Centre of Excellence in Quark Matter

NNPDF

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

Research Council of Finland
HELSINKI INSTITUTE OF PHYSICS

Funded by the European Union

erc

Towards a new NNPDF global analysis

Felix Hekhorn

DIS 2026

Overview

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

1. Introduction
2. Theory
3. Experimental data
4. Methodology
5. Summary

Introduction

NNPDF is open source! [EPJC81.958]

https://github.com/NNPDF/

https://github.com/NNPDF/nnpdf

/nnpdf_data: all experimental data
/n3fit: core fitting methodology
/validphys2: (overarching) methodology

https://github.com/NNPDF/pineappl

Fast interpolation grid library

https://github.com/NNPDF/eko

DGLAP solver

NNPDF timeline

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

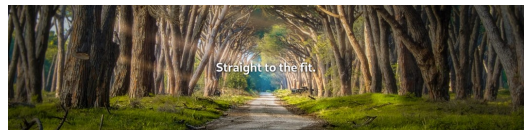
- 2021-09 NNPDF4.0 [EPJC82.428]
- 2022-08 Intrinsic charm [Nature608.483]
- 2023-11 Intrinsic valence charm [PRD109.L091501]
- 2024-01 MHOU [EPJC84.517]
- 2024-01 QED [EPJC84.540]
- 2024-02 aN^{3}LO [EPJC84.659]
- 2024-06 Monte Carlo PDF [JHEP09.088]
- 2024-10 GPU-based hyperoptimization [MLST6.025027]
- 2024-11 aN^{3}LO combination [JPG52.065002]
- 2025-01 LHC data study [JHEP07.067]
- 2025-06 α_s determination [EPJC85.1001]

$\Rightarrow$ 2026(?) autumn(?) NNPDF4.1

Theory

New Theory Prediction ~~Pipeline~~ Pineline [CPC297.109061]

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ



Madgraph

vrap

yadism

top++

Matrix

NNLOJET

PineAPPL

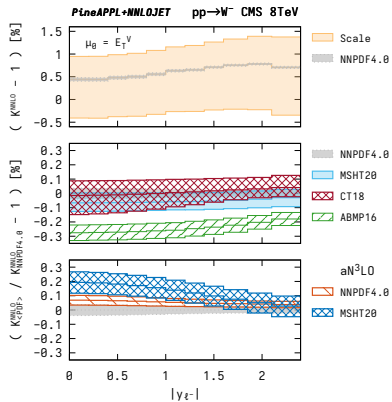
EKO
Evolution Kernel Operators

https://github.com/NNPDF

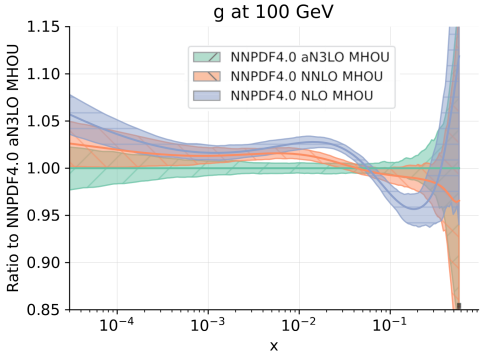
Exact NNLO grids

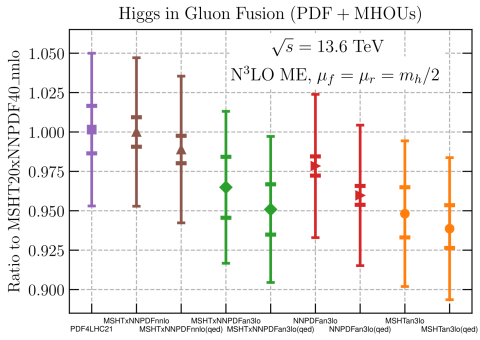
For example DY from [EPJC85.459]:

- Use exact NNLO grids for all hadronic processes and N3LO for DIS
- K-factors are computed with a fixed PDF (typically not the one which is fitted)



N^3LO QCD $\otimes$ NLO QED evolution [EPJC84.659] [JPG52.065002]





Plus all recent updates!

Theory setup

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

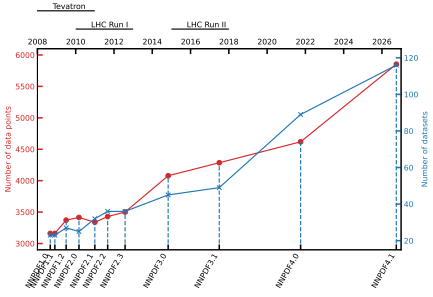
- account for missing higher order uncertainties (MHOU) [EPJC84.517] and other theory uncertainties
- update scale choices for some datasets (e.g. DY: $M_V \to E_t$)
- update treatment of EW corrections
- update input parameters, e.g. quark masses, electro-weak parameters, CKM, etc.

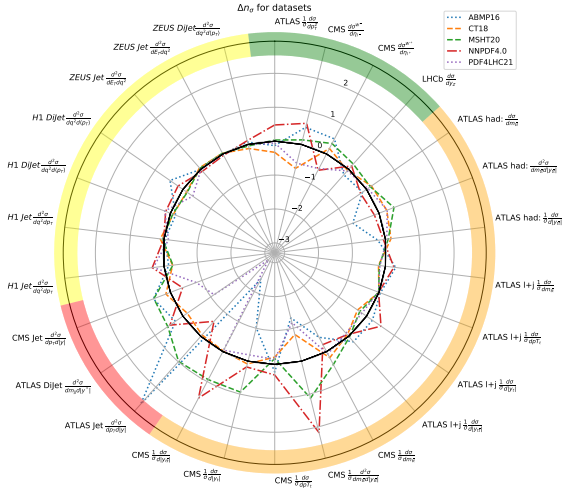
$\Rightarrow$ use best available theory

Experimental data

Overview experimental data

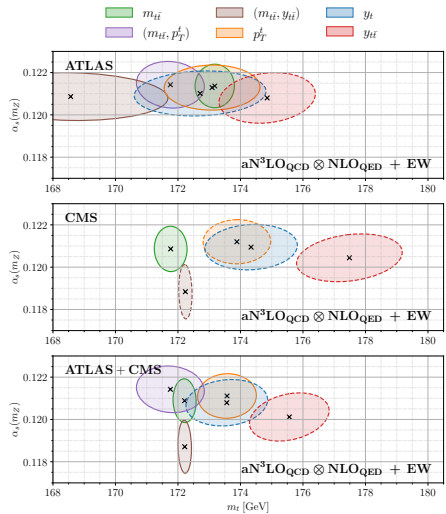
- full rework of implementation
- all data from [JHEP07.067] + more
- published < 2025-01-01
- other data → validation





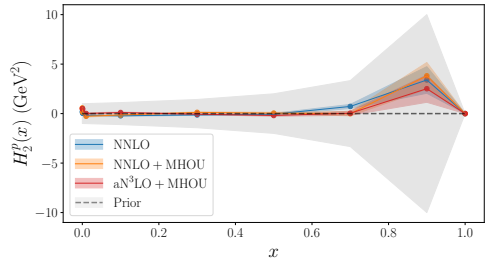
Top data [2603.28865]

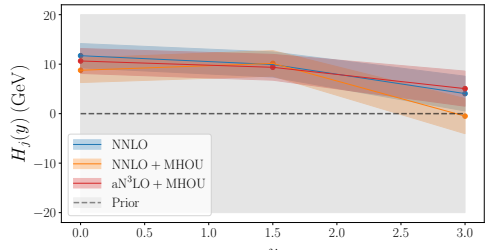
- [2603.28865] extracts $\alpha_s + m_t$
  $\rightarrow$Jaco ter Hoeve, Wed. 17:05
- the default NNPDF fit will assume a given $\alpha_s(m_Z) = 0.118$ and $m_t = 172.5\,\text{GeV}$
- top measurements come as different observables: $p_T, y_t, y_{t\bar{t}}, m_{t\bar{t}}$
- choosing a “good” set is non-trivial and may be different for extraction of PDF or m_t



Higher twist and power corrections [EPJC86.281]

- [EPJC86.281] extracts HT/PC
- Identify contaminated kinematic regions and either deweight them or cut them away
  - (fully inclusive) NC and CC DIS
  - single jets and dijets





Dataset selection [2503.17447]

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

Lessons learnt from closure testing with artificially underestimated uncertainties:

- the NN can correct for moderate uncertainty underestimation
- Update dataset selection:
  - Perform a weighted fit for each datasets
  - *IF* the (weighted) χ^2 for that dataset deviates too much *OR* the χ^2 of any other dataset deviates too much from it's "normal" value
    - Flag dataset as inconsistent

Methodology

Hyperoptimization on GPU [MLST6.025027]

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

# Replicas	10	50	100
Energy reduction	78%	87%	91%
Cost reduction	−45%	47%	55%

NVIDIA H100 GPU vs 16 AMD EPYC Genoa CPU on SURF's SNELLIUS cluster

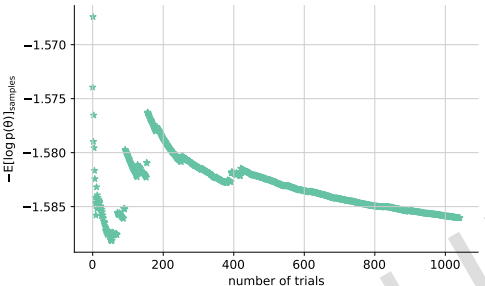
Technical changes:

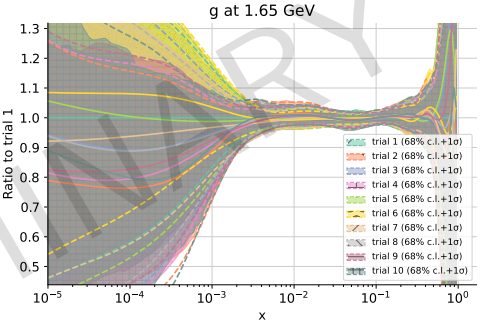
- Single NN model for all samples
- Share memory-heavy objects
- Single hyperopt database shared by GPUs

Results:

- Memory usage scales only weakly with number of replicas, enabling a 100 replica fit in a single GPU
- 90% energy reduction: faster and more affordable fits!

More hyperoptimization





NNPDF1.0:
manual hyperopt
1 architecture

→

NNPDF4.0:
hyperopt with 1 replica
1 architecture

→

NNPDF4.1:
hyperopt with ensemble
10 architectures

Other methodological improvements

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

- Parametrize charm and anti-charm independently [Nature608.483] [PRD109.L091501]
- Impose PDF positivity at $Q = 100$ GeV during the fit $\rightarrow$ sufficient to ensure positivity of physical observables [EPJC84.335]
- No pre-processing needed at large-x [EPJC82.163]

Summary

Towards NNPDF4.1

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

- theory:
  - use exact NNLO grids (or better if we can)
  - use N^{3}LO QCD $\otimes$ NLO QED evolution
- data:
  - include all available data, which appeared before 2025
  - use data after cut-off date for validation [APPB52.243]
  - account for power corrections/higher twist contamination [EPJC86.281]
- methodology:
  - use refined hyperoptimization strategy
  - parametrize c and $\bar{c}$
  - use closure testing [2503.17447]

Towards NNPDF4.1

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

- theory:
  - use exact NNLO grids (or better if we can)
  - use N^{3}LO QCD $\otimes$ NLO QED evolution
- data:
  - include all available data, which appeared before 2025
  - use data after cut-off date for validation [APPB52.243]
  - account for power corrections/higher twist contamination [EPJC86.281]
- methodology:
  - use refined hyperoptimization strategy
  - parametrize c and $\bar{c}$
  - use closure testing [2503.17447]

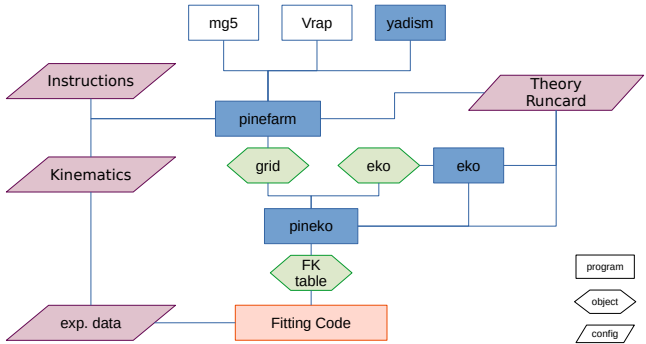
Danke! Thanks! Kiitos!

Backup slides

New Theory Prediction ~~Pipeline~~ Pineline [CPC297.109061]

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

Produce FastKernel (FK) tables!



The workhorse in the background: PineAPPL

ATLAS 8TeV Z production [JHEP07.067]

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

Dataset	$n_{\rm dat}$	$\sqrt{2/n_{\rm dat}}$		*ABMP16*	*CT18*	*CT18A*	*CT18Z*	*MSHT20*	*NNPDF3.1*	*NNPDF4.0*	*PDF4LHC15*	*PDF4LHC21*
ATLAS 8 TeV Z $\frac{d\sigma}{d\lvert y\rvert}$	7	0.53	$\chi^2_{\rm exp+th}$	4.25	1.52	1.52	1.18	1.37	1.61	3.83	1.23	1.09
			$\chi^2_{\rm exp}$	7.36	14.0	4.63	4.31	2.14	4.70	7.90	7.41	1.93

Table 4. Same as table 3 for the ATLAS Drell-Yan gauge boson production measurements at the LHC 8 TeV [42].

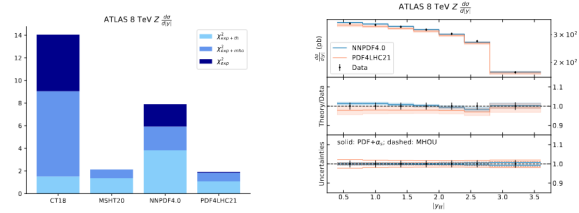
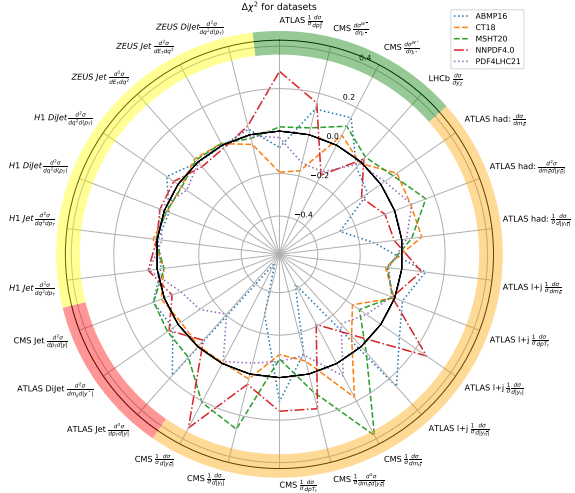


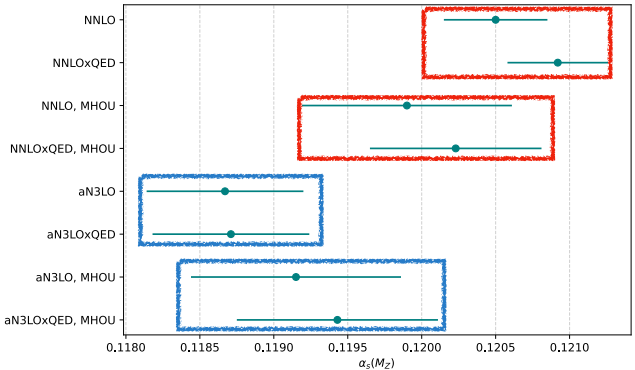
Figure 4. Same as figures 2 (left) and 3 (right) for the ATLAS Drell-Yan gauge boson production measurements at the LHC 8 TeV [42] ($n_{\rm dat} = 7$, $\sqrt{2/n_{\rm dat}} = 0.53$).

Generalization to unseen data [JHEP07.067]

$$\Delta\chi^{2(i)} = \frac{\chi^{2(i)}_{\text{exp+th}} - \left\langle \chi^{2}_{\text{exp+th}} \right\rangle_{\text{pdfs}}}{\left\langle \chi^{2}_{\text{exp+th}} \right\rangle_{\text{pdfs}}}$$



α_s at more PTO [EPJC85.1001]

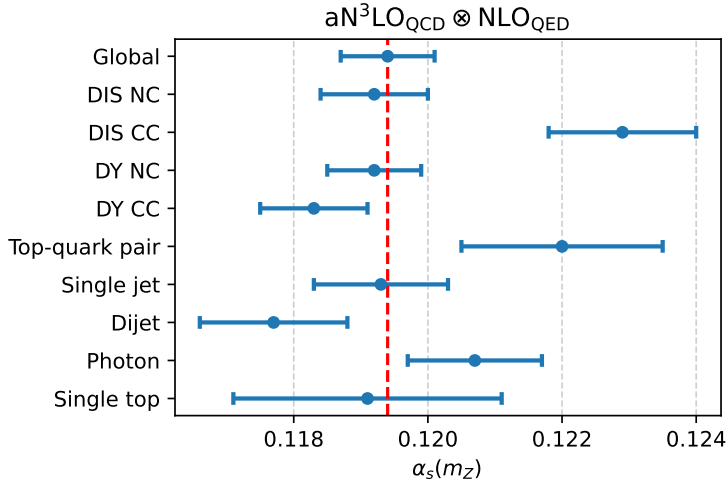


QED means
- NLO QED corrections to DGLAP evolution
- Determine also the photon PDF

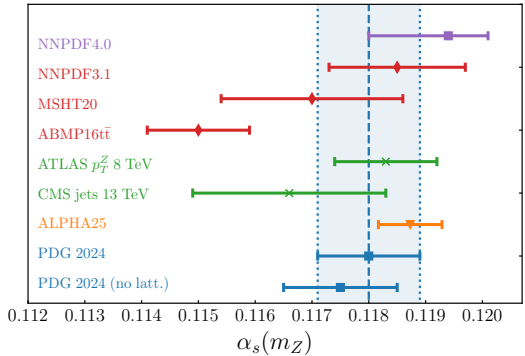
QED has a bigger impact at **NNLO** than at **aN3LO**

$$\alpha_s(M_Z)^{aN^3LO,QED,MHOU} = 0.1194^{+0.0007}_{-0.0014}$$

α_s by process [EPJC85.1001]



α_s to others [EPJC85.1001]



Positivity