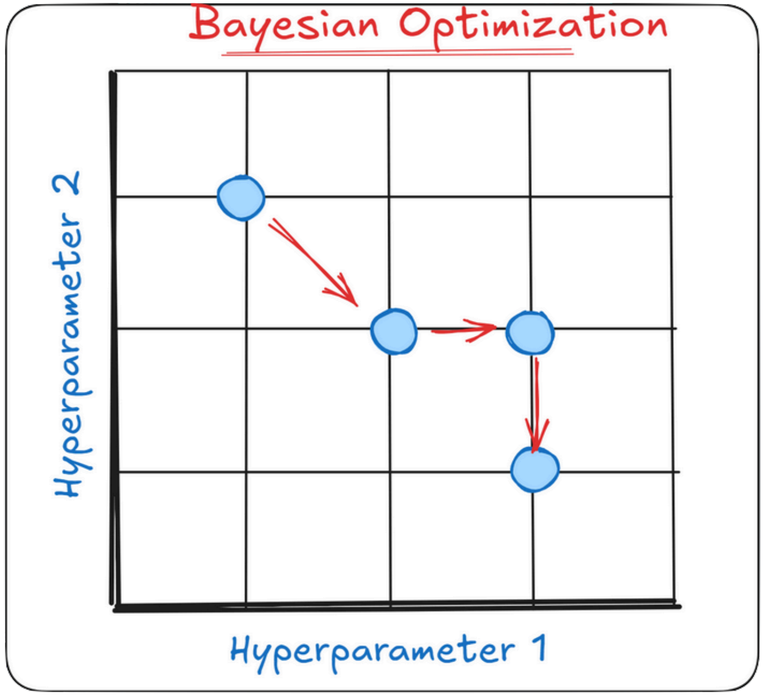
- discrete variables (number of nodes and layers …)
- continuous variables (learning rate, clipnorm …)
- Conditional choices (optimizer specific parameters …)

- Complex, tree-structure hyperparameter space
- single evaluation of loss requires full training of NNs across different folds and replicas

→ Random or grid-based scans computationally prohibitive ✗

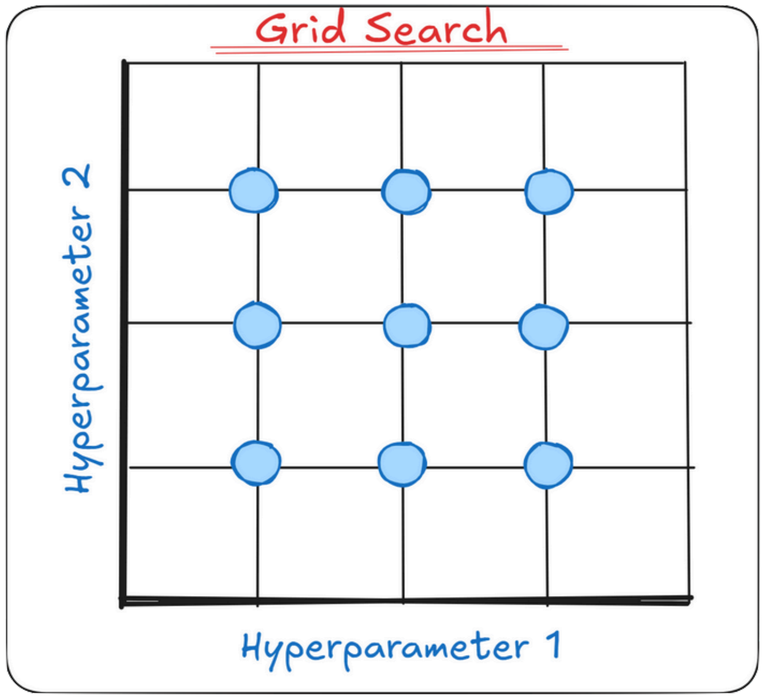


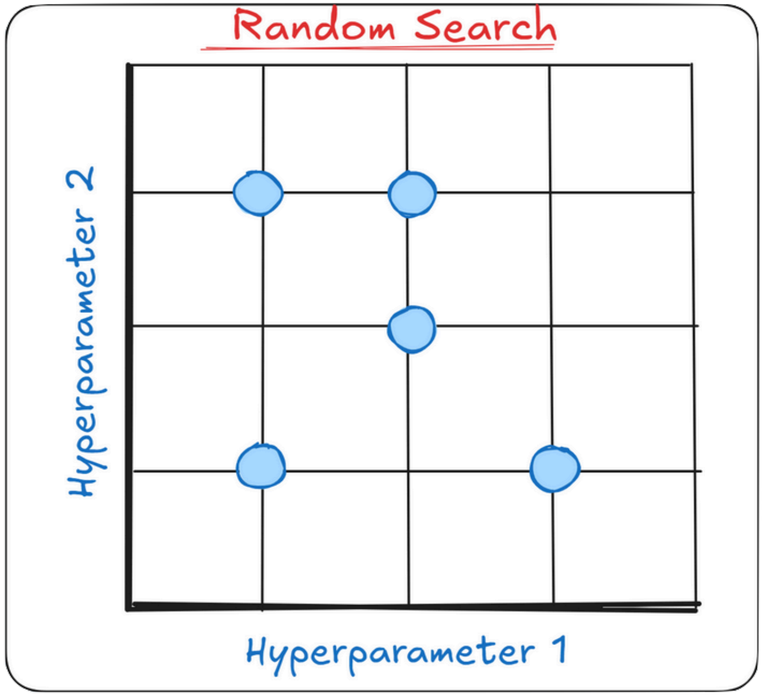


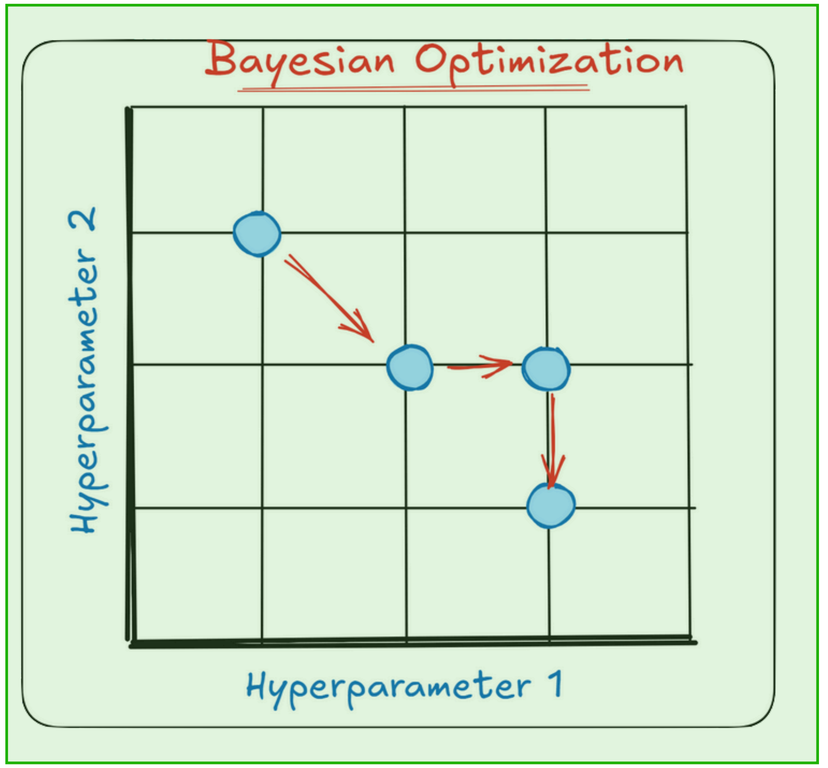


- **Static Search Algorithms:**
  Grid Search, Random Search

- **Adaptive & Model-Based Search Algorithms:**
  accumulate evaluations of (θ_i, L_i), model probability density $p(\theta|L)$ and sample new hyperparameter configurations in regions dictated by $p(\theta|L)$.



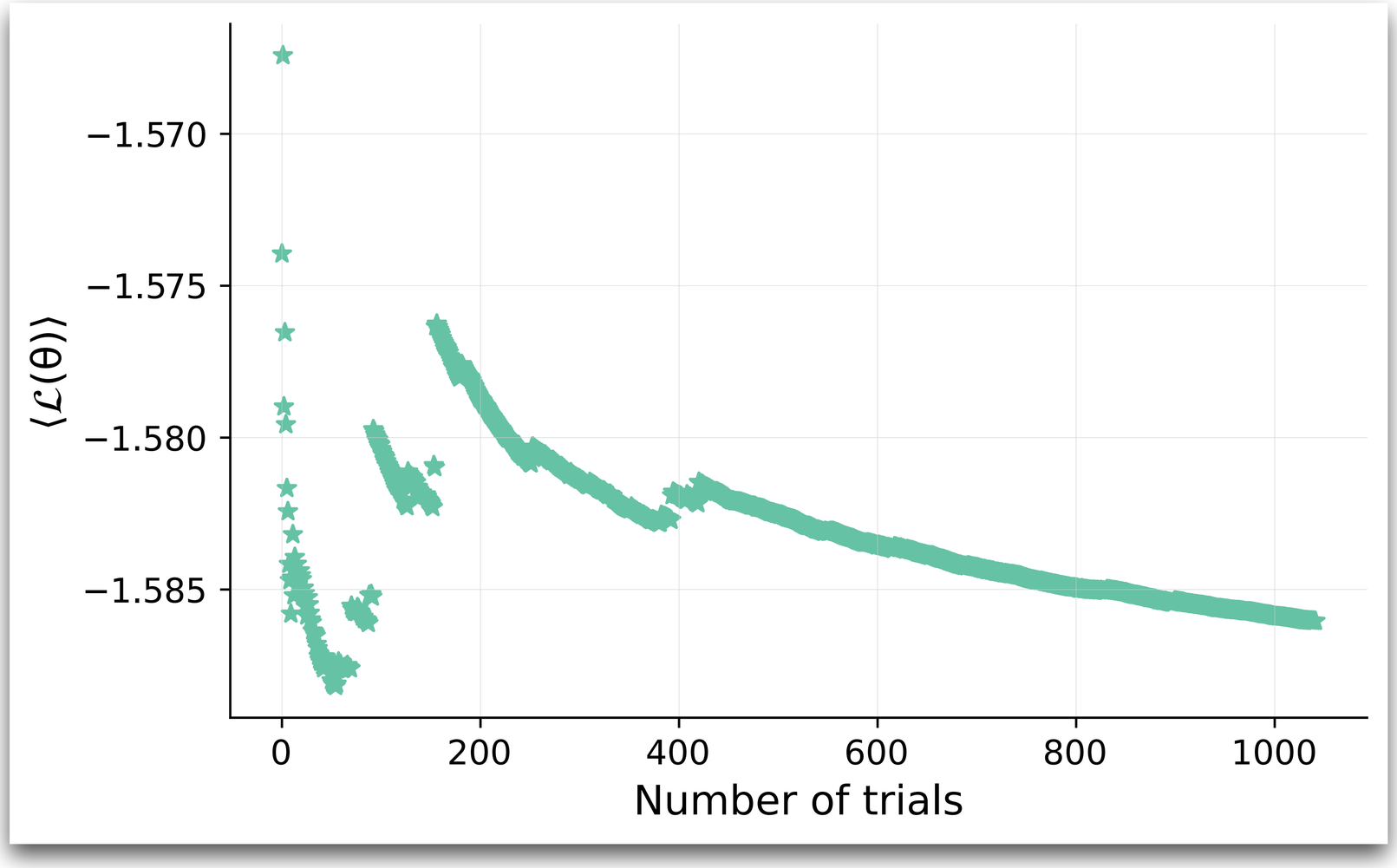




- **Static Search Algorithms:**
  Grid Search, Random Search

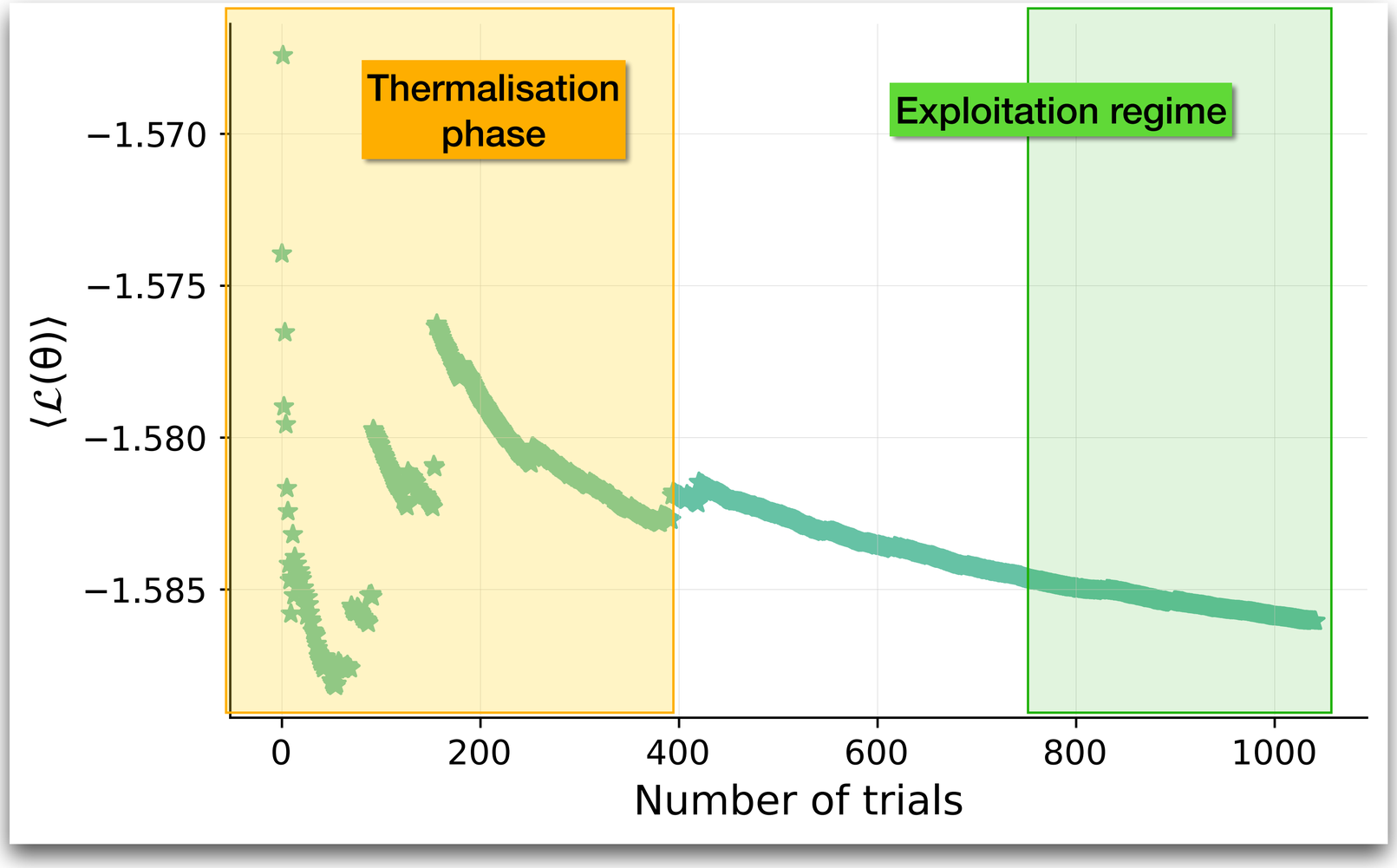
- **Adaptive & Model-Based Search Algorithms:**
  accumulate evaluations of (θ_i, L_i), model probability density $p(\theta|L)$ and sample new hyperparameter configurations in regions dictated by $p(\theta|L)$.

2) Run Tree-structured Parzen Estimator (TPE) algorithm



- We use **hyperopt library** [Bergstra, J., Yamins, D., Cox, D. D. (2013) Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures.]

2) Run Tree-structured Parzen Estimator (TPE) algorithm

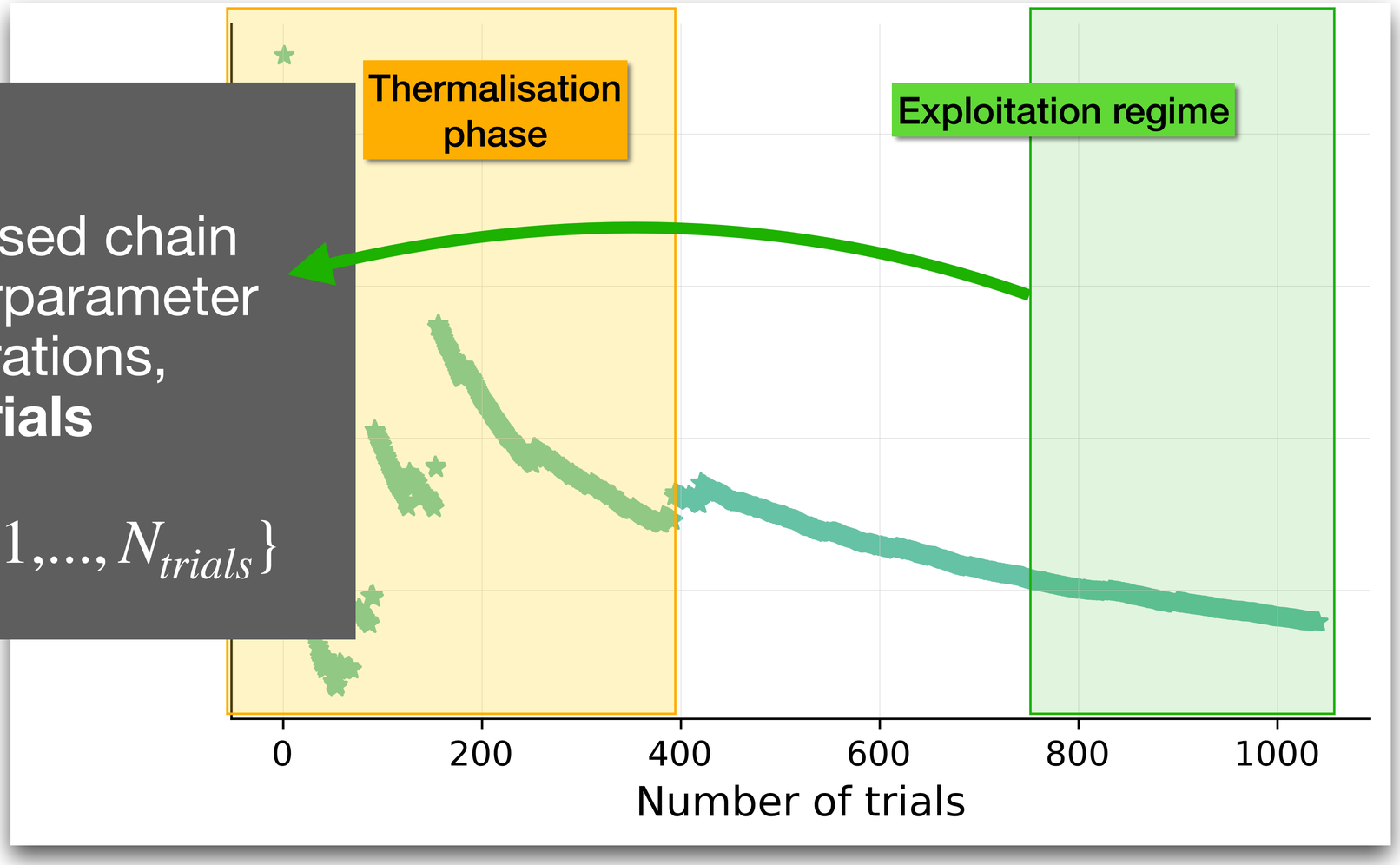


- We use **hyperopt library** [Bergstra, J., Yamins, D., Cox, D. D. (2013) Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures.]

2) Run Tree-structured Parzen Estimator (TPE) algorithm

Output:
thermalised chain of hyperparameter configurations, called **trials**

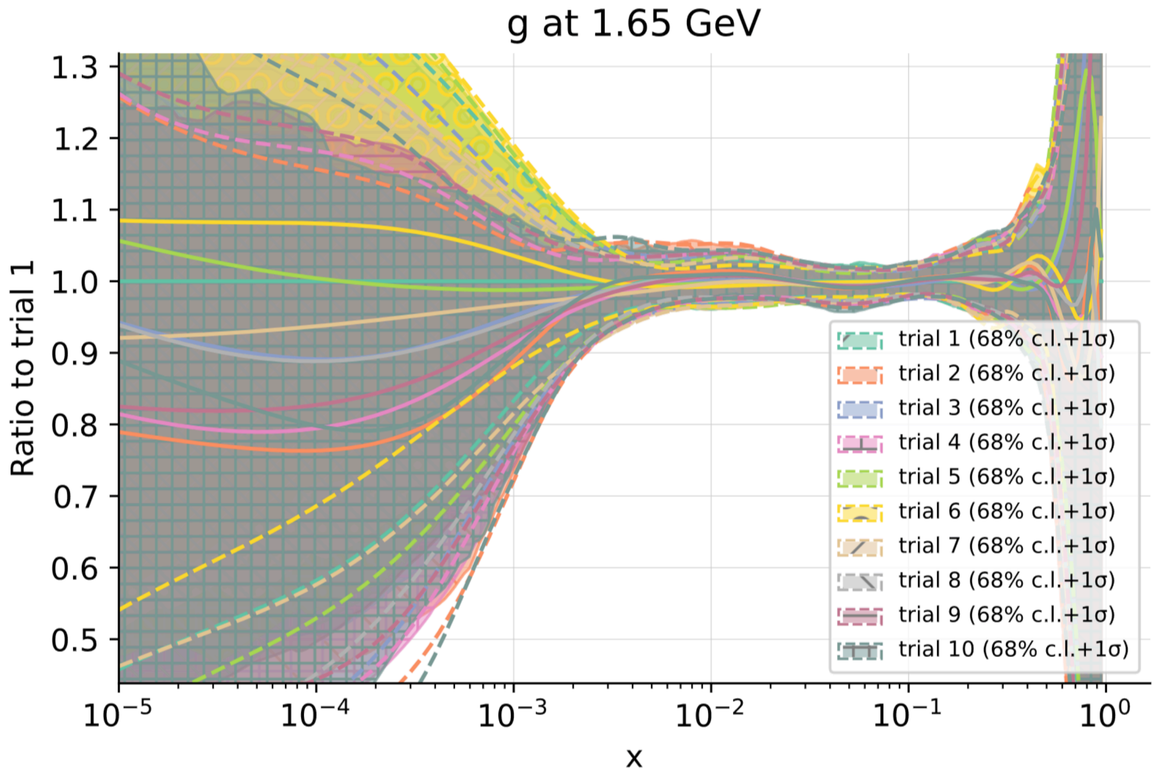
$$\{\theta_i, i = 1,..., N_{trials}\}$$

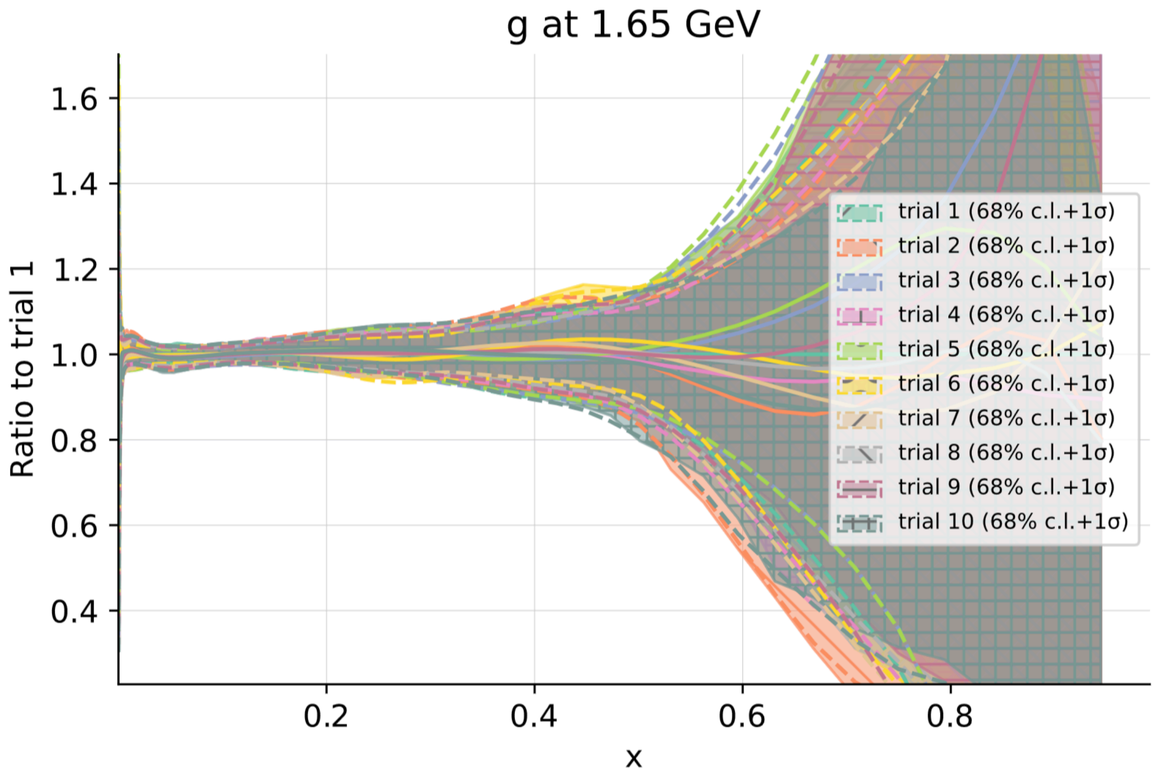


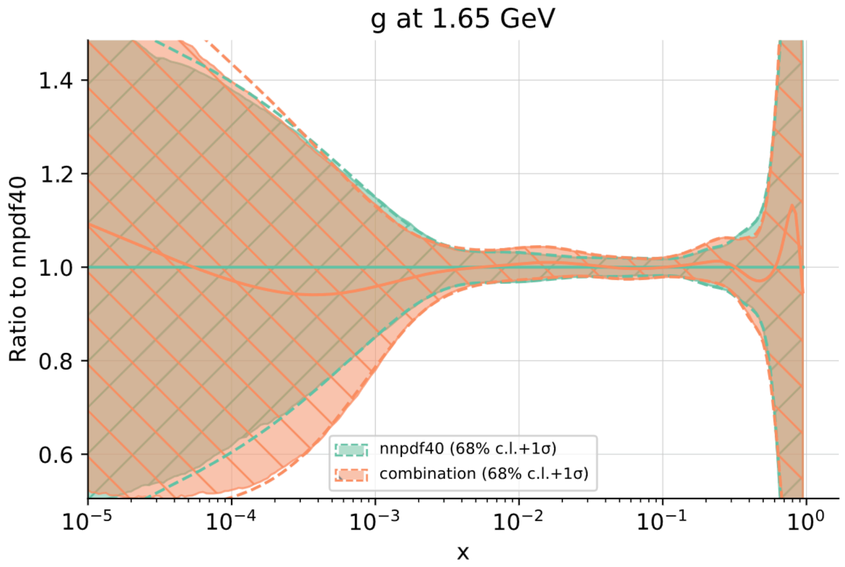
- We use **hyperopt library** [Bergstra, J., Yamins, D., Cox, D. D. (2013) Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures.]

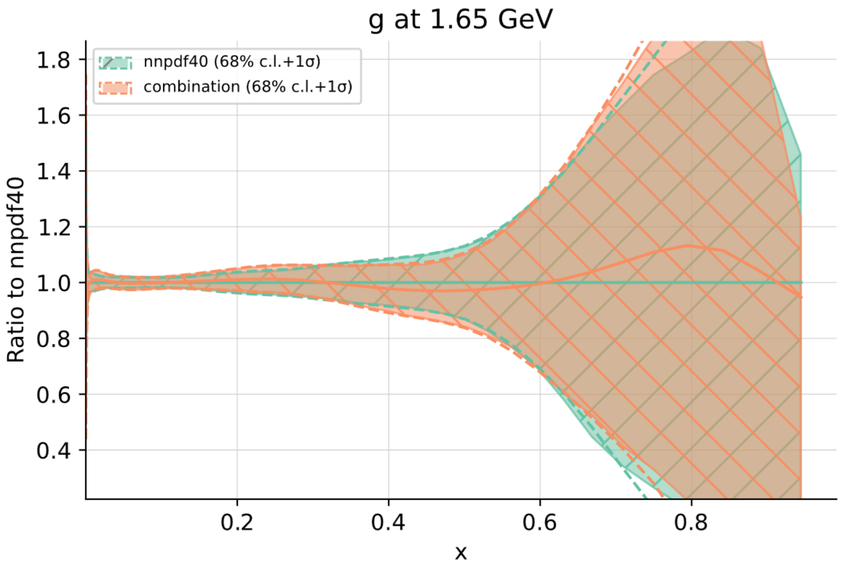
Trial	Architecture	Activation	Optimizer	Learning rate	Clipnorm	Epochs
925	[15, 16, 19, 8]	tanh	Adam	0.00618	5.98e-7	18704
550	[13, 16, 18, 8]	tanh	Adam	0.00816	4.67e-7	20089
870	[15, 16, 18, 8]	tanh	Adam	0.00785	6.06e-7	19298
760	[22, 16, 16, 8]	tanh	Adam	0.00763	8.96e-7	17993
859	[15, 17, 17, 8]	tanh	Adam	0.00850	4.97e-7	19303
463	[12, 16, 19, 8]	tanh	Adam	0.00979	1.93e-6	18683
729	[19, 17, 17, 8]	tanh	Adam	0.00695	6.21e-7	18880
666	[12, 12, 17, 8]	tanh	Adam	0.00491	8.02e-7	22204
898	[24, 15, 19, 8]	tanh	Adam	0.00891	4.65e-7	19564
438	[14, 15, 19, 8]	sigmoid	Adam	0.01000	1.95e-6	19737

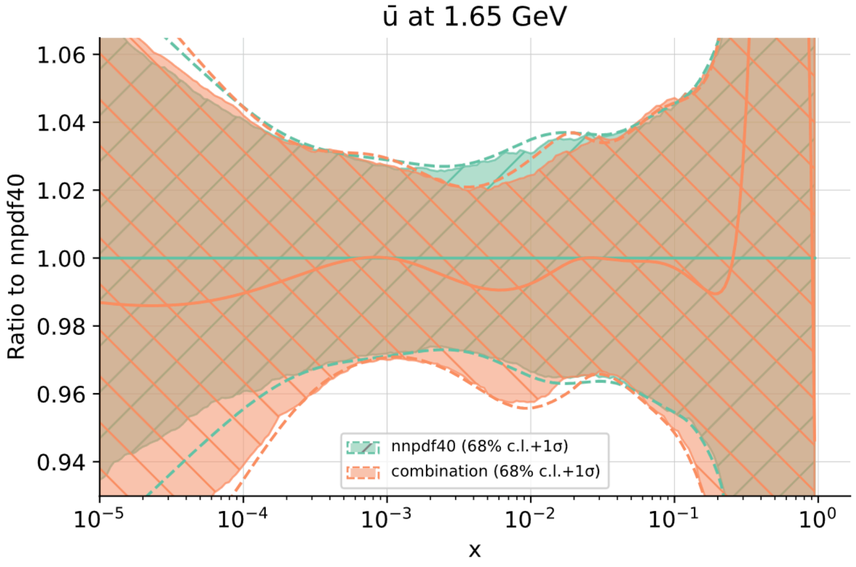
- 10 best trials characterised by **different architecture, optimisers, etc.**,
- the best trials yield PDF distributions that are in **excellent agreement** with the **differences well within the uncertainties**
- Spread of central values in some regions: bigger PDF uncertainty upon combination

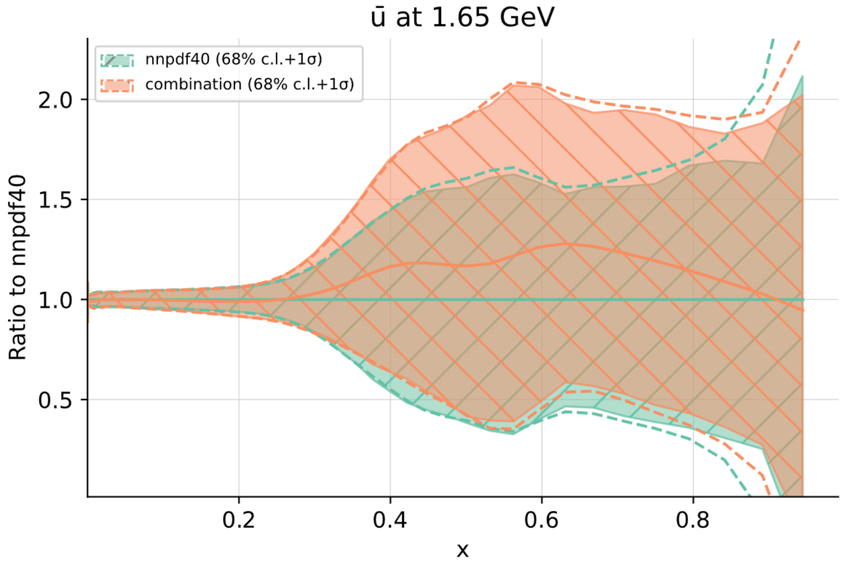












- Samples from the best trials are used to construct a **combined** PDF distribution
- The combined results are **consistent with the published baseline NNPDF4.0** but with **slightly larger** uncertainties (in the extrapolation regions)

GPU parallelisation



- Monte Carlo (MC) PDF fits require **hundreds** to **thousands** of independent replica trainings
- Each fit is **computationally expensive**: theory evaluations, χ^2 minimisation, NN training
- New hyperoptimization **multiplies the total number of required runs**



- Replica fits are **naturally parallel** and ideal for distributed computing
- GPUs accelerate training
- Faster turnaround enables rapid methodology development and validation studies
- Precision PDF physics i**ncreasingly depends on advanced computing** infrastructure

Summary

- **NNPDF framework**: public fitting framework for PDF determination, based on NN parameterisation and Monte Carlo replicas approach
- Monte Carlo (MC) PDF extractions requires large ensembles of independent fits for robust uncertainty propagation
- **New hyperoptimization algorithm based on ensemble regression**: reduce methodological bias, using a a metric based on full likelihood and including in final PDF error effect due to hyeprparameter variation but
- New hyperopt further increases the computational cost: GPUs and HPC systems provide the parallelism and scalability needed to make these workflows practical

NNPDF

All tests NNPDF passing | Documentation passing | Test CommonData passing

Eur.Phys.J.C 81 (2021) 10 | DOI 10.5281/zenodo.15691478 | HSF Affiliated

NNPDF: An open-source machine learning framework for global analyses of parton distributions

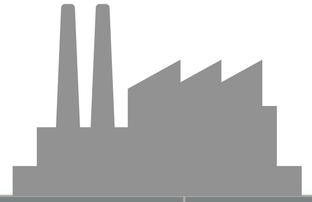
Github: https://github.com/NNPDF/nnpdf

Documentation: https://docs.nnpdf.science/



Backup slides

Different approaches and releases



	NNPDF4.0	MSHT20	CT18	ABMP16
Released	*Sept 2021*: LHAPDF grids + **fitting code**	*Dec 2020*: LHAPDF grids	*Dec 2019*: LHAPDF grids	*Jan 2017*: LHAPDF grids
Parametrisation	Neural networks (**hyperoptimised**)	Functional form + Chebyshev	Functional form + Bernstein	Functional form
Error estimate	Monte Carlo (**closure + future tested**)	Hessian (dynamic tolerance)	Hessian (dynamic tolerance) + Lagrange mult.	Hessian (no tolerance)
Theory settings	NNLO QCD, GM-VFN (+ NLO electroweak)	NNLO QCD GM-VFN	NNLO QCD GM-VFN	NNLO QCD, FFN

What is the role of θ in central values and uncertainties?

$$P\left(f \mid D\right) \rightarrow P\left(f \mid D, \theta\right)$$

What is the role of θ in central values and uncertainties?

$$P\left(f \mid D\right) \rightarrow P\left(f \mid D, \theta\right)$$

Let s assume that hyperparameters θ are distributed according to probability distribution $\eta\left(\theta\right)$

What is the role of θ in central values and uncertainties?

$$P\left(f \mid D\right) \rightarrow P\left(f \mid D, \theta\right)$$

Let s assume that hyperparameters θ are distributed according to probability distribution $\eta\left(\theta\right)$

$$\langle \mathcal{O} \rangle_{f,\theta} = \int d\theta \, Df \; \mathcal{O}(f) \, P\left(f \mid D, \theta\right) \eta\left(\theta\right) = \int d\theta \; \overline{\mathcal{O}}\left(\theta\right) \;\; \eta\left(\theta\right)$$

$$\langle \left(\mathcal{O} - \langle \mathcal{O} \rangle_{f,\theta} \right)^2 \rangle_{f,\theta} = \int d\theta \, \eta\left(\theta\right) \left(K\left(\theta\right) + \left(\overline{\mathcal{O}}\left(\theta\right) - \langle \overline{\mathcal{O}} \rangle_{\theta} \right)^2 \right)$$

What is the role of θ in central values and uncertainties?

$$P\left(f \mid D\right) \rightarrow P\left(f \mid D, \theta\right)$$

Let s assume that hyperparameters θ are distributed according to probability distribution $\eta\left(\theta\right)$

Average of the mean value on hyperparamters distributions

$$\langle \mathcal{O} \rangle_{f,\theta} = \int d\theta \, Df \; \mathcal{O}(f) \, P\left(f \mid D, \theta\right) \eta\left(\theta\right) = \int d\theta \, \overline{\mathcal{O}}\left(\theta\right) \; \eta\left(\theta\right)$$

$$\langle \left(\mathcal{O} - \langle \mathcal{O} \rangle_{f,\theta} \right)^2 \rangle_{f,\theta} = \int d\theta \, \eta\left(\theta\right) \left(K\left(\theta\right) + \left(\overline{\mathcal{O}}\left(\theta\right) - \langle \overline{\mathcal{O}} \rangle_{\theta} \right)^2 \right)$$

What is the role of θ in central values and uncertainties?

$$P\left(f \mid D\right) \rightarrow P\left(f \mid D, \theta\right)$$

Let s assume that hyperparameters θ are distributed according to probability distribution $\eta\left(\theta\right)$

$$\langle \mathcal{O} \rangle_{f,\theta} = \int d\theta \, Df \; \mathcal{O}(f) \, P\left(f \mid D, \theta\right) \eta\left(\theta\right) = \int d\theta \; \overline{\mathcal{O}}\left(\theta\right) \;\; \eta\left(\theta\right)$$

Average of the mean value on hyperparamters distributions

$$\langle \left(\mathcal{O} - \langle \mathcal{O} \rangle_{f,\theta} \right)^2 \rangle_{f,\theta} = \int d\theta \, \eta\left(\theta\right) \left(K\left(\theta\right) + \left(\overline{\mathcal{O}}\left(\theta\right) - \langle \overline{\mathcal{O}} \rangle_{\theta} \right)^2 \right)$$

PDF error with fixed hyperparameter values

Spread of central values when varying hyperparameters

Likelihood of the excluded fold

$$p_k(\theta) = \frac{1}{N(\theta)} \exp\left[-\frac{1}{2}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)\left(C_y^{(k)} + C_{PDF}^{(k)}(\theta)\right)^{-1}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)^T\right]$$

$$L^{(k)}(\theta) = -2\log p_k(\theta) = \log\det\left(C_y^{(k)} + C_{\text{PDF}}^{(k)}(\theta)\right) + \left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)\left(C_y^{(k)} + C_{PDF}^{(k)}(\theta)\right)^{-1}\left(y^{(k)} - \overline{T}^{(k)}(\theta)\right)^T$$

Penalty term

Chi2 of the excluded fold, including PDF error

New NNPDF4.1 metric

$$L_{4.1}(\theta) = \frac{1}{k}\sum_i^k L^{(i)}(\theta)$$