MARIA UBIALI
UNIVERSITY OF CAMBRIDGE

ROBUST PDF DETERMINATION FOR PRECISE ELECTROWEAK MEASUREMENTS

European Research Council
Established by the European Commission

OUTLINE

- Introduction

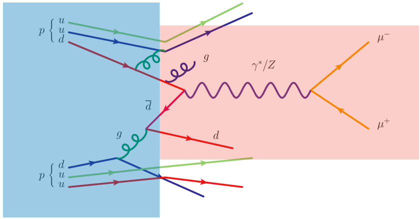
- Towards 1% precision in PDFs
  - ➡ Progress in PDF determination
  - ➡ Experimental data & Methodological advances
  - ➡ PDFs and M_W determination

- Beyond the state of the art
  - ➡ Theory: missing higher order uncertainties and N3LO PDFs
  - ➡ Simultaneous fits and interplay between PDFs and indirect new physics searches in high mass tails

- Conclusions and outlook

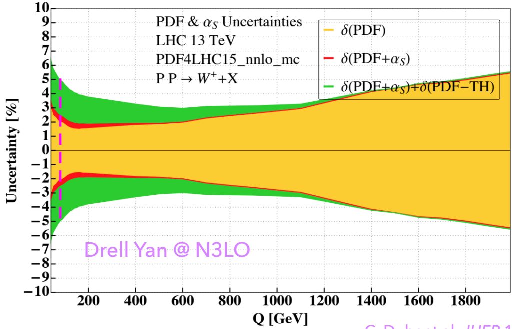
INTRODUCTION

WHY ARE PDFS SO IMPORTANT FOR PRECISION PHYSICS ...



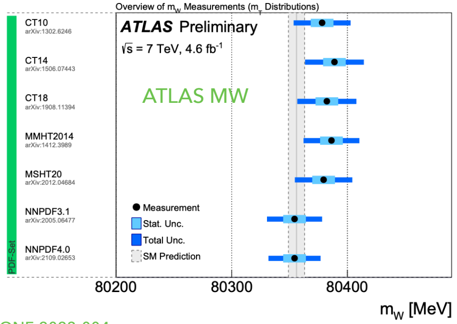
$$d\sigma^{pp\to ab} = \sum_{i,j} f_i \otimes f_j \otimes d\hat{\sigma}^{ij\to ab} + \ldots$$

#1: Theory uncertainty of SM predictions



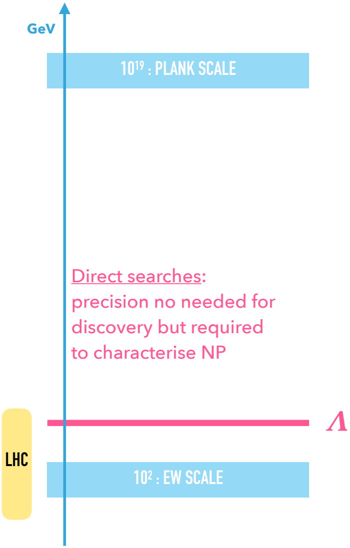
C. Duhr et al, *JHEP* 11 (2020) 143

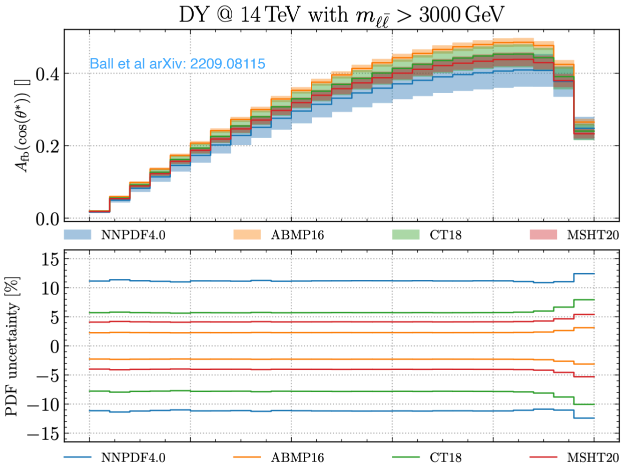
#2: Determination of SM parameters



ATLAS-CONF-2023-004

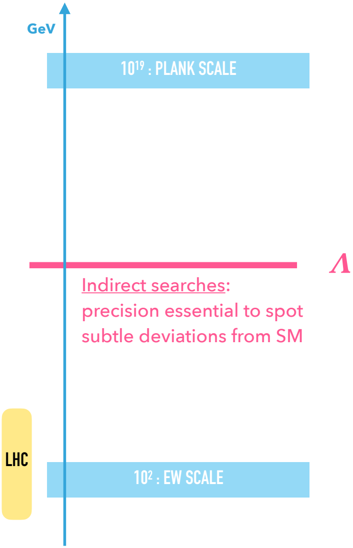
. . . AND FOR NEW PHYSICS SEARCHES

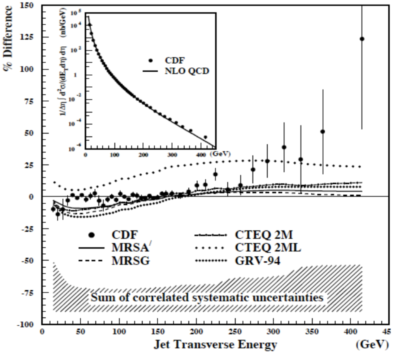


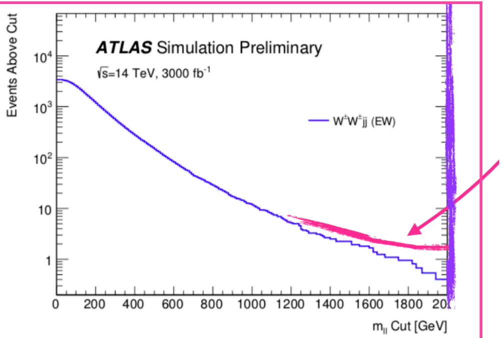


#1: Direct searches: heavy Z′ search in DY Forward-Backward asymmetry

. . . AND FOR NEW PHYSICS SEARCHES







CA Lee, HL/HE-LHC Jamboree, 1 March 2019

Discrepancy betweenQCD calculation and CDF jets data (1995)

At that time no information on PDF uncertainties and theory predictions strongly depended on gluon shape at $x>0.1$. Once data included in the CTEQ fit, discrepancy disappeared.

Deviations from SM predictions in high energy tails (>2023)

New physics or limited understanding of proton structure?

#2: Indirect searches

TOWARDS 1% PRECISION IN PDF DETERMINATION

DETERMINATION OF PDFS

$$f_i(x, \mu)$$

Perturbative QCD

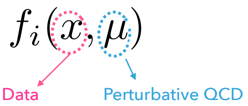
- DGLAP evolution equations fully known up to NNLO and partially at N3LO
- Most theory predictions for processes in PDF fits known at NNLO (some at N3LO)

[Moch, Vermaseren, Vogt (2004)]

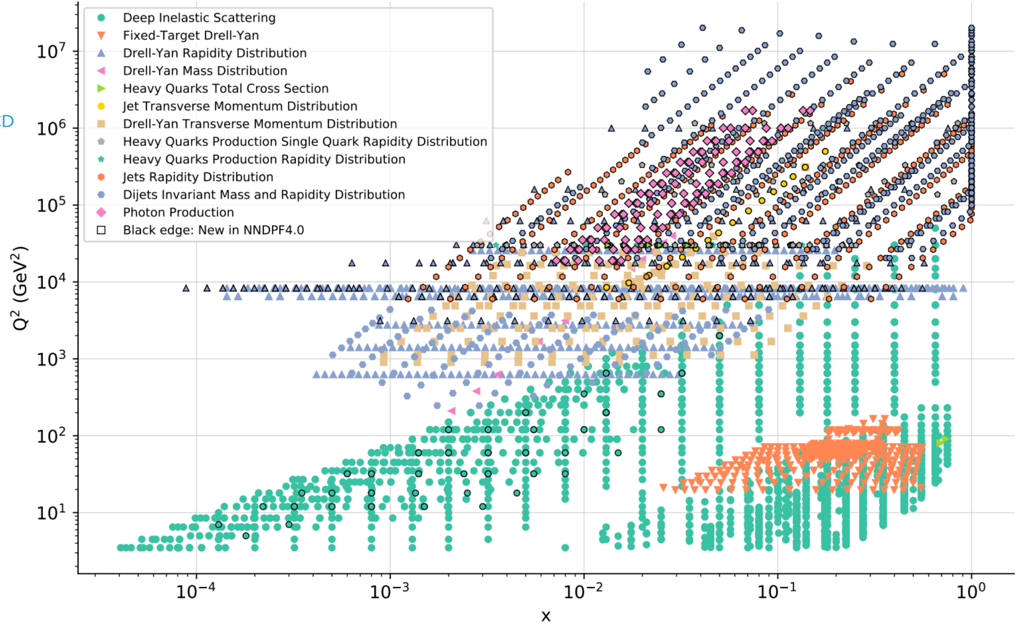
[Herzog, Moch, Vermaseren, Vogt, Bonvini…]

DETERMINATION OF PDFS

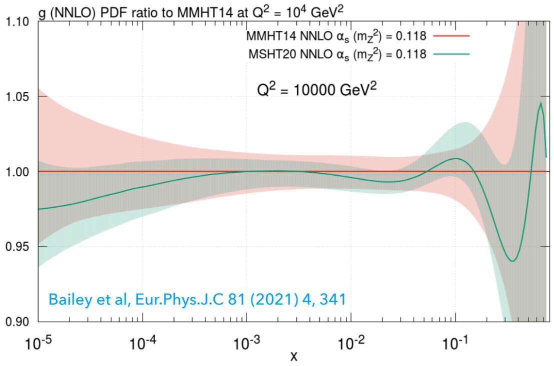
$$\chi^2 = \sum_{i,j=1}^{N_{\rm dat}} (D_i - T_i(\theta_{\rm PDF}))(\mathrm{cov})^{-1}_{ij}(D_j - T_j(\theta_{\rm PDF}))$$

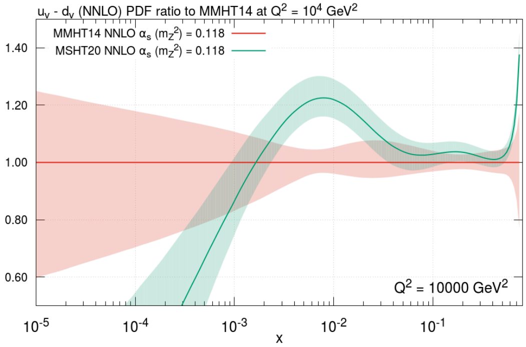


- DGLAP evolution equations fully known up to NNLO and partially at N3LO
- Most theory predictions for processes in PDF fits known at NNLO (some at N3LO)
- Precision of the most data of the order of a few percents
- Mostly from correlated systematic uncertainties

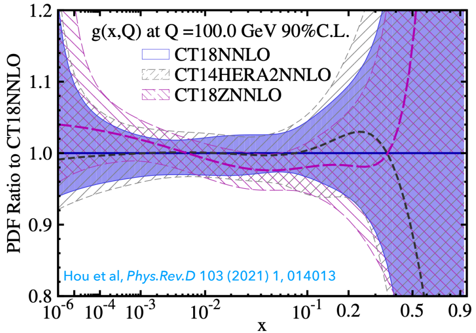


A LOT OF RECENT PROGRESS

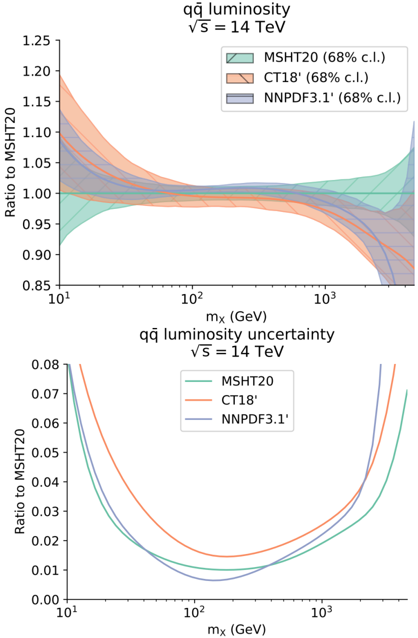




- In recent updates from global PDF fitting collaborations (NNPDF3.1, CT18, MSHT20) the effect of LHC data is driving the PDF uncertainties down, although highly correlated data provide big challenge
- The increased precision of the data and their strong correlation demands methodological improvements (e. g. more flexible parametrisation, refined statistical techniques)
- Some tension among data observed (CT18 vs CT18Z)

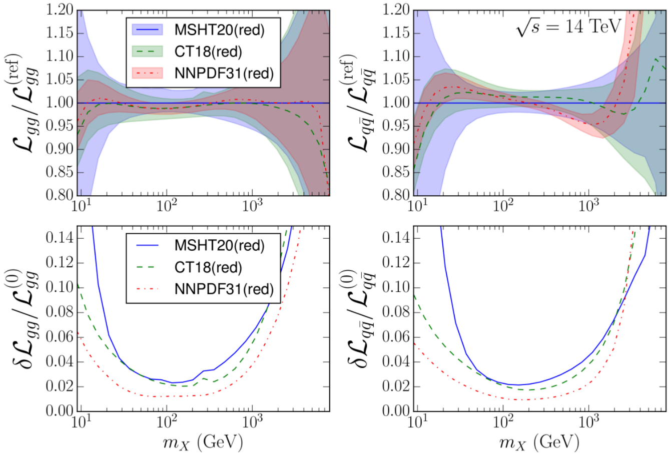


PDF4LHC21 EXERCISE



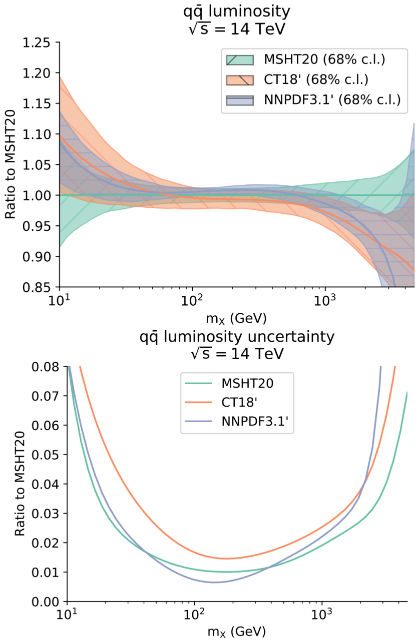
PDF4LHC WG, arXiv: 2203.05506

- Benchmark among NNPDF3.1, MSHT20 and CT18
- Overall agreement, which improves once common dataset is used, differences in uncertainties with ΔCT ≳ ΔMHST ≳ ΔNN

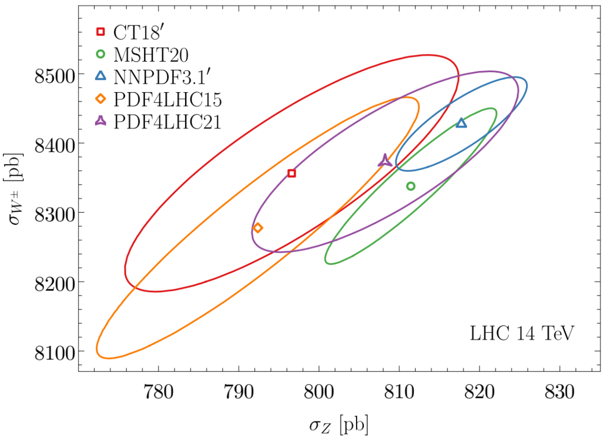


PDF4LHC21 EXERCISE

PDF4LHC WG, arXiv: 2203.05506



- Benchmark among NNPDF3.1, MSHT20 and CT18
- Overall agreement, which improves once common dataset is used, differences in uncertainties with ΔCT ≳ ΔMHST ≳ ΔNN



NNPDF4.0: THE ROLE OF LHC DATA

Ball et al, arXi: 2109.02653

Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
ATLAS W, Z 7 TeV (2010)	✓	✓	✓	✓	✓
ATLAS W, Z 7 TeV (2011)	✓	✓	✗	✓	✓
ATLAS low-mass DY 7 TeV	✓	✓	✗	✗	✗
ATLAS high-mass DY 7 TeV	✓	✓	✗	✗	✓
ATLAS W 8 TeV	✓	✗	✗	✗	✓
ATLAS DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS high-mass DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS $\sigma_{W,Z}$ 13 TeV	✓	✗	✓	✗	✗
ATLAS W^++jet 8 TeV	✓	✗	✗	✗	✓
ATLAS Z p_T 8 TeV	✓	✓	✗	✓	✓
ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	✓	✓	✓	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV	✓	✓	✓	✗	✗
ATLAS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	✓	✗	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, R=0.6	✗	✓	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS dijets 7 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS direct photon production 13 TeV	✓	✗	✗	✗	✗
ATLAS single top R_t 7, 8, 13 TeV	✓	✗	✓	✗	✗
ATLAS single top diff. 7, 8 TeV	✓	✗	✗	✗	✗
ATLAS single top diff. 8 TeV	✓	✗	✗	✗	✗

Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
LHCb Z 940 pb	✓	✓	✗	✗	✓
LHCb $Z \to ee$ 2 fb	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 7 TeV	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 8 TeV	✓	✓	✓	✓	✓
LHCb $Z \to \mu\mu, ee$ 13 TeV	✓	✗	✗	✗	✗

Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
CMS W electron asymmetry 7 TeV	✓	✓	✗	✓	✓
CMS W muon asymmetry 7 TeV	✓	✓	✓	✓	✗
CMS Drell-Yan 2D 7 TeV	✓	✓	✗	✗	✓
CMS W rapidity 8 TeV	✓	✓	✓	✓	✓
CMS Z p_T 8 TeV	✓	✓	✗	✓	✗
CMS $W+c$ 7 TeV	✓	✓	✗	✗	✓
CMS $W+c$ 13 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 2.76 TeV	✗	✓	✗	✗	✓
CMS single-inclusive jets 7 TeV	✗	✓	✗	✓	✓
CMS dijets 7 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 8 TeV	✗	✗	✗	✓	✓
CMS 3D dijets 8 TeV	✓	✗	✗	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 5 TeV	✓	✗	✓	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	✓	✓	✓	✗	✓
CMS $\sigma_{tt}^{\rm tot}$ 13 TeV	✓	✓	✓	✗	✗
CMS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✗	✓
CMS $t\bar{t}$ 2D dilepton 8 TeV	✓	✗	✗	✓	✓
CMS $t\bar{t}$ lepton+jet 13 TeV	✓	✗	✗	✗	✗
CMS $t\bar{t}$ dilepton 13 TeV	✓	✗	✗	✗	✗
CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV	✓	✗	✓	✗	✗
CMS single top R_t 8, 13 TeV	✓	✗	✓	✗	✗

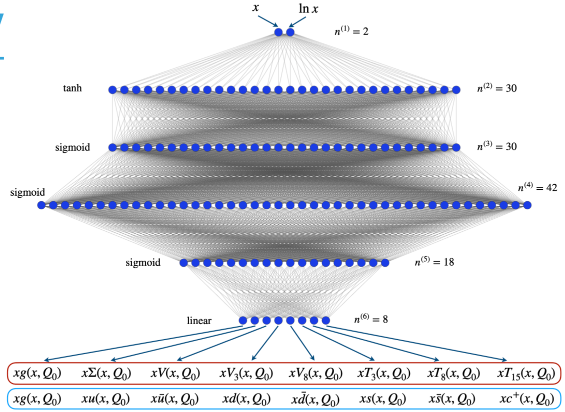
- NNPDF4.0 based on larger set of data from LHC Run I and Run II
- O(4500) data points and first time inclusion of di-jets preferred over inclusive jets based on perturbative stability and impact on gluon [Khalek et al, 2005.11327] j
- See also talk by L. Harland-Lang at DIS2023

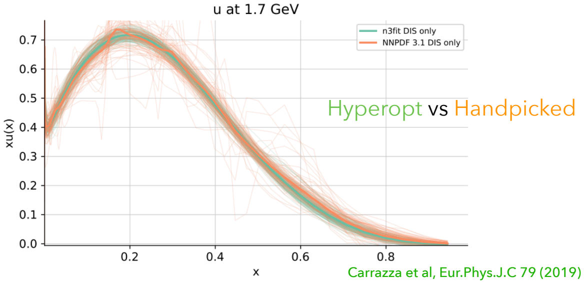
NNPDF4.0: THE ROLE OF METHODOLOGY

The NNPDF4.0 methodology

- Single neural network to parametrise 8 independent PDF combinations (g, u, d, s, u~, d~, s~, c=c~)
- New optimisation strategy based on gradient descent rather than genetic algorithm
- Hyper-optimised methodology: scan of the hyper parameter space to find optimal minimisation settings (optimiser, initialiser, stopping patience, number of layers, learning rate, epochs, activation function) by minimising $\chi^2_{\rm val}$
  [Carrazza et al, Eur.Phys.J.C 79 (2019) 8, 676]
- Statistical validation of PDF uncertainties via closure tests (data region)
  [Del Debbio et al, Eur.Phys.J.C 82 (2022) 4, 330]
  and future test (extrapolation region)
  [J. Cruz-Martinez et al, Acta Phys.Polon.B 52 (2021) 243]

Ball et al, arXi: 2109.02653

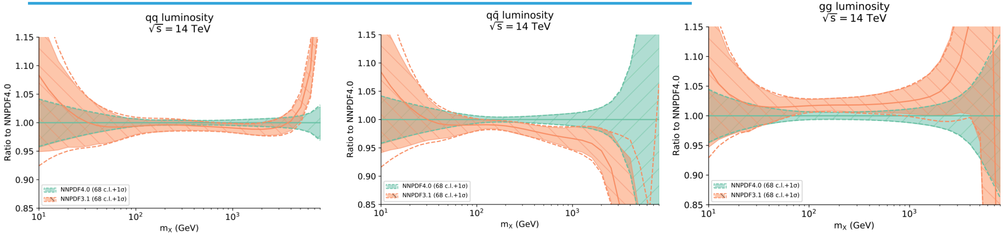




Carrazza et al, Eur.Phys.J.C 79 (2019) 8, 676

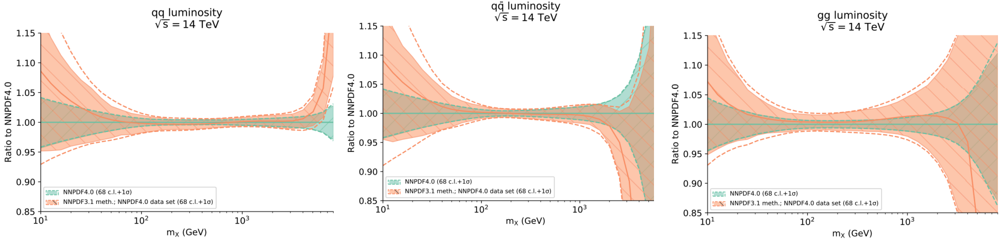
NNPDF4.0: THE ROLE OF METHODOLOGY

NNPDF3.1 vs NNPDF4.0



DATA \ METHODOLOGY	NNPDF3.1	NNPDF4.0
NNPDF3.1 (4093)	1.19	1.12
NNPDF4.0 (4491)	1.25	1.17

NNPDF4.0 data - NNPDF3.1 vs NNPDF4.0 methodology



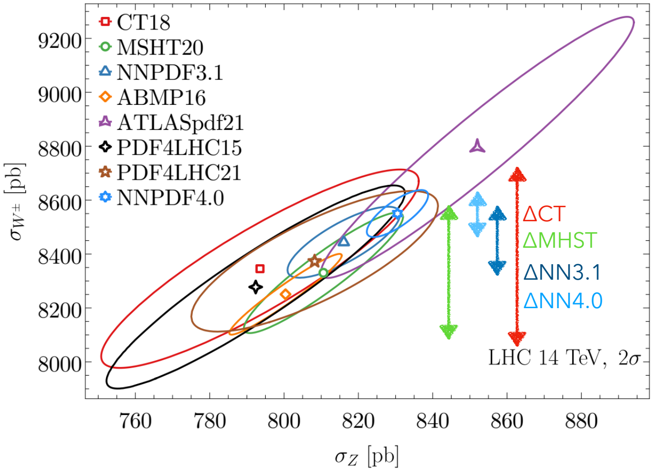
DATA \ METHODOLOGY	NNPDF3.1	NNPDF4.0
NNPDF3.1 (4093)	1.19	1.12
NNPDF4.0 (4491)	1.25	1.17

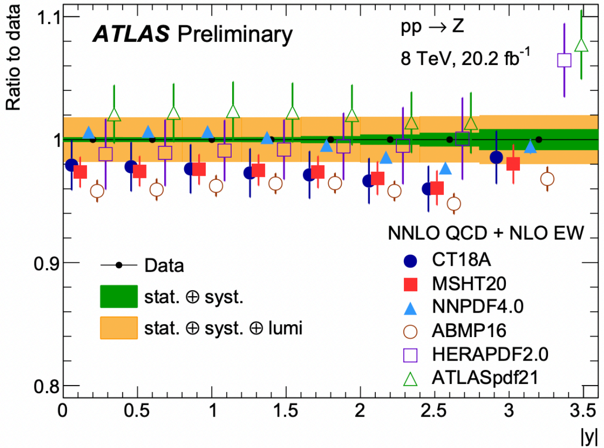
Ball et al, arXi: 2109.02653

- Shift in parton luminosities mostly due to inclusion of O(500) more data points
- Parton luminosities based on same dataset are consistent with each other but 4.0 methodology displays smaller uncertainty than 3.1 methodology: NNPDF4.0 more accurate and superior to previous methodology

NNPDF4.0: PHENOMENOLOGICAL IMPLICATIONS

S. Camarda's talk



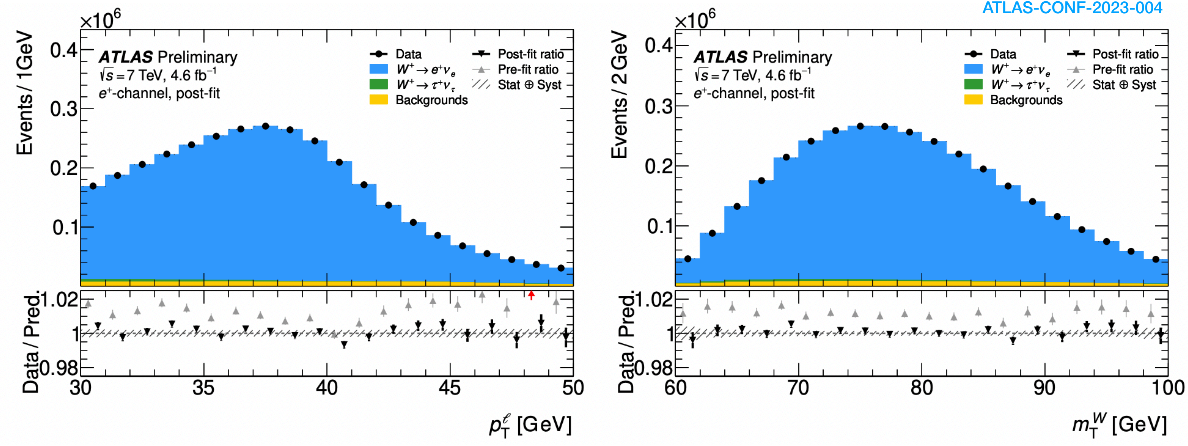


Snowmass review - arXiv:2211.12961

ATLAS-CONF-2023-013

ROLE OF PDFS IN MW DETERMINATION

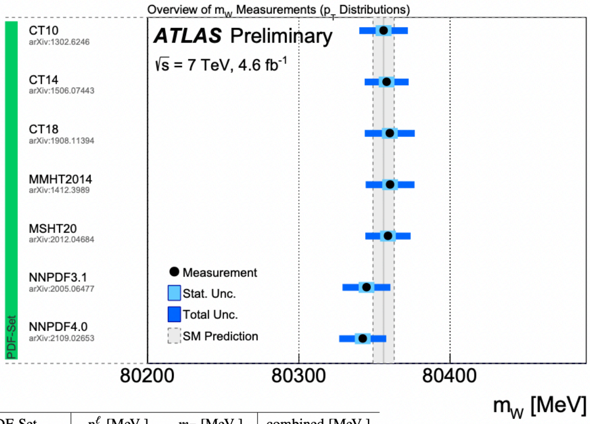
ATLAS-CONF-2023-004

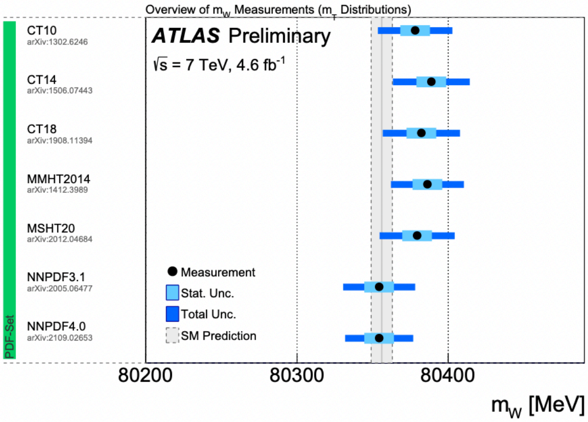


- Template method: W mass extracted by comparing observed spectra to LHC $p_{T,l}$ and $m_{T,W}$ spectra obtained with a given input theory (pQCD and PDF set)
- Shape and normalisation depend on $p_{T,l}$: larger Mw, larger cross section, faster drop at high $p_{T,l}$
- ATLAS note discuss PDF dependence

ON THE ATLAS W-MASS ANALYSIS

ATLAS-CONF-2023-004





PDF-Set	p_T^ℓ [MeV]	m_T [MeV]	combined [MeV]
CT10	$80355.6^{+15.8}_{-15.7}$	$80378.1^{+24.4}_{-24.8}$	$80355.8^{+15.7}_{-15.7}$
CT14	$80358.0^{+16.3}_{-16.3}$	$80388.8^{+25.2}_{-25.5}$	$80358.4^{+16.3}_{-16.3}$
CT18	$80360.1^{+16.3}_{-16.3}$	$80382.2^{+25.3}_{-25.3}$	$80360.4^{+16.3}_{-16.3}$
MMHT2014	$80360.3^{+15.9}_{-15.9}$	$80386.2^{+23.9}_{-24.4}$	$80361.0^{+15.9}_{-15.9}$
MSHT20	$80358.9^{+13.0}_{-16.3}$	$80379.4^{+24.6}_{-25.1}$	$80356.3^{+14.6}_{-14.6}$
NNPDF3.1	$80344.7^{+15.6}_{-15.5}$	$80354.3^{+23.6}_{-23.7}$	$80345.0^{+15.5}_{-15.5}$
NNPDF4.0	$80342.2^{+15.3}_{-15.3}$	$80354.3^{+22.3}_{-22.4}$	$80342.9^{+15.3}_{-15.3}$

- Similar uncertainties obtained with different PDF sets
- Differences of ~18 MeV between central values obtained with NNPDF4.0 and CT18 comparable (larger) than measurement uncertainty
- Important to confirm results and understand the reason behind the differences

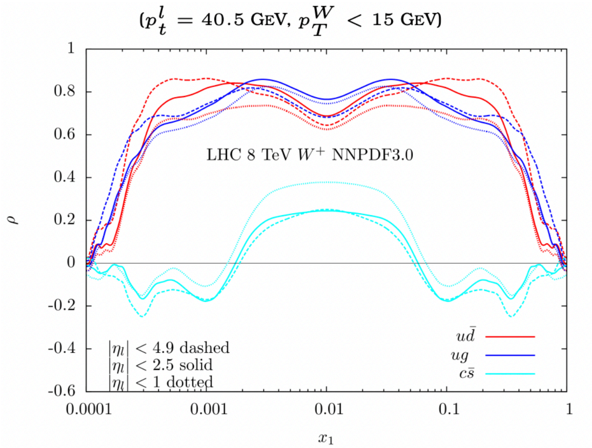
ON THE ATLAS W-MASS ANALYSIS

- PDF uncertainty dominant especially in the m_T channel (although reduced compared to previous analysis)

ATLAS-CONF-2023-004

Obs.	Mean [MeV]	Elec. Unc.	PDF Unc.	Muon Unc.	EW Unc.	PS & A_i Unc.	Bkg. Unc.	Γ_W Unc.	MC stat. Unc.	Lumi Unc.	Recoil Unc.	Total sys.	Data stat.	Total Unc.
p_T^ℓ	80360.1	8.0	7.7	7.0	6.0	4.7	2.4	2.0	1.9	1.2	0.6	15.5	4.9	16.3
m_T	80382.2	9.2	14.6	9.8	5.9	10.3	6.0	7.0	2.4	1.8	11.7	24.4	6.7	25.3

ON THE ATLAS W-MASS ANALYSIS



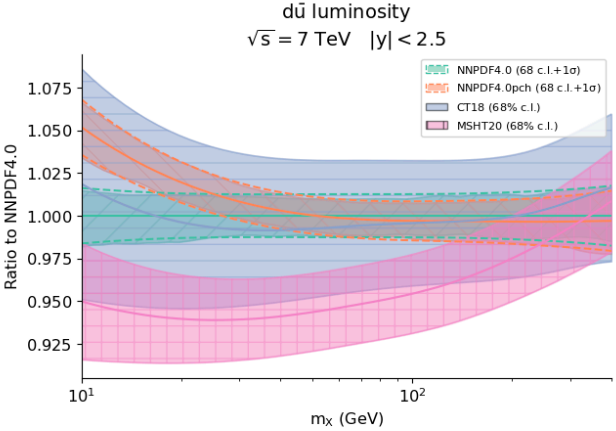
Bozzi, Citelli, Vicini, arXiv: 1501.05587

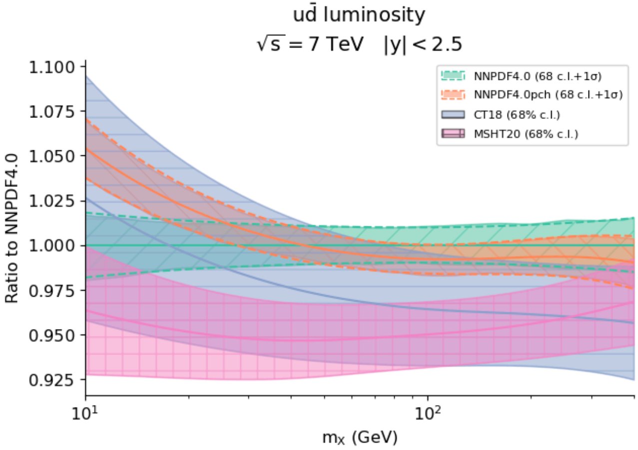
- PDF uncertainty dominant especially in the m_T channel (although reduced compared to previous analysis)
- At LHC (contrarily to Tevatron) antiquark play an important role as well as gluon and heavy quarks (approx 25% of W boson production induced by charm - strange)
- Fitted charm might have an important role

ATLAS-CONF-2023-004

Obs.	Mean [MeV]	Elec. Unc.	PDF Unc.	Muon Unc.	EW Unc.	PS & A_i Unc.	Bkg. Unc.	Γ_W Unc.	MC stat. Unc.	Lumi Unc.	Recoil Unc.	Total sys.	Data stat.	Total Unc.
p_T^ℓ	80360.1	8.0	7.7	7.0	6.0	4.7	2.4	2.0	1.9	1.2	0.6	15.5	4.9	16.3
m_T	80382.2	9.2	14.6	9.8	5.9	10.3	6.0	7.0	2.4	1.8	11.7	24.4	6.7	25.3

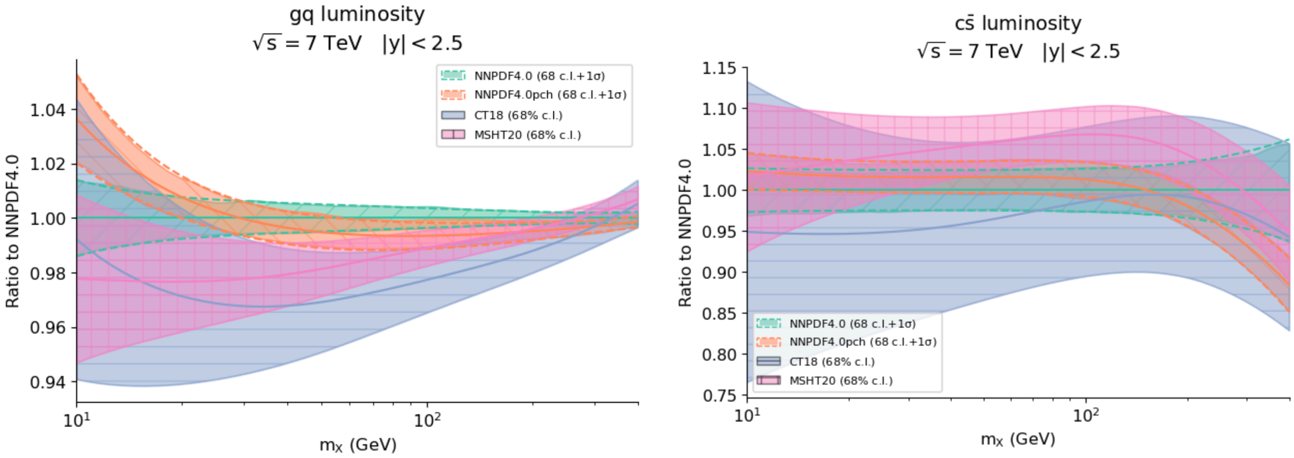
LUMINOSITY COMPARISON





- NNPDF4.0 uncertainty in light quark-antiquark luminosity half of MSHT and a third of CT18
- Perturbative charm versus fitted charm increases luminosities at M ≲ 30 GeV

LUMINOSITY COMPARISON



- NNPDF4.0 uncertainty in light quark-gluon luminosity larger than MSHT at smaller M and comparable at larger M, CT18 larger uncertainty across the whole range
- Perturbative charm versus fitted charm big effect on charm-strange luminosity

Dataset	N_{dat}	χ^2	n_σ	Z	ω	decision
CDF Z differential	29	1.23	+0.87	5.966	—	
D0 Z differential	28	0.65	−1.31	1.000	—	
D0 W electron asymmetry	11	3.54	+5.97	1.939	429	remove
D0 W muon asymmetry	9	1.64	+1.35	1.402	—	
ATLAS low-mass DY 7 TeV	6	0.89	−0.18	3.696	—	
ATLAS high-mass DY 7 TeV	5	1.67	+1.06	3.110	—	
ATLAS W, Z 7 TeV ($\mathcal{L} = 35$ pb^{-1})	30	0.94	−0.24	3.451	—	
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6$ fb^{-1}) central	46	1.86	+4.13	9.013	102	retain
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6$ fb^{-1}) forward	15	1.04	+0.10	2.838	—	
CMS W electron asymmetry 7 TeV	11	0.97	−0.07	1.061	—	
CMS W muon asymmetry 7 TeV	11	1.69	+1.61	1.246	—	
CMS DY 2D 7 TeV	110	1.34	+2.50	8.785	—	
LHCb $Z \to ee$ 7 TeV	17	1.25	+0.72	1.436	—	
LHCb $W, Z \to \mu$ 7 TeV	29	2.32	+5.04	2.890	162	retain
ATLAS W 8 TeV	22	3.50	+8.29	11.28	214	remove
ATLAS low-mass DY 2D 8 TeV	60	1.26	+1.42	1.120	—	
ATLAS high-mass DY 2D 8 TeV	48	1.11	+0.53	2.568	—	
CMS W rapidity 8 TeV	22	1.20	+0.65	13.51	—	
LHCb $Z \to ee$ 8 TeV	17	1.25	+0.72	1.436	—	
LHCb $W, Z \to \mu$ 8 TeV	30	1.39	+1.51	2.542	—	
LHCb $W \to e$ 8 TeV	8	2.61	+3.22	1.005	590	remove
ATLAS $\sigma^{tot}_{W,Z}$ 13 TeV	3	0.97	−0.03	4.961	—	
LHCb $Z \to ee$ 13 TeV	16	0.94	−0.16	2.354	—	
LHCb $Z \to \mu\mu$ 13 TeV	15	1.66	+1.80	1.608	—	
ATLAS W^++jet 8 TeV	15	0.76	−0.65	4.020	—	
ATLAS W^-+jet 8 TeV	15	1.50	+1.36	5.679	—	
ATLAS Z p_T 8 TeV ($p_T, m_{\ell\ell}$)	44	0.91	−0.42	3.325	—	
ATLAS Z p_T 8 TeV (p_T, y_Z)	48	0.89	−0.52	8.815	—	
CMS Z p_T 8 TeV	28	1.38	+1.41	9.521	—	

Only in NNPDF4.0
In NNPDF4.0 and MHST20
Excluded from NNPDF4.0

- NNPDF4.0 dataset includes largest set of Drell-Yan data

  66 inclusive Tevatron data
  491 inclusive LHC data
  150 V+jet LHC data

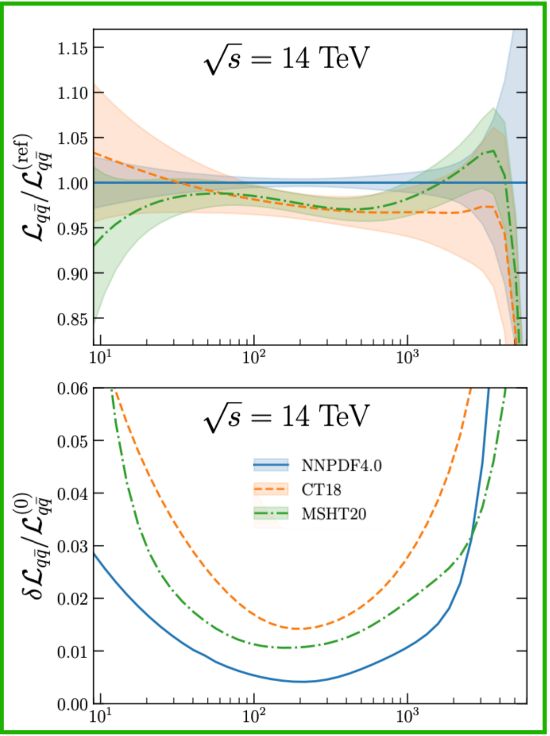
- Data with suboptimal chi2 retained or discarded on the basis of consistency with the global dataset

- CT18 has **~300** less Drell-Yan data than NNPDF4.0

UNDERSTAND THE ORIGIN OF DIFFERENCES

Ball et al, arXi: 2109.02653

Dataset
CDF Z differential
D0 Z differential
D0 W electron asymmetry
D0 W muon asymmetry
ATLAS low-mass DY 7 TeV
ATLAS high-mass DY 7 TeV
ATLAS W, Z 7 TeV ($\mathcal{L} = 35$ pb^{-1})
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6$ fb^{-1}) central
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6$ fb^{-1}) forward
CMS W electron asymmetry 7 TeV
CMS W muon asymmetry 7 TeV
CMS DY 2D 7 TeV
LHCb $Z \to ee$ 7 TeV
LHCb $W, Z \to \mu$ 7 TeV
ATLAS W 8 TeV
ATLAS low-mass DY 2D 8 TeV
ATLAS high-mass DY 2D 8 TeV
CMS W rapidity 8 TeV
LHCb $Z \to ee$ 8 TeV
LHCb $W, Z \to \mu$ 8 TeV
LHCb $W \to e$ 8 TeV
ATLAS $\sigma^{tot}_{W,Z}$ 13 TeV
LHCb $Z \to ee$ 13 TeV
LHCb $Z \to \mu\mu$ 13 TeV
ATLAS W^++jet 8 TeV
ATLAS W^-+jet 8 TeV
ATLAS Z p_T 8 TeV ($p_T, m_{\ell\ell}$)
ATLAS Z p_T 8 TeV (p_T, y_Z)
CMS Z p_T 8 TeV



- NNPDF4.0 dataset includes largest set of Drell-Yan data

  66 inclusive Tevatron data
  491 inclusive LHC data
  150 V+jet LHC data

- Data with suboptimal chi2 retained or discarded on the basis of consistency with the global dataset

- CT18 has **~300** less Drell-Yan data than NNPDF4.0

- When comparing to CT18 and MSHT20 luminosities one should consider both impact of data (and selection criteria) and impact of methodology

- Public codes would make benchmarking accessible

DATA–THEORY AGREEMENT

Dataset	NNPDF31	NNPDF40	MMHT14	MSHT20	CT18NNLO	ABMP16
CDF Z rapidity	24|28 / 28	28|30 / 28	30|31 / 28	32|32 / 28	27|27 / 28	31|31 / 28
CDF W asymmetry	11|57 / 13	14|17 / 13	12|13 / 13	28|27 / 13	11|35 / 13	21|43 / 13
D0 Z rapidity	22|22 / 28	23|23 / 28	23|23 / 28	24|23 / 28	22|22 / 28	22|22 / 28
D0 $We\nu$ lepton asymmetry	22|32 / 13	23|29 / 13	52|51 / 13	42|40 / 13	19|32 / 13	26|24 / 13
D0 $W\mu\nu$ lepton asymmetry	12|14 / 10	12|16 / 10	11|14 / 10	11|13 / 10	12|13 / 10	11|12 / 10
ATLAS peak CC Z rapidity	13|18 / 12	13|17 / 12	58|89 / 12	17|19 / 12	11|77 / 12	18|32 / 12
ATLAS W^- lepton rapidity	12|18 / 11	12|15 / 11	33|33 / 11	16|17 / 11	9.9|28 / 11	14|17 / 11
ATLAS W^+ lepton rapidity	8.9|13 / 11	8.6|11 / 11	15|21 / 11	12|13 / 11	9.4|16 / 11	10|12 / 11
Correlated χ^2	76|110	63|83	212|236	91|102	43|251	86|108
Log penalty χ^2	-0.62|-0.62	-0.58|-0.58	-1.62|-1.62	-2.89|-2.89	-1.68|-1.68	-2.72|-2.72
Total χ^2 / dof	200|312 / 126	195|242 / 126	445|509 / 126	270|283 / 126	163|499 / 126	236|300 / 126
χ^2 p-value	0.00	0.00	0.00	0.00	0.02	0.00

NNPDF4.0 and MSHT20 best data-theory agreement
(without PDF uncertainty in the χ^2 computation)

CERN-LPCC-2022-06 ; FERMILAB-TM-2779-V

CT18 and NNPDF4.0 best data-theory agreement
(with PDF uncertainty in the χ^2 computation)

DATA-THEORY AGREEMENT

This measurement incompatible with NNPDF4.0 bulk of data

The MSHT20 chi2 increases when PDF error included

Dataset	NNPDF31	NNPDF40	MMHT14	MSHT20	CT18NNLO	ABMP16
CDF Z rapidity	24|28 / 28	28|30 / 28	30|31 / 28	32|32 / 28	27|27 / 28	31|31 / 28
CDF W asymmetry	11|57 / 13	14|17 / 13	12|13 / 13	28|27 / 13	11|35 / 13	21|43 / 13
D0 Z rapidity	22|22 / 28	23|23 / 28	23|23 / 28	24|23 / 28	22|22 / 28	22|22 / 28
D0 $We\nu$ lepton asymmetry	22|32 / 13	23|29 / 13	52|51 / 13	42|40 / 13	19|32 / 13	26|24 / 13
D0 $W\mu\nu$ lepton asymmetry	12|14 / 10	12|16 / 10	11|14 / 10	11|13 / 10	12|13 / 10	11|12 / 10
ATLAS peak CC Z rapidity	13|18 / 12	13|17 / 12	58|89 / 12	17|19 / 12	11|77 / 12	18|32 / 12
ATLAS W^- lepton rapidity	12|18 / 11	12|15 / 11	33|33 / 11	16|17 / 11	9.9|28 / 11	14|17 / 11
ATLAS W^+ lepton rapidity	8.9|13 / 11	8.6|11 / 11	15|21 / 11	12|13 / 11	9.4|16 / 11	10|12 / 11
Correlated χ^2	76|110	63|83	212|236	91|102	43|251	86|108
Log penalty χ^2	-0.62|-0.62	-0.58|-0.58	-1.62|-1.62	-2.89|-2.89	-1.68|-1.68	-2.72|-2.72
Total χ^2 / dof	200|312 / 126	195|242 / 126	445|509 / 126	270|283 / 126	163|499 / 126	236|300 / 126
χ^2 p-value	0.00	0.00	0.00	0.00	0.02	0.00

Poor description of ATLAS data

NNPDF4.0 and MSHT20 best data-theory agreement
(without PDF uncertainty in the χ^2 computation)

CERN-LPCC-2022-06 ; FERMILAB-TM-2779-V

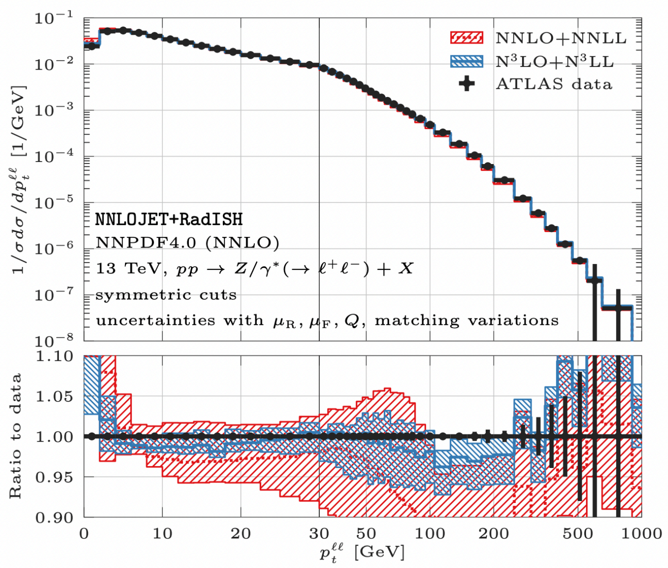
CT18 and NNPDF4.0 best data-theory agreement
(with PDF uncertainty in the χ^2 computation)

DATA-THEORY AGREEMENT

- NNPDF4.0 shows perfect agreement with the ATLAS 13 TeV data once N3LO+ N3LL prediction is used in the data theory comparison, without the need to introduce intrinsic kT or other non perturbative modelling

- Would a re-tune of Monte Carlo change the situation when going from Z pT to W pT?

Chen et al arXiv: 2203.01565



THEORY FRONTIERS

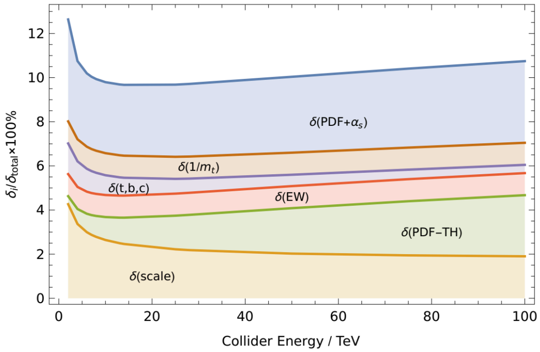
THEORY UNCERTAINTIES IN PDF FITS

$$\sigma = \alpha_s^p \sigma_0 + \alpha_s^{p+1}\sigma_1 + \alpha_s^{p+2}\sigma_2 + \mathcal{O}(\alpha_s^{p+3})$$

- Standard global PDF fits based on fixed-order NNLO QCD calculations (using fast interpolation grid for NLO predictions accompanied by local K-factors for NNLO). PDF uncertainty reflects experimental uncertainty.
- N3LO is now the precision frontier for partonic cross sections (N3LO splitting functions partially known)
- Mismatch between perturbative order of partonic cross section and PDFs becoming significant source of uncertainty

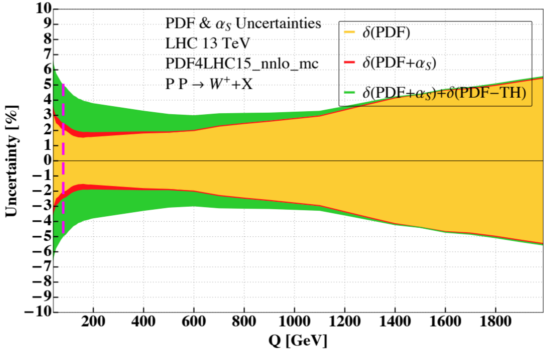
$$\delta(PDF - TH) = \frac{1}{2}\left|\frac{\sigma^{(2)}_{\mathrm{NNLO-PDFs}} - \sigma^{(2)}_{\mathrm{NLO-PDFs}}}{\sigma^{(2)}_{\mathrm{NNLO-PDFs}}}\right|$$

Gluon-gluon fusion into Higgs



M. Cepeda et al. [HL/HE WG2 group], arXiv:1902.00134

Drell-Yan



C. Duhr et al, *JHEP* 11 (2020) 143

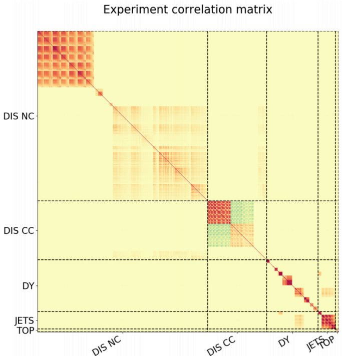
THEORY UNCERTAINTIES IN PDF FITS

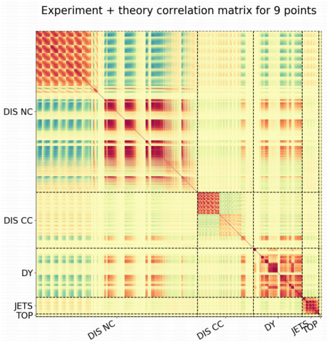
How to include Missing Higher Order Uncertainties in a PDF uncertainties? Lots of activity in PDF community

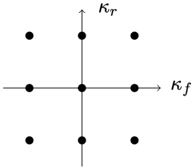
- Option 1 - theory covmat [NNPDF: 1906.10698]
  Construct a theory covariance matrix from scale-varied cross sections and combine it with the experimental covariance matrix

$$\chi^2 = \sum_{m,n=1}^{N} (d_m - t_m)(\text{cov}_{\text{exp}} + \text{cov}_{\text{th}})^{-1}_{mn}(d_n - t_n)$$

- Assumptions: experimental and theoretical errors independent and Gaussian
- Assumptions on correlation of scales and scale ratio will determine the specific form of the covariance matrix



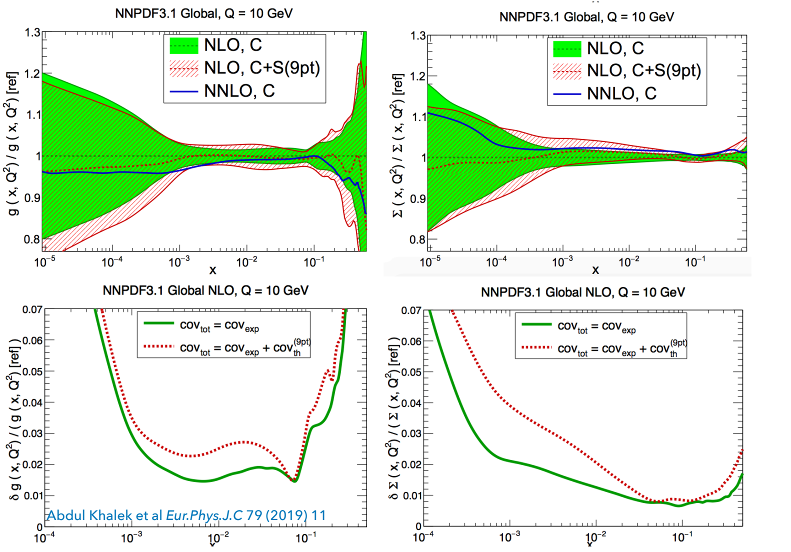


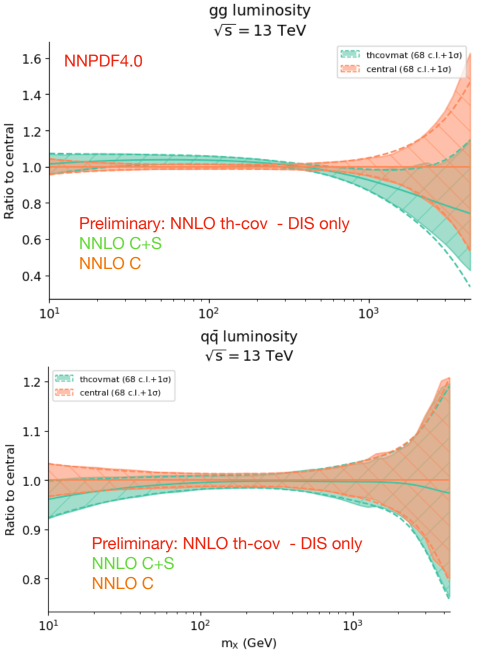


9-pts variation

Abdul Khalek et al
Eur.Phys.J.C 79 (2019) 11

THEORY UNCERTAINTIES IN PDF FITS





➡ Missing correlation between scale variation in PDF fits and hard cross sections [Harland-Lang, Thorne Eur.Phys.J.C 79 (2019) 3, 225] (See also Ball, Pearson arXiv:2105.05114)

➡ NNPDF4.0 NNLO with MHOU will be out before summer

THEORY UNCERTAINTIES IN PDF FITS

How to include Missing Higher Order Uncertainties in a PDF uncertainties? Lots of activity in PDF community

- Option 2 - MCscales [Kassabov et al: 2207.07616]
  - Main idea: renormalisation and factorisation scales are free parameters of the fixed-order theory, that induce an uncertainty on the theory predictions included in a PDF fit & need to be propagated
  - Joint sampling of experimental uncertainty (propagated to PDF uncertainty by MC sampling) by specifying a suitable prior probability distribution of all possible scale choices & a-posteriori criterion based on agreement with the data.

$$P\left(k_f = \xi_f, k_{r_1} = \xi_1, \ldots, k_{r_{N_\mathrm{p}}} = \xi_{N_\mathrm{p}}\right) = P(\omega)$$

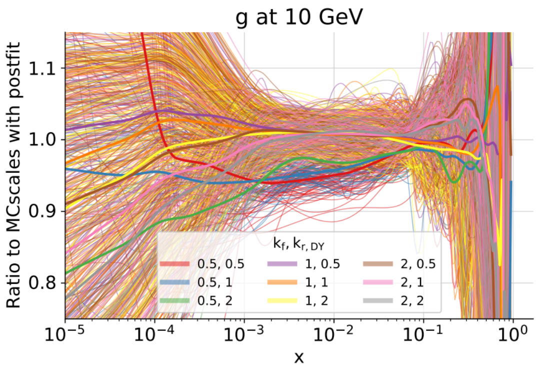
with $\omega \in \Omega = \{(\xi_f, \xi_1, \ldots, \xi_{N_\mathrm{p}}) \;\forall\; \xi_f, \xi_1, \ldots, \xi_{N_\mathrm{p}} \in \Xi\}$

3^{1+N_p} elements, with N_p = 5, p=DIS NC, DIS CC, DY, JET, TOP

Choose prior = choose P($\boldsymbol{\omega}$)

Posterior

$$\chi^2_n > \langle \chi^2 \rangle_{n\,|\,\omega^{(n)}=\{1,\ldots,1\}} + 4\,\mathrm{std}(\chi^2)_{n\,|\,\omega^{(n)}=\{1,\ldots,1\}}$$



THEORY UNCERTAINTIES IN PDF FITS

How to include Missing Higher Order Uncertainties in a PDF uncertainties? Lots of activity in PDF community

- Option 3 - theory nuisance parameters [MMHT: 2207.04739]
  - Main idea: add MHOU as nuisance parameters
  - Fit nuisance parameters from data

T. Cridge - DIS2023

Consider usual PDF fit probability:

Theory, Data, Hessian matrix - contains uncorrelated (s_k) and correlated uncertainties (β_k)

$$P(T|D) \propto \exp(-\chi^2) \propto \exp\left(-\frac{1}{2}(T-D)^T H_0 (T-D)\right)$$

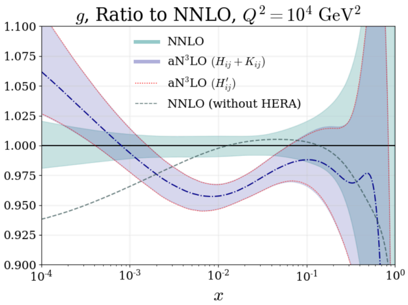
$$\propto \exp\left(-\frac{1}{2}\sum_{k=1}^{N_{pt}} \frac{1}{s_k^2}\left(D_k - T_k - \sum_{\alpha=1}^{N_{corr}} \beta_{k,\alpha}\lambda_\alpha\right)^2 + \sum_{\alpha=1}^{N_{corr}} \lambda_\alpha{}^2\right)$$

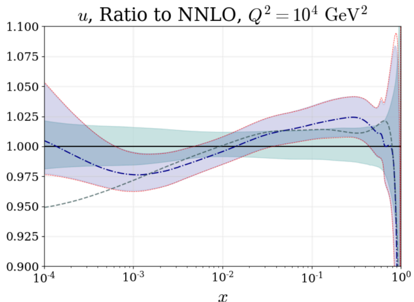
Experimental Nuisance parameters

- Unknown terms in N3LO theory added via a Gaussian prior and parameters fitted from the data give rise to aN3LO MSHT PDFs

$$T' = T + tu + (\theta - t)u = T'_0 + \theta' u$$

$$P(\theta') = \frac{1}{\sqrt{2\pi}\sigma_{\theta'}} \exp\left(-\theta'^2/2\sigma^2_{\theta'}\right)$$



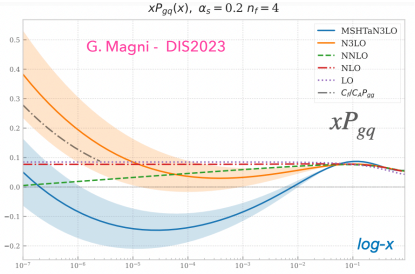


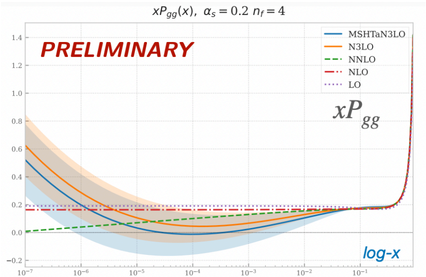
APPROXIMATED N3LO PDFS

MSHT - arXiv:2207.04739
NNPDF - in preparation

- MSHT & NNPDF: inclusion of available theoretical ingredients at N3LO (non-singlet splitting functions, singlet splitting function in the large nf limit, small-x limit, large-x limit, Mellin moments + DIS structure functions in the massless limit and approximate heavy flavour structure functions between known limits + hadronic N3LO K-factors)
- MSHT: MHOU (missing higher order uncertainty) and IHOU (incomplete higher order uncertainty) added as nuisance parameters and fitted from the data
- NNPDF: MHOU added via theory covariance matrix, IHOU added as extra additional theory uncertainties computed by varying each of the possible parametrisation that interpolate the known ingredients

$$\delta^2_{th} = \delta^2_{MHOU} + \delta^2_{aN3LO} \qquad \delta^2_{aN3LO_{mn}} = \frac{1}{N_{dat}}(\sigma_{i,n} - \sigma_{0,n})(\sigma_{i,m} - \sigma_{0,m}), \quad n = \{0,\ldots,N_{dat}\} \; i = \{0,\ldots,N_{var}\}$$

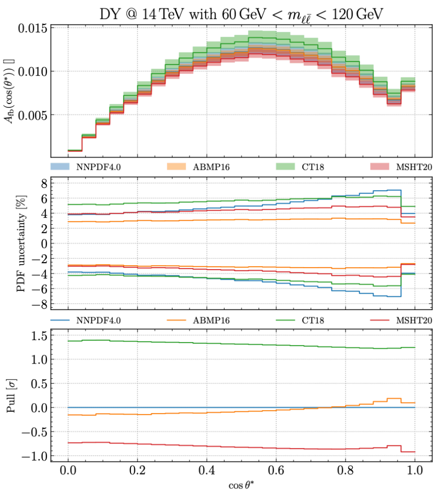


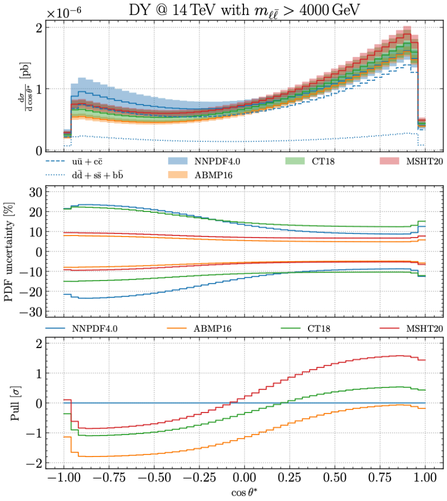


PDFS AND NEW PHYSICS

FORWARD BACKWARD ASYMMETRY

- ✓ High mass Drell-Yan tails affected by large PDF uncertainties
- ✓ This affects searches for new physics
- ✓ Need data constraining large-x to see and characterise new physics (at the LHC high energy and/or high rapidity - high-x))



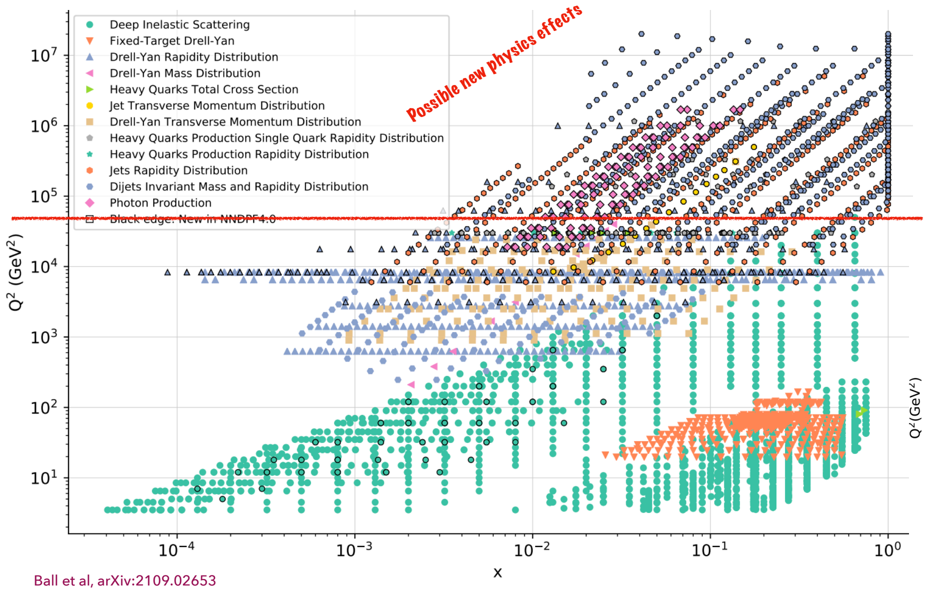


$$\frac{d\sigma}{d\cos\theta^*} = \int_{m_{\ell\bar{\ell}}^{\min}}^{\sqrt{s}} dm_{\ell\bar{\ell}} \int_{\ln(m_{\ell\bar{\ell}}/\sqrt{s})}^{\ln(\sqrt{s}/m_{\ell\bar{\ell}})} dy_{\ell\bar{\ell}} \frac{d^3\sigma}{dm_{\ell\bar{\ell}}\,dy_{\ell\bar{\ell}}\,d\cos\theta^*}$$

Collins-Soper angle

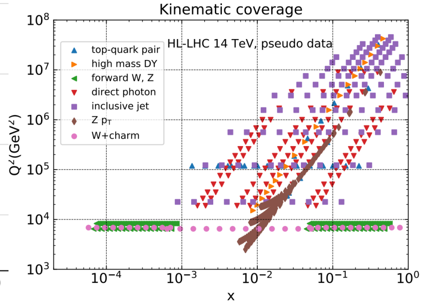
Ball et al, arXiv: 2209:08115

INTERPLAY WITH INDIRECT NEW PHYSICS SEARCHES



Ball et al, arXiv:2109.02653

- Top pair production and single top data included in SMEFT analysis [Hartland et al 1901.05965] [Ellis et al 2012.02779]
- Dijets data in [Bordone et al 2103.10332] [Alioli et al 1706.03068]
- Drell-Yan data in [Farina et al 1609.08157] [Torre et al 2008.12978]
- Inclusive jets in [Alte et al 1711.07484]
- Overlap enhanced in HL-LHC projections [Abdul Khalek et al,1810.03639]



Abdul Khalek et al, arXiv:1810.03639

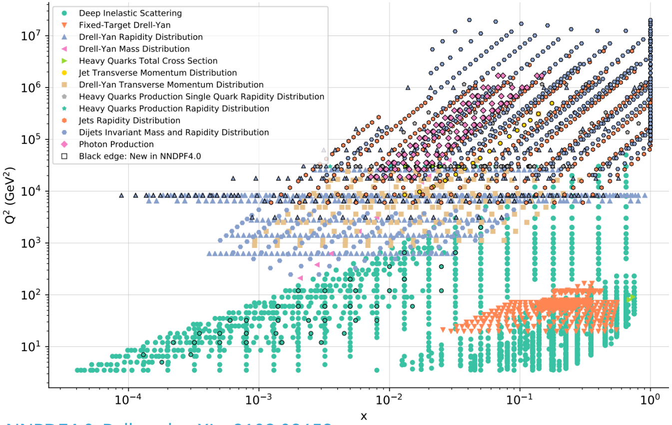
EXTRACTING PHYSICS PARAMETERS FROM LHC DATA

$$\chi^2 = \frac{1}{N_{\rm dat}} \sum_{i=1}^{N_{\rm dat}} (T_i(\{\theta\},\{c\}) - D_i)\, {\rm cov}^{-1}_{ij}\, (T_j(\{\theta\},\{c\}) - D_j)$$

$$T_i(\{\theta\},\{c\}) = \mathrm{PDFs}(\{\theta\},\{c\}) \otimes \hat{\sigma}_i(\{c\})$$

(B)SM parameters: $\alpha_s(M_Z)$, M_W, θ_W ,SMEFT WCs…..

Parameters determining PDFs at initial scale



NNPDF4.0, Ball et al, arXiv: 2109.02653

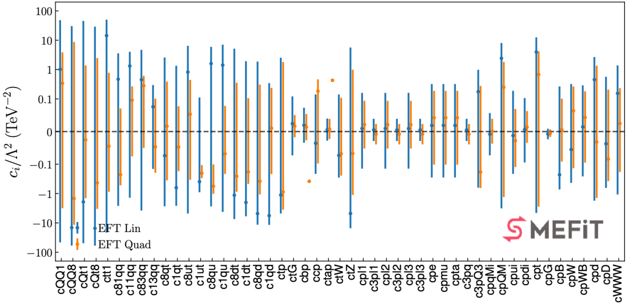
✓ In a PDF fit typically

$$T_i(\{\theta\}) = \mathrm{PDFs}(\{\theta\},\{c=\bar{c}\}) \otimes \hat{\sigma}_i(\{c=\bar{c}\})$$

EXTRACTING PHYSICS PARAMETERS FROM LHC DATA

$$\chi^2 = \frac{1}{N_{\rm dat}} \sum_{i=1}^{N_{\rm dat}} (T_i(\{\theta\},\{c\}) - D_i)\, {\rm cov}^{-1}_{ij}\, (T_j(\{\theta\},\{c\}) - D_j)$$

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \cdots$$



$$T_i(\{\theta\},\{c\}) = {\rm PDFs}(\{\theta\},\{c\}) \otimes \hat{\sigma}_i(\{c\})$$

(B)SM parameters: $\alpha_s(M_Z)$, M_W, θ_W, SMEFT WCs…..

Parameters determining PDFs at initial scale

✓ In a PDF fit typically

$$T_i(\{\theta\}) = {\rm PDFs}(\{\theta\},\{c=\bar{c}\}) \otimes \hat{\sigma}_i(\{c=\bar{c}\})$$

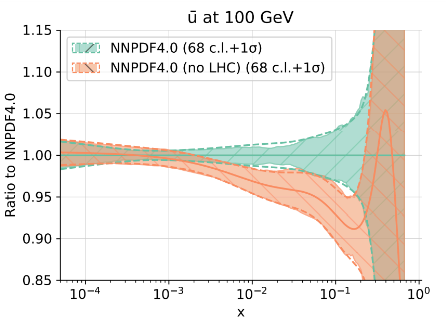
✓ In a fit of (B)SM parameters

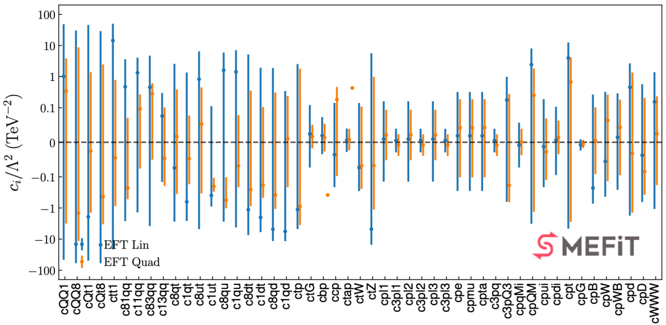
$$T_i(\{c\}) = {\rm PDFs}(\{\bar{\theta}\},\{\bar{c}\}) \otimes \hat{\sigma}_i(\{c\})$$

Ethier et al, arXiv: 2105.00006

INTERPLAY BETWEEN PDF FITS AND SMEFT FITS

- In principle low-scale physics is separable from high-scale physics, BUT the complexity of the LHC environment might well intertwine them.
- PDFs are low-scale quantities extracted from experimental data at all scales, without considering any potential contamination due to new physics in high-energy data.
- (SM)EFT fits are performed by assuming a priori that PDFs are SM-like.





Ball et al, arXiv:2109.02653

Ethier et al, arXiv: 2105.00006

INTERPLAY BETWEEN PDF FITS AND SMEFT FITS

- From the point of view of PDF fits:
  - ➡ How to make sure that new physics effects are not inadvertently fitted away in a PDF fit?
- From the point of view of SMEFT fits:
  - ➡ Should I make sure I am using a clean set of PDFs in a SMEFT analysis? How to define it? Is it enough?
  - ➡ How would the bounds change if I was consistently using PDFs that include in the fit the same operators that I am fitting?

T $$d\sigma^{pp\to ab} = \sum_{i,j} f_i \otimes f_j \otimes d\hat{\sigma}^{ij\to ab} + \ldots$$

$$f(\{\theta_k\})$$

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \cdots$$

➡ Simultaneous fits can shed light on their interplay

$$T(\{\theta_k\},\{c_i\})$$

SIMUNET: A DEEP-LEARNING BASED SIMULTANEOUS FIT

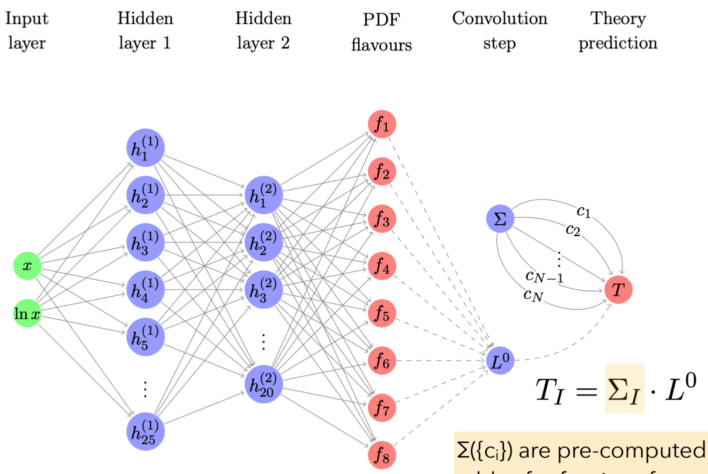
- The idea: take a PDF fit based on NNPDF4.0 methodology and make dependence of observables on physics parameters $\{c_i\}$ explicit via fast interface before computing the loss function (e.g. adding SMEFT corrections, expanding observables in terms of SM precision parameters)

- Perform minimisation of loss function over

$$\hat{\theta} = \theta \bigcup \{c_i\}$$

by adding new layer to the deep neural network used in NNPDF4.0

- Can expand dependence on c_i beyond linear terms in T (up to generic power in polynomial expansion) by adding non-trainable edges
- Can be done both for SM parameters and SMEFT coefficients.

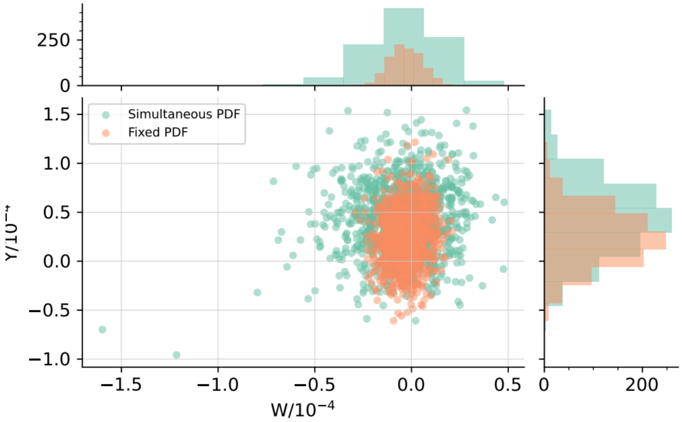


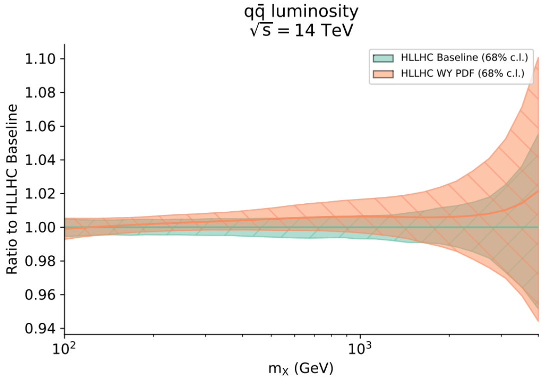
$\Sigma(\{c_i\})$ are pre-computed tables for fast interface accounting for PDF evolution and part. xsec

S. Iranipour, MU - arXiv: 2201.07240

SIMUNET: APPLICATION TO DRELL-YAN

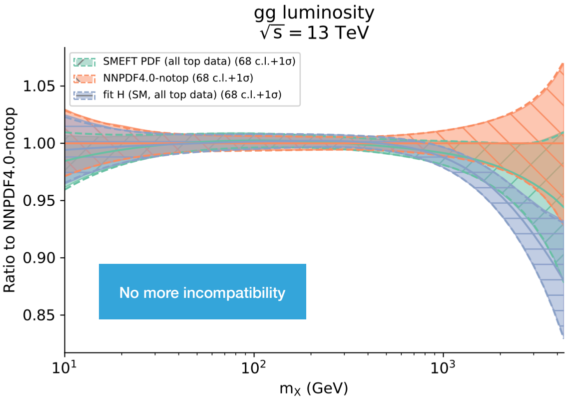
S. Iranipour, MU - arXiv: 2201.07240



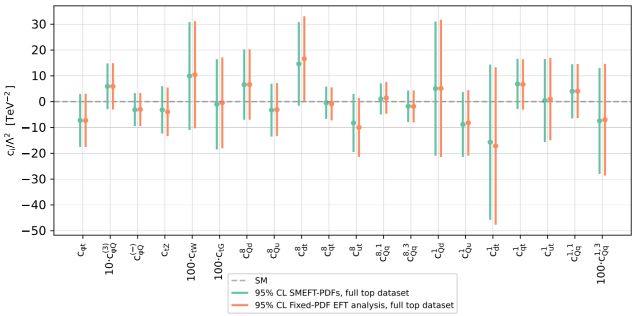


- ✓ Simultaneous analysis of PDFs and W&Y SMEFT coefficient of DIS + DY (including HL-LHC projections) using simuNET method shows that at HL-LHC the effect of interplay becomes important as WCs bounds broaden and PDFs change significantly once SMEFT effects allowed in theory predictions entering PDF fit
- ✓ Stress-tested and shown robustness with closure tests

SIMUNET: APPLICATION TO TOP



25 (21) dim-6 operators at the quadratic (linear) SMEFT



- Huge amount of Run II top quark data from ATLAS and CMS
- Four basic processes: inclusive tt~ and asymmetry (inclusive and differential), single top (inclusive and differential, associated ttV production, associated single top production
- Recent analysis of Run II top legacy data in PDF fit, SMEFT fit and simultaneous fit shows change in large M luminosity and stability in the SMEFT determination

Z. Kassabov, M. Madigan, L. Mantani, J. Moore, M. Morales, J. Rojo, MU - arXiv: 2303.06159

CONCLUSIONS

- The path towards 1% accurate PDFs opens up new challenges
- Precise and accurate determination of the proton structure is key to make progress in precision phenomenology - precision level highlights critical points of the formalism
- Need: robust methodology & solid statistical tests of methodology
- Benchmark effort would benefit of public releases of PDF codes and inputs!

- M_W determination opens up questions about different results obtained with different PDFs.
- For precision's sake, should move out of simplistic paradigm "conservative = good: and understand what goes in the fits that makes a difference

- Current uncertainties partially underestimated due to theory uncertainty not accounted for at NNLO & other effects (fixed-order calculations [Alekhin et al 2104.02400], full account of electroweak corrections, resummations, SMEFT effects, nuclear corrections [Ball et al 2105.05114])
- Inclusion of missing higher order uncertainties and approximate N3LO PDFs hot topics

- Simultaneous fits can shed light on interplay between high energy tails in PDF fits and SMEFT fits (could be expanded to simultaneous fits of PDFs and EW parameters)
- In parallel a more careful investigation of definition of "safe" PDF sets & account for PDF uncertainties