Recent Developements and Open Problems in Parton Distributions

Juan Rojo

Dipartimento di Fisica, Università degli Studi di Milano

Les Rencontres de Physique de la Valle d'Aoste 2011
La Thuile, 01/03/2011

Outline

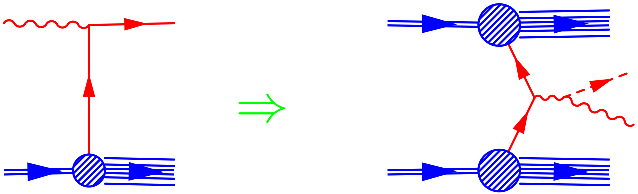
Motivation

- The accurate determination of parton distributions (PDFs) is a crucial requirement for the LHC physics program
- PDFs and their uncertainties are relevant for standard candles ($W, Z, t\bar{t}$), SM backgrounds (inclusive jets, photons) and new physics searches (Higgs, SUSY)
- Using accurate PDFs specially relevant for Higgs exclusion bounds (and discovery claims?) at Tevatron and LHC

PARTONS FOR LHC:

THE ACCURATE COMPUTATION OF PHYSICAL PROCESS AT A HADRON COLLIDER REQUIRES GOOD KNOWLEDGE OF PARTON DISTRIBUTIONS OF THE NUCLEON

FACTORIZATION



IN ORDER TO EXTRACT THE RELEVANT PHYSICS SIGNAL,

WE NEED TO KNOW THE PARTON DISTRIBUTIONS AND THEIR UNCERTAINTY

IS THIS ASPECT OF LHC PHYSICS UNDER CONTROL?

An ideal PDF set

S. Forte, *Parton Distributions at the Dawn of the LHC*, arxiv:1011.5247

1. Use a wide dataset → ensure all relevant experimental information retained
2. Use sufficiently general and unbiased parton parametrization or estimate effect of parton parametrization variations
3. Provide PDF uncertainty bands checked to provide consistently-sized confidence levels for individual experiments
4. Include heavy quark mass effects through a GM-VFN scheme, and provide estimate of uncertainties due to subleading terms
5. Use exact computations performed at the highest available perturbative order
6. Provide PDFs for a variety of values of $\alpha_s\left(M_Z\right)$ and for the heavy quark masses m_c and m_b

PDF determination was for many years a *qualitative art*
In the LHC era it must become a *quantitative science*

PDFs: Status of the Art

	Data	Pert. Order	HQ scheme	α_s	Parametrisation	Stat. treatment
CT10(w)	global eP & PP data (combined HERA)	NLO	SACOT-chi	external parameter - several as values	6 independent PDFs Polynomial param 26 pars	Hessian with dynamical tolerance
MSTW08	global eP & PP data (HERA jets)	LO NLO NNLO (*)	ACOT + TR'	external parameter - several as + fitted	7 independent PDFs Polynomial param 20(28) pars	Hessian with dynamic tolerance
NNPDF2.1	global eP & PP data (combined HERA)	NLO	FONLL-A	external parameter - several as values	7 independent PDFs Neural Networks 259 pars	Monte Carlo sampling + cross validation
HERAPDF1.0 HERAPDF1.5	DIS HERA data [1.5] Added prel HERA II	NLO NNLO	ACOT + TR'	external parameter	5 independent PDFs Polynomial param + Chebyshev polynomial	Conventional Hessian, Monte Carlo
ABKM09 ABM10	global DIS + Fixed-Target DY (combined HERA)	NLO NNLO	FFNS n_f=4 matched with BMSN	fitted,not external parameter	6 independent PDFs Polynomial param	Hessian with no tolerance
JR08	DIS + Jets + Fixed-Target DY	NLO NNLO	FFN, nf=3 VFN	fitted,not external parameter	5 independent PDFs Valence-like assumpt 20 pars	Hessian with fixed tolerance

M. Ubiali, *Challenges for precision physics at the LHC*, Paris 2010

An ideal PDF set

S. Forte, *Parton Distributions at the Dawn of the LHC*, arxiv:1011.5247

1. **Use a wide dataset $\longrightarrow$ ensure all relevant experimental information retained**
2. Use sufficiently general and unbiased parton parametrization or estimate effect of parton parametrization variations
3. Provide PDF uncertainty bands checked to provide consistently-sized confidence levels for individual experiments
4. Include heavy quark mass effects through a GM-VFN scheme, and provide estimate of uncertainties due to subleading terms
5. Use exact computations performed at the highest available perturbative order
6. Provide PDFs for a variety of values of $\alpha_s(M_Z)$ and for the heavy quark masses m_c and m_b

Datasets in PDF analysis

Process	Subprocess	Partons	x range
$\ell^{\pm}\,\{p,n\} \to \ell^{\pm}\,X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
$\ell^{\pm}\,n/p \to \ell^{\pm}\,X$	$\gamma^*\,d/u \to d/u$	d/u	$x \gtrsim 0.01$
$pp \to \mu^+\mu^-\,X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
$pn/pp \to \mu^+\mu^-\,X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
$\nu(\bar{\nu})\,N \to \mu^-(\mu^+)\,X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
$\nu\,N \to \mu^-\mu^+\,X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
$\bar{\nu}\,N \to \mu^+\mu^-\,X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
$e^{\pm}\,p \to e^{\pm}\,X$	$\gamma^* q \to q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
$e^+\,p \to \bar{\nu}\,X$	$W^+\,\{d,s\} \to \{u,c\}$	d, s	$x \gtrsim 0.01$
$e^{\pm} p \to e^{\pm}\,c\bar{c}\,X$	$\gamma^* c \to c$, $\gamma^* g \to c\bar{c}$	c, g	$0.0001 \lesssim x \lesssim 0.01$
$e^{\pm} p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g, q	$0.01 \lesssim x \lesssim 0.5$
$p\bar{p} \to (W^{\pm} \to \ell^{\pm}\nu)\,X$	$ud \to W$, $\bar{u}\bar{d} \to W$	$u, d, \bar{u}, \bar{d}$	$x \gtrsim 0.05$
$p\bar{p} \to (Z \to \ell^+\ell^-)\,X$	$uu, dd \to Z$	d	$x \gtrsim 0.05$

MSTW08, arXiv:0901.0002

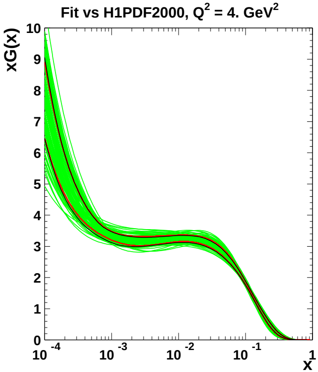
- Global PDF sets **NNPDF2.1, CT10, MSTW08** include all relevant datasets from DIS, Drell-Yan, vector-boson production and inclusive jet production
- Different data constrain different PDF combinations $\to$ Use all available experimental information for PDF determinations
- PDF sets based on reduced datasets might have some PDFs unconstrained

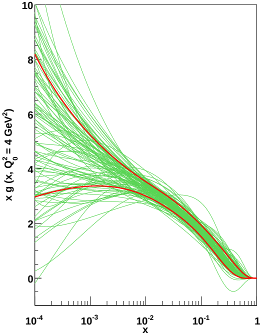
An ideal PDF set

S. Forte, *Parton Distributions at the Dawn of the LHC*, arxiv:1011.5247

1. Use a wide dataset → ensure all relevant experimental information retained
2. **Use sufficiently general and unbiased parton parametrization or estimate effect of parton parametrization variations**
3. Provide PDF uncertainty bands checked to provide consistently-sized confidence levels for individual experiments
4. Include heavy quark mass effects through a GM-VFN scheme, and provide estimate of uncertainties due to subleading terms
5. Use exact computations performed at the highest available perturbative order
6. Provide PDFs for a variety of values of $\alpha_s(M_Z)$ and for the heavy quark masses m_c and m_b

Simple functional froms vs. NeuralNets





- Simple functional forms $q(x) = Ax^b(1-x)^c P(x)$ (CT, MSTW, ABKM, HERAPDF) $\longrightarrow$ systematic underestimation of uncertainties
- Artificial Neural Networks as universal interpolants (NNPDF) $\longrightarrow$ avoid theoretical bias from choice of PDF functional form
- Compare $\mathcal{O}$(300) parms in NNPDF with $\mathcal{O}$(10-25) parms in standard PDF sets

also PDFs from Chebichev polynomials, Glazov, Moch, Radescu, arXiv:1009.6170 and Pumplin, arXiv:0909.5176

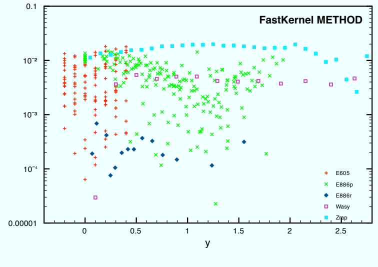
An ideal PDF set

S. Forte, *Parton Distributions at the Dawn of the LHC*, arxiv:

1. Use a wide dataset $\rightarrow$ ensure all relevant experimental information retained
2. Use sufficiently general and unbiased parton parametrization or estimate effect of parton parametrization variations
3. Provide PDF uncertainty bands checked to provide consistently-sized confidence levels for individual experiments
4. Include heavy quark mass effects through a GM-VFN scheme, and provide estimate of uncertainties due to subleading terms
5. **Use exact computations performed at the highest available perturbative order**
6. Provide PDFs for a variety of values of $\alpha_s(M_Z)$ and for the heavy quark masses m_c and m_b

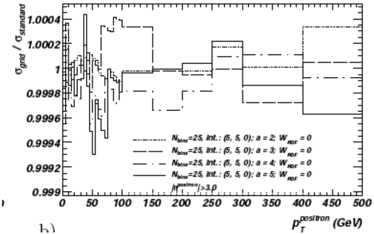
Exact NLO/NNLO computations

- NLO exact computation of hadronic observables often too slow for PDF fits.
- MSTW and CT include Drell-Yan and Jets NLO/NNLO as (local) K-factors rescaling the LO cross section
- NNPDF uses exact fast NLO calculations



- Exact fast computations of hadronic observables now available:
- **FastNLO**: interface for jet inclusive cross-section (hep-ph/0609285)
- **FastKernel**: computation of DY observables (NNPDF arXiv:1002.4407)
- **APPLgrid**: general purpose interface (arXiv:0911.2985)

APPLGRID, NLO $W^+ \rightarrow e^+\nu$ distribution



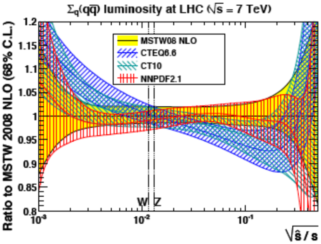
Exact NLO/NNLO computations should always be used in global PDF fits

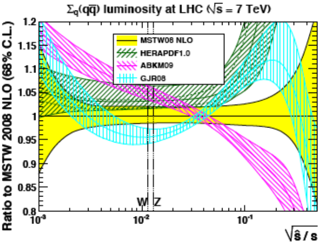
Outline

Benchmark LHC observables

Study and plots from G. Watt, Les Houches QCD 2011

$$\frac{\partial \mathcal{L}_{\Sigma_q(q\bar{q})}}{\partial \hat{s}} = \frac{1}{s}\int_\tau^1 \frac{dx}{x} \sum_{q=d,u,s,c,b} \left[f_q(x,\hat{s}) f_{\bar{q}}(\tau/x,\hat{s}) + (q \leftrightarrow \bar{q}) \right], \quad \tau \equiv \frac{\hat{s}}{s}$$

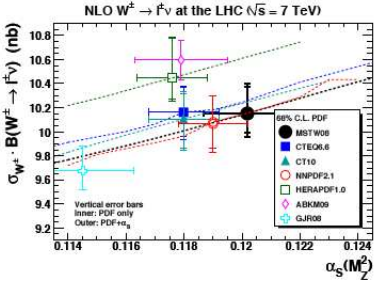


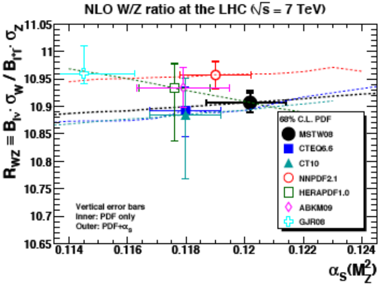


Good agreement within global PDF sets
Larger differences for non-global sets

Benchmark LHC observables

Study and plots from G. Watt, Les Houches QCD 2011



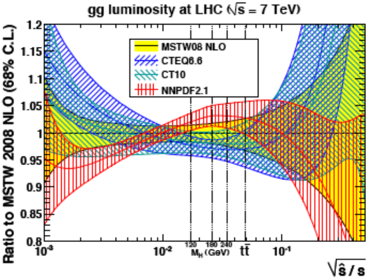


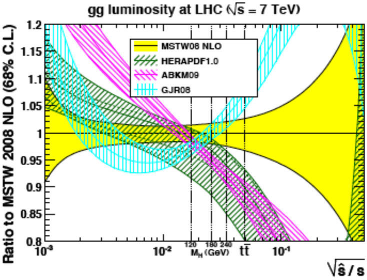
Good agreement within global PDF sets
Larger differences for non-global sets

Benchmark LHC observables

Study and plots from G. Watt, Les Houches QCD 2011

$$\frac{\partial \mathcal{L}_{gg}}{\partial \hat{s}} = \frac{1}{s}\int_{\tau}^{1}\frac{\mathrm{d}x}{x}\, f_g(x,\hat{s}) f_g(\tau/x,\hat{s}), \quad \tau \equiv \frac{\hat{s}}{s}$$

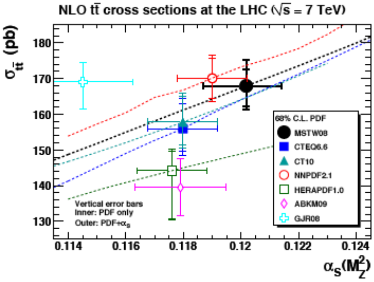


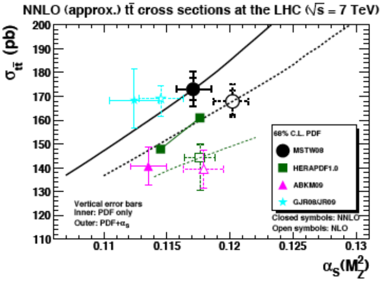


Same trend as for $q\bar{q}$
NNPDF2.1 in good agreement with MSTW08, CT10 lower (also for Higgs)

Benchmark LHC observables

Study and plots from G. Watt, Les Houches QCD 2011





Same trend as for $q\bar{q}$
NNPDF2.1 in good agreement with MSTW08, CT10 lower (also for Higgs)

The PDF4LHC working group

The PDF4LHC group (CERN management mandate) coordinates studies and research in PDF determinations from different groups and is responsible for providing official recommendations for PDF use in LHC experiments
Current NLO recommendation for LHC analysis:

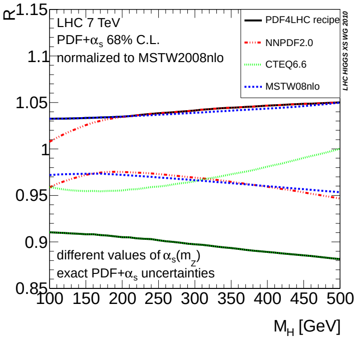
NLO Summary:

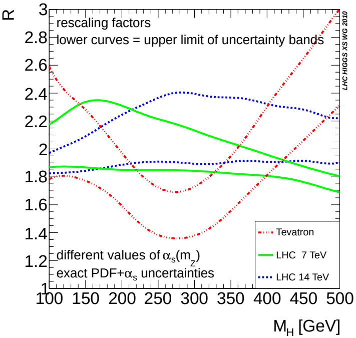
For the calculation of uncertainties at the LHC, use the envelope provided by the central values and PDF+α_s errors from the MSTW08, CTEQ6.6 and NNPDF2.0 PDFs, using each group's prescriptions for combining the two types of errors. We propose this definition of an envelope because the deviations between the predictions are currently greater than their uncertainties would strictly suggest. As a central value, use the midpoint of this

http://www.hep.ucl.ac.uk/pdf4lhc/
At NNLO: MSTW2008 NNLO central value + NLO envelope
Updated when new data / PDF sets / theoretical developements require so

The PDF4LHC recommendation and Higgs searches

PDF4LHC recommendation adopted by ATLAS, CMS and the LHC Higgs cross section working group (CERN Yellow Report, arxiv:1101.0593)

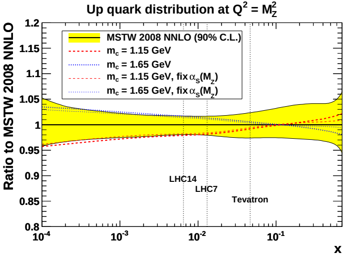


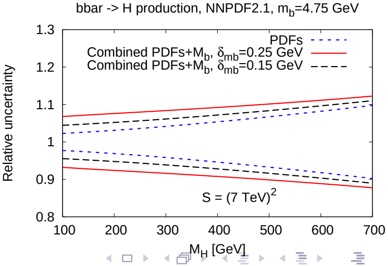


PDF4LHC recipee should be used in all LHC analysis where PDFs are relevant!

Heavy quark mass effects

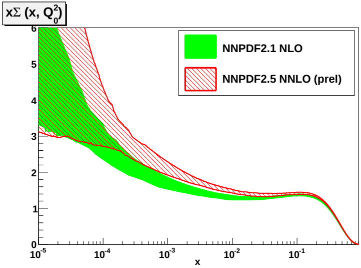
- Most recent PDF sets all based on a General-Mass VFN scheme for heavy quarks
- Precision LHC pheno requires to estimate theo uncertainties from a) Subleading terms in the GM-VFN definition and b) choice of m_c and m_b values
- NNPDF (arXiv:1101.1300) and MSTW (arXiv:1007.2624): detailed studies of heavy quark mass effects $\longrightarrow$ Important to take into account for:
    1. Process sensitive to small-x quarks: $W^{\pm}, Z$
    2. HQ initiated process: $b\bar{b} \longrightarrow H$, single top
- Also HERAPDF (arXiv:1006.4471)
- Dependence on m_h reduced if MSbar mass used in the GM-VFN (Alekhin and Moch, arXiv:1011.5790)

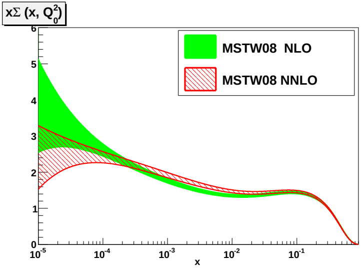




NNPDF@NNLO

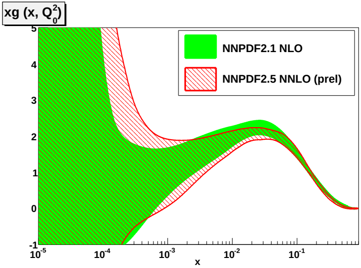
- NNPDF2.5: (preliminary) NNLO analysis based on FONLL-C for DIS structure functions
- NNLO only shifts central values by 1–1.5–σ at most, PDF uncertainties unchanged
- Harder small–x quarks, stable gluon at small-x (larger differences for MSTW)

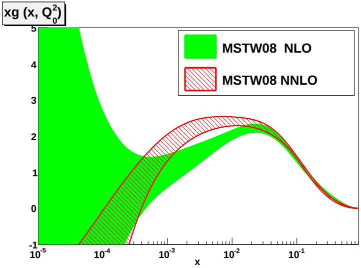




NNPDF@NNLO

- NNPDF2.5: (preliminary) NNLO analysis based on FONLL-C for DIS structure functions
- NNLO only shifts central values by 1–1.5–σ at most, PDF uncertainties unchanged
- Harder small–x quarks, stable gluon at small-x (larger differences for MSTW)

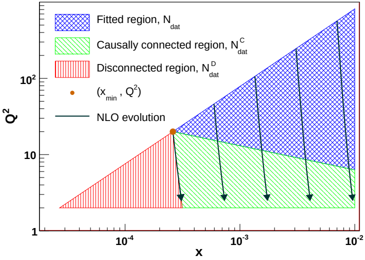


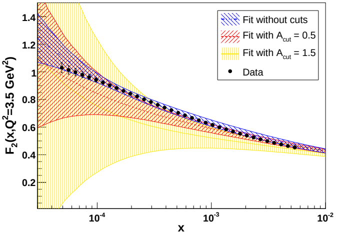


NNLO NNPDFs soon ready for LHC phenomenology

A new QCD regime in HERA data

- Deviations from NLO DGLAP in small-x small-Q^2 HERA-I data have been reported
- NNPDF finds that small–x and Q^2 data cannot be recovered by backwards DGLAP evolution when excluded from global PDF fit (arXiv:0910.3143,1007.5405)
- Prediction: NNLO should be worse than NLO for small-x HERA





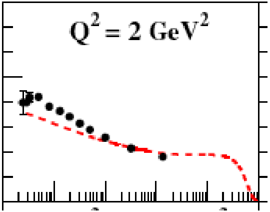
A new QCD regime in HERA data

- Deviations from NLO DGLAP in small-x small-Q^2 HERA-I data have been reported
- NNPDF finds that small–x and Q^2 data cannot be recovered by backwards DGLAP evolution when excluded from global PDF fit (arXiv:0910.3143,1007.5405)
- Prediction: NNLO should be worse than NLO for small-x HERA
- Independent confirmation from HERAPDF (H1 prelim-10-044)

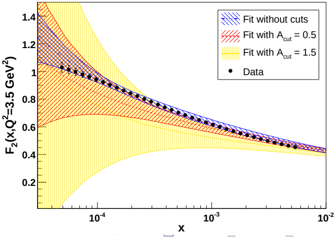
$\alpha_S(M_Z)$	NNLO fits	without low energy data
0.1176	$\Delta\chi^2$	64
0.1145	$\Delta\chi^2$	49

New QCD regime: Non–linear dynamics? High–energy resummation?

HERAPDF



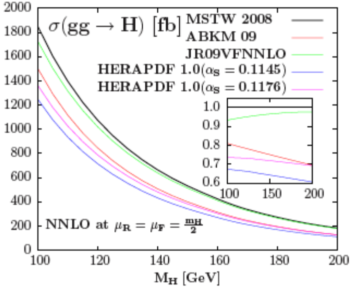
NNPDF

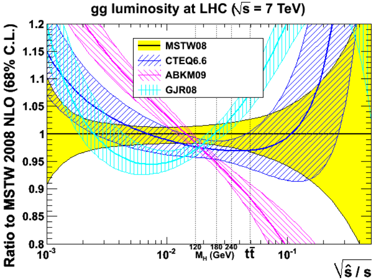


Outline

PDFs and Higgs production

- The Tevatron reports SM Higgs exclusion bounds of $158 \leq M_H \leq 175$ GeV 95% C.L., theory prediction used MSTW 2008 NNLO
- Challenged by Djoaudi et al. (arXiv:1101.1832): ABKM09 and HERAPDF lead to cross sections smaller by $\sim$ 20–50% due to both smaller α_s ($\alpha_s^{\mathrm{ABKM}} = 0.1135$, $\alpha_s^{\mathrm{MSTW}} = 0.1171$) and smaller gg lumi





The Tevatron excluded mass range should be reopened?

PDFs and Higgs production

- The Tevatron reports SM Higgs exclusion bounds of $158 \leq M_H \leq 175$ GeV 95% C.L., theory prediction used MSTW 2008 NNLO
- Challenged by Djoaudi et al. (arXiv:1101.1832): ABKM09 and HERAPDF lead to cross sections smaller by $\sim$ 20–50% due to both smaller α_s ($\alpha_s^{\rm ABKM} = 0.1135$, $\alpha_s^{\rm MSTW} = 0.1171$) and smaller gg lumi but ....
- ... ABKM09 and HERAPDF do not include Tevatron Run II jet data, crucial to pin down large-x gluon

χ^2 Description of CDF Run-II inclusive jet data

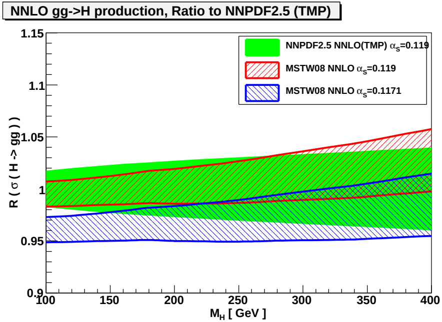
NNLO PDF	$\mu = p_T/2$	$\mu = p_T$	$\mu = 2p_T$
MSTW08	1.39 (0.42)	0.69 (0.44)	0.97 (0.48)
HERAPDF1.0 ($\alpha_S = 0.1145$)	2.64 (0.36)	2.15 (0.36)	2.20 (0.46)
HERAPDF1.0 ($\alpha_S = 0.1176$)	2.24 (0.35)	1.17 (0.32)	1.23 (0.31)
ABKM09	2.55 (0.82)	2.76 (0.89)	3.41 (1.17)

G. Watt, Les Houches QCD 2011

Non-global PDF sets: poor description of jet data

PDFs and Higgs production

- The Tevatron reports SM Higgs exclusion bounds of $158 \leq M_H \leq 175$ GeV 95% C.L., theory prediction used MSTW 2008 NNLO
- Challenged by Djoaudi et al. (arXiv:1101.1832): ABKM09 and HERAPDF lead to cross sections smaller by $\sim$ 20–50%.
- Instead, the two NNLO global fits, MSTW08 and NNPDF2.5, that include Tevatron jets, are in very good agreement



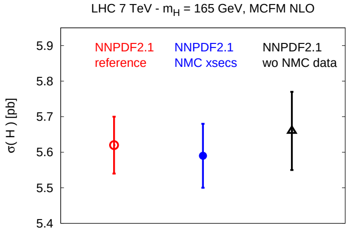
NNPDF2.5/MSTW08 agree for the NNLO Higgs at $\pm 5\%$
Tevatron exclusion limits solid!

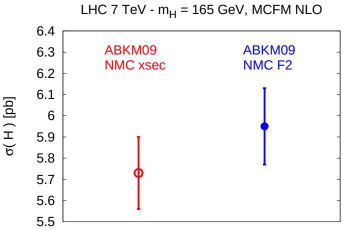
NMC data and Higgs production

- ABKM report a 3(1)–σ shift at NNLO (NLO) on the Higgs production cross section in gluon fusion at the LHC (and Tevatron) (arXiv:1101.5261)
- Claim is different treatment of fixed target DIS NMC data: used as structure functions (MSTW, NNPDF, CT) or cross sections (ABKM) $\to$ Origin of ABKM/MSTW discrepancy?

$$\widetilde{\sigma}(x,y,Q^2) = F_2(x,Q^2)\left(2-2y+y^2/\left[1+R\left(x,Q^2\right)\right]\right)$$

- NNPDF finds negligible impact of the treatment of NMC data for Higgs production, both at NLO and at NNLO



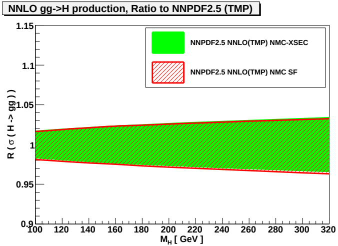


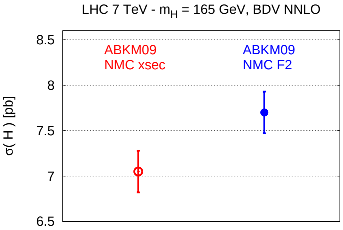
NMC data and Higgs production

- ABKM report a 3(1)–σ shift at NNLO (NLO) on the Higgs production cross section in gluon fusion at the LHC (and Tevatron) (arXiv:1101.5261)
- Claim is different treatment of fixed target DIS NMC data: used as structure functions (MSTW, NNPDF, CT) or cross sections (ABKM) → Origin of ABKM/MSTW discrepancy?

$$\widetilde{\sigma}(x, y, Q^2) = F_2(x, Q^2)\left(2 - 2y + y^2/\left[1 + R\left(x, Q^2\right)\right]\right)$$

- NNPDF finds negligible impact of the treatment of NMC data for Higgs production, both at NLO and at NNLO

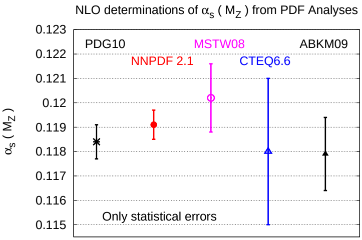


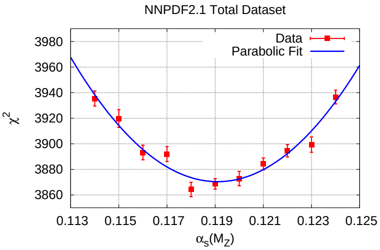


The treatment of NMC data has negligible impact on collider Higgs production
Again, the Tevatron exclusion bounds are solid!

$\alpha_s(M_Z)$ from PDF analysis

- Good: Small statistical errors from large dataset
- Bad: Bias from PDF parametrization? Dependence on dataset?
- NLO: reasonable agreement, NNPDF2.1 smallest statistical uncertainties without any theoretical bias
- Compare with current PDF average $\alpha_s^{\rm PDG}(M_Z) = 0.1184 \pm 0.0007$

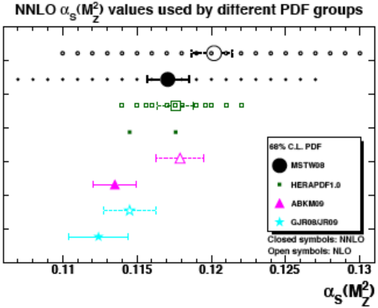




N.B.: CT, NNPDF and MSTW provide PDF sets for a wide range of $\alpha_s(M_Z)$ values

$\alpha_s(M_Z)$ from PDF analysis

- Good: Small statistical errors from large dataset
- Bad: Bias from PDF parametrization? Dependence on dataset?
- NNLO: larger discrepancies, α_s lower than $\alpha_s^{\rm PDG}$
- Is it meaningful to use $\alpha_s\,(M_Z) = 0.1135$ in LHC phenomenology?



G. Watt 2011

Crucial to provide PDF sets with varying $\alpha_s\,(M_Z)$
This includes $\alpha_s^{\rm PDG}\,(M_Z)$ for reliable LHC phenomenology

The Tevatron W lepton asymmetry

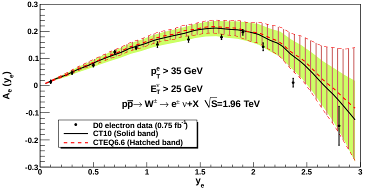
- Accurate Run-II Tevatron data on the $p\bar{p} \to W^{\pm} \to l^{\pm}\nu$ asymmetry available

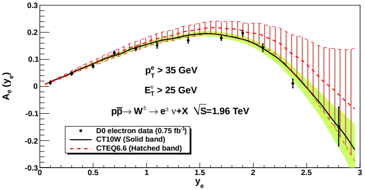
$$A_{\ell}(y_{\ell}) = \frac{d\sigma(W^{+})/dy_{\ell} \; - \; d\sigma(W^{-})/dy_{\ell}}{d\sigma(W^{+})/dy_{\ell} \; + \; d\sigma(W^{-})/dy_{\ell}}$$

- Some groups (CT10, MSTW) report tension with other datasets, specially DIS deuteron data. Problem with deuteron nuclear corrections?
- CT10 produced two sets, CT10 and CT10W, with and without D0 W asy data (arXiv:1007.2241)

ℓ	Cut (GeV)	Points	χ^2 CT10	CT10W
e	$p_T^{\ell} > 25$	12	79.5	25.3
e	$25 < p_T^{\ell} < 35$	12	20.7	25.5
e	$p_T^{\ell} > 35$	12	91.4	26.5
μ	$p_T^{\ell} > 20$	9	8.3	13.5

What do we do with Tevatron Wasy data?



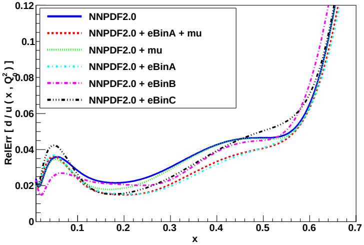


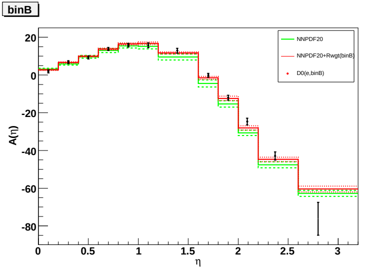
The Tevatron W lepton asymmetry

- Accurate Run-II Tevatron data on the $p\bar{p} \to W^{\pm} \to l^{\pm}\nu$ asymmetry available
- Some groups report tension with other datasets, specially DIS deuteron data. Problem with deuteron nuclear corrections?
- NNPDF (arXiv:1012.0836): D0 $\mu + e$ inclusive data (not binned in p_t^l) perfectly consistent with global fit, reduction in valence PDF uncertainty + no tension exclusive to deuteron data found

Set	χ^2	$\chi^2_{\rm rw}$	$\chi^2_{\rm tot}$
D0 μ (E_T > 20 GeV)	0.62	0.51	1.14
D0 e bin A (E_T > 25 GeV)	2.12	1.55	1.13
D0 μ + e bin A	1.44	1.11	1.13
D0 e bin B (25< E_T < 35 GeV)	4.75	1.12	1.16
D0 e bin C (E_T > 35 GeV)	5.06	2.51	1.16

TeV W asy data: useful constrains on PDFs
No issues with DIS deuteron data, but care with systematics for binned p_T^e data





Outline

Conclusions

- Very intense progress in PDF determinations in the last years
- Systematic benchmarking, heavy quark mass effects, many new datasets, unbiased NNPDFs, among many other theoretical and computational developements
- The LHC is performing extremely well $\longrightarrow$ Precision measurements around the corner $\longrightarrow$ Need precision PDFs to match the data accuracy
- Global PDF sets in reasonable agreement $\longrightarrow$ PDF4LHC recipee
  Dangerous to use non-global PDF sets for LHC phenomenology
- Not discussed here, but very relevant: Impact of PDFs for precision determination of ElectroWeak parameters at the LHC

Thanks for your attention!

Conclusions

- Very intense progress in PDF determinations in the last years
- Systematic benchmarking, heavy quark mass effects, many new datasets, unbiased NNPDFs, among many other theoretical and computational developements
- The LHC is performing extremely well $\longrightarrow$ Precision measurements around the corner $\longrightarrow$ Need precision PDFs to match the data accuracy
- Global PDF sets in reasonable agreement $\longrightarrow$ PDF4LHC recipee
  Dangerous to use non-global PDF sets for LHC phenomenology
- Not discussed here, but very relevant: Impact of PDFs for precision determination of ElectroWeak parameters at the LHC

Thanks for your attention!

Outline

Bayesian reweighting

- Given the probability distribution of the PDFs represented by $N_{\rm rep}$ instances, $\{\rm PDF_i\},\ i=1,N_{rep}$
- Given a set of new experimental data (or simulated pseudo-data) $y=\{y_1,y_2,\cdots,y_n\}$
- Given a general functional of the PDFs, $\mathcal{O}\left[\{\rm PDF\}\right]$

The impact of the new data y on the PDFs can be determined using Bayesian inference without refitting : PDF reweighting

$$\langle\mathcal{O}\rangle_{\rm old}=\sum_{k=1}^{N_{\rm rep}}\frac{1}{N_{\rm rep}}\mathcal{O}[f_k]\quad\rightarrow\quad\langle\mathcal{O}\rangle_{\rm new}=\sum_{k=1}^{N}w_k\,\mathcal{O}[f_k]$$

$$w_k\equiv\mathcal{N}_\chi(\chi^2(y,f_k))^{n/2-1}e^{-\frac{1}{2}\chi^2(y,f_k)},$$

$\chi^2(y,f_k)\rightarrow\chi^2$ of new data for the prediction of $\rm PDF_k$

No need of any refitting!

Bayesian reweighting

- Given the probability distribution of the PDFs represented by $N_{\rm rep}$ instances, $\{\rm PDF_i\}$, $\rm i = 1, N_{rep}$
- Given a set of new experimental data (or simulated pseudo-data) $y = \{y_1, y_2, \cdots, y_n\}$
- Given a general functional of the PDFs, $\mathcal{O}\left[\{\rm PDF\}\right]$

The impact of the new data y on the PDFs can be determined using Bayesian inference without refitting : PDF reweighting

$$\langle \mathcal{O} \rangle_{\rm old} = \sum_{k=1}^{N_{\rm rep}} \frac{1}{N_{\rm rep}} \mathcal{O}[f_k] \quad \to \quad \langle \mathcal{O} \rangle_{\rm new} = \sum_{k=1}^{N} w_k \, \mathcal{O}[f_k]$$

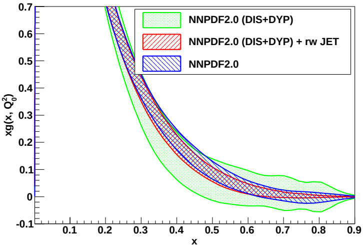
$$w_k \equiv \mathcal{N}_\chi (\chi^2(y, f_k))^{n/2-1} e^{-\frac{1}{2}\chi^2(y, f_k)},$$

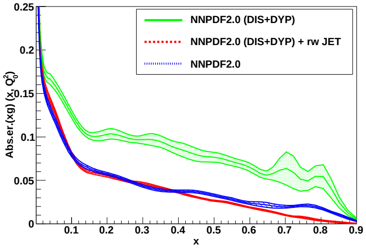
$\chi^2(y, f_k) \to \chi^2$ of new data for the prediction of $\rm PDF_k$

No need of any refitting!

Validation: Tevatron inclusive jets

- Tevatron inclusive jets constrain large-x gluon
- Compare refitting with reweighting → Statistically identical results





NNPDFs provides a *bona-fide*, statistically solid representation of the PDF uncertainties
NNPDF uncertainties behave as required by Bayes Theorem

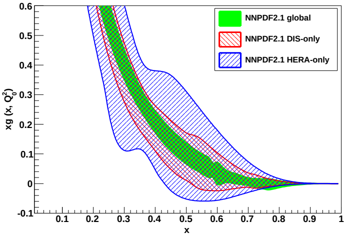
An ideal PDF set

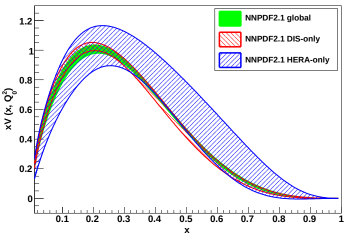
S. Forte, *Parton Distributions at the Dawn of the LHC*, arxiv:1011.5247

1. Use a wide dataset $\longrightarrow$ ensure all relevant experimental information retained
2. Use sufficiently general and unbiased parton parametrization or estimate effect of parton parametrization variations
3. **Provide PDF uncertainty bands checked to provide consistently-sized confidence levels for individual experiments**
4. Include heavy quark mass effects through a GM-VFN scheme, and provide estimate of uncertainties due to subleading terms
5. Use exact computations performed at the highest available perturbative order
6. Provide PDFs for a variety of values of $\alpha_s(M_Z)$ and for the heavy quark masses m_c and m_b

PDF sets from reduced datasets

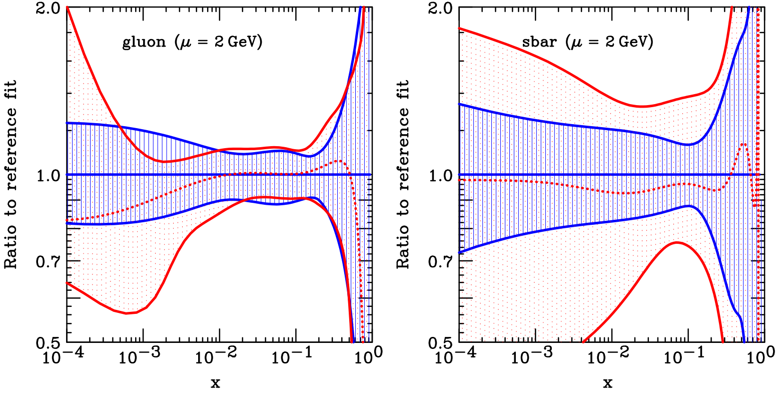
$\mathbf{xg(x, Q_0^2)}$ $\mathbf{xV(x, Q_0^2)}$





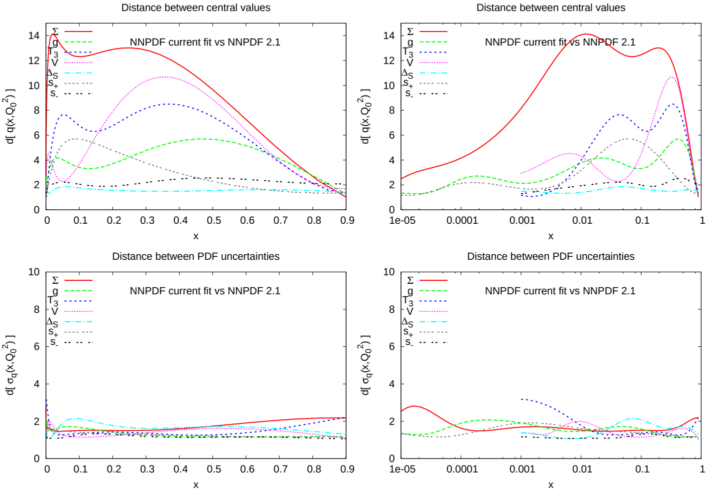
- If PDF errors are reliably estimated, PDF uncertainties consistently grow when datasets are removed
- When PDF parametrizations not flexible enough, adding new data $\longrightarrow$ larger PDF errors

PDF sets from reduced datasets



- If PDF errors are reliably estimated, PDF uncertainties consistently grow when datasets are removed
- When PDF parametrizations not flexible enough, adding new data $\rightarrow$ larger PDF errors *example* CT10 (26 params) vs. CTEQ6.6 (22 params)

Distances between NNPDF NLO and NNLO

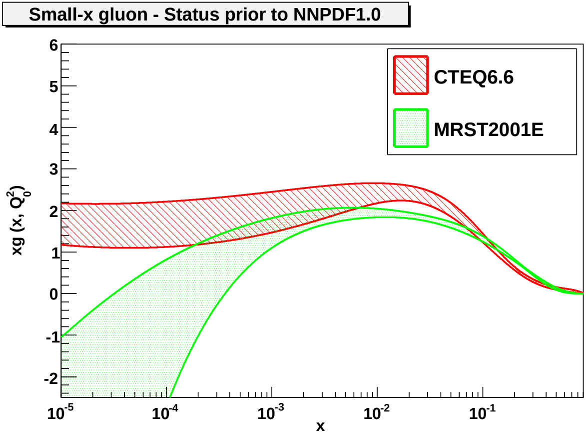


Towards PDF convergence

Global PDF sets have improved convergence in recent releases

But still some important disagreements to be understood

LOW-X GLUON

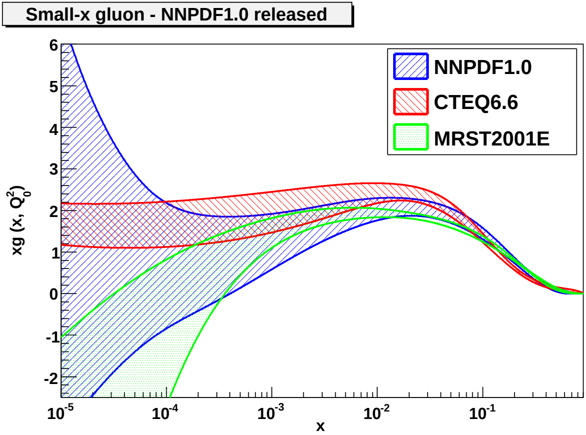


Towards PDF convergence

Global PDF sets have improved convergence in recent releases

But still some important disagreements to be understood

LOW-X GLUON

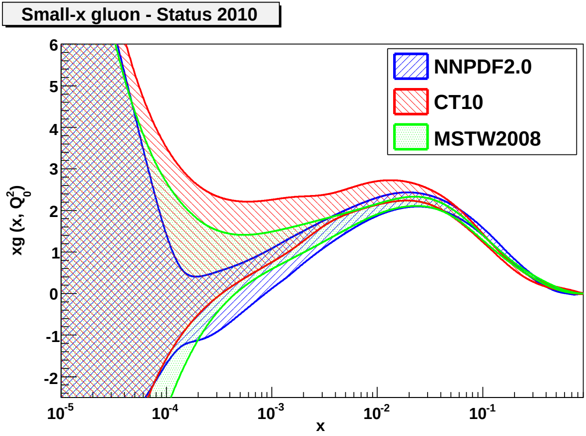


Towards PDF convergence

Global PDF sets have improved convergence in recent releases

But still some important disagreements to be understood

LOW-X GLUON



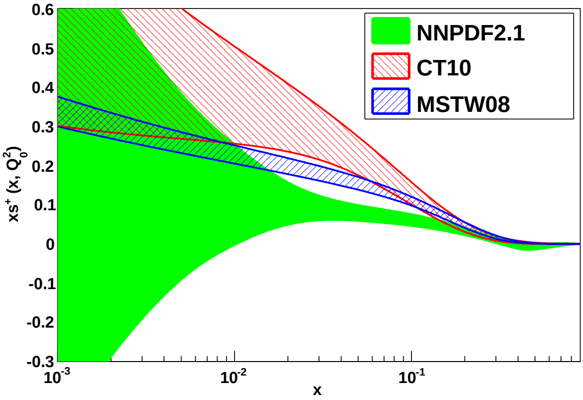
GOOD!

Towards PDF convergence

Global PDF sets have improved convergence in recent releases

But still some important disagreements to be understood

STRANGENESS



BAD! To be understood!

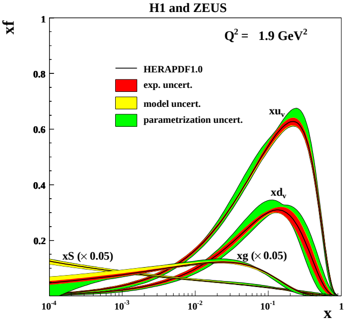
PDF uncertainties in Hessian approach

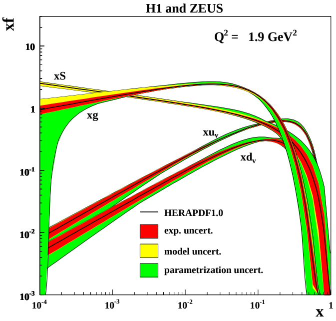
- PDF parametrization bias are an important (often dominant) source of uncertainty:

> *... parametrization error is in many places as large as the uncertainty estimated, via $\Delta\chi^2 = 10$, on the basis of conflicts observed between the different data sets used in the analysis. At very large x, ..., the parametrization error even becomes large compared to the modified $\Delta\chi^2 = 100$ uncertainty estimate used in CT10 ... These parametrization effects persist up to scales of $Q = 100$ GeV and beyond, so they are significant for predictions of important background and discovery processes at the Tevatron and LHC. (J. Pumplin, arXiv:0909.5176)*

PDF uncertainties in Hessian approach

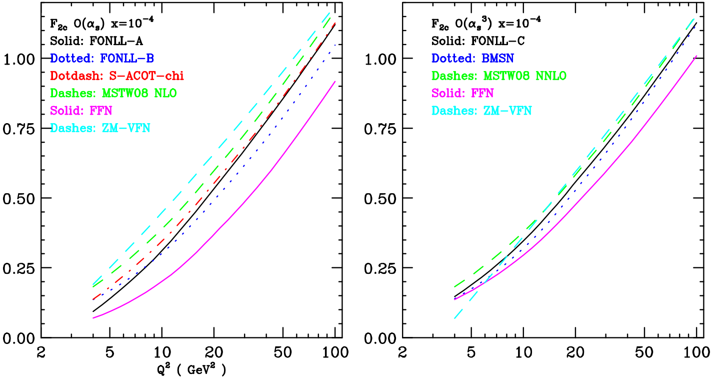
- PDF parametrization bias are an important (often dominant) source of uncertainty:
- Hessian PDF sets need to account for this effect. Two options:
  1. Effective larger tolerances $\Delta\chi^2 \gg 1$ (CT10, MSTW, JR)
  2. Explicit parametrization uncertainties (HERAPDF1.0)
- Without accounting for parametrization bias $\to$ substantial underestimation of PDF uncertainties





Heavy quark schemes for $F_2^c(x, Q^2)$

NNPDF based on FONLL-A (at NLO) and FONLL-C (at NNLO) GM-VFN
Forte, Nason, Laenen and Rojo, arxiv:1001.2312
GM-VFN interpolate between FFN at low Q^2 and ZM–VFN at high-Q^2



Benchmark comparison of GM-VFN schemes: J. Rojo et al., arXiv:1003.1241

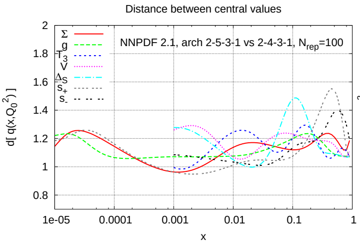
FONLL originally for hadronic collisions: Cacciari, Greco and Nason, hep-ph/9803400

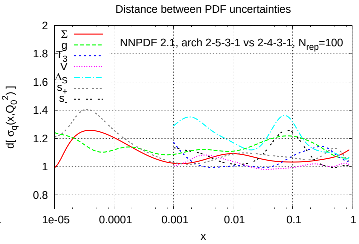
Parametrization independence

Compute distances between NNPDF2.1 with varying NN architecture
From arch. 2-5-3-1 to arch. 2-4-3-1 $\rightarrow$ From **289** to **230** parameters

$$d^2\left(\langle q^{(1)}\rangle, \langle q^{(2)}\rangle\right) = \frac{\left(\langle q^{(1)}\rangle_{(1)} - \langle q^{(2)}\rangle_{(2)}\right)^2}{\sigma^2_{(1)}[\langle q^{(1)}\rangle] + \sigma^2_{(2)}[\langle q^{(2)}\rangle]}$$

All distances $d \sim 1$ $\rightarrow$ The two PDF sets are statistically identical
(S. Forte, PHYSTAT 2011)





Impact on NuTeV anomaly

Accurate determination of xs^- $\to$ Important phenomenological implications:

NuTeV anomaly: Discrepancy ($\geq 3\sigma$) between indirect (global fit)and direct (NuTeV neutrino scattering) determinations of $\sin^2\theta_W$

EW fit

$\sin^2\theta_W = 0.2223 \pm 0.0003$

NuTeV (assumes $[S^-] = 0$)

$\sin^2\theta_W = 0.2277 \pm 0.0017$

Impact on NuTeV anomaly

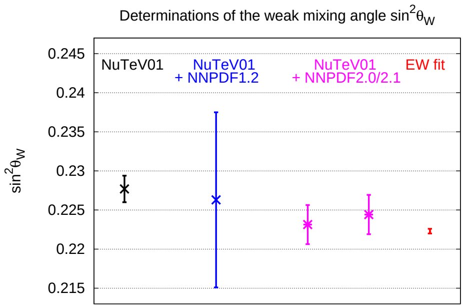
Accurate determination of xs^- $\rightarrow$ Important phenomenological implications:
NuTeV anomaly: Discrepancy ($\geq 3\sigma$) between indirect (global fit)and direct (NuTeV neutrino scattering) determinations of $\sin^2\theta_W$

EW fit

$\sin^2\theta_W = 0.2223 \pm 0.0003$

NuTeV (assumes $[S^-] = 0$)

$\sin^2\theta_W = 0.2277 \pm 0.0017$



Once NNPDF $\left[S^-\right]$ accounted for the NuTeV anomaly disappears

NuTeV + $[S^-]_{\rm nnpdf2.1}$

$\sin^2\theta_W = 0.2244 \pm 0.0025$

Direct $|V_{cs}|$ determination

- The CKM matrix element $|V_{cs}|$ is a fundamental SM parameter
- Its precision determination relevant for BSM indirect searches via CKM unitarity violations
- $|V_{cs}|$ in neutrino DIS affected by large PDF uncertainties in $s(x, Q^2)$

$$\sigma^{\nu,c} \sim |V_{cs}|^2 s(x, Q^2)$$

- PDF determinations useless for $|V_{cs}|^2$?

CKM global fit

$$V_{cs} = 0.97334 \pm 0.00023, \ \Delta V_{cs} \sim 0.02\%$$

Direct determination-D and B decays

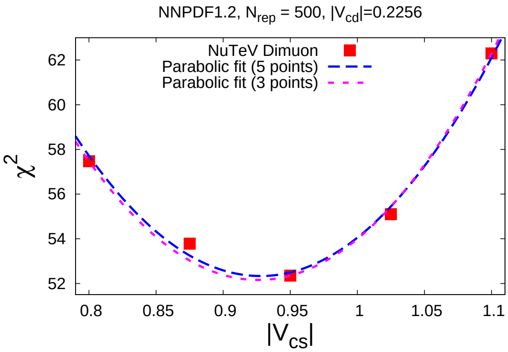
$$V_{cs} = 1.04 \pm 0.06, \qquad \Delta V_{cs} \sim 6\%$$

Direct det from $\nu-$DIS (CCFR)

$$V_{cs} \geq 0.74 \quad (90\%\mathrm{CL})$$

[PDG, Amsler et al, Phys. Lett. B67(2008) 1.]

Direct $|V_{cs}|$ determination



Direct det from $\nu-$DIS (CCFR)

$$V_{cs} \geq 0.74 \quad (90\%\mathrm{CL})$$

Direct determination-D and B decays

$$V_{cs} = 1.04 \pm 0.06, \qquad \Delta V_{cs} \sim 6\%$$

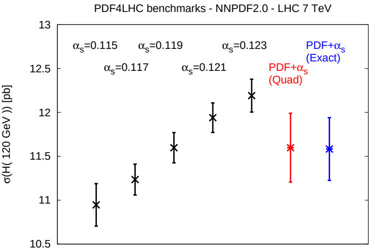
Direct det NNPDF1.2

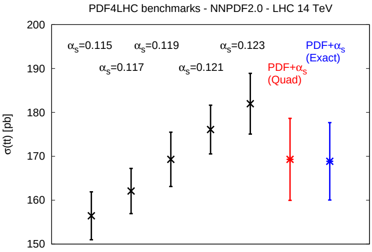
$$V_{cs} = 0.96 \pm 0.07, \ \Delta V_{cd} \sim 7\%$$

[PDG, Amsler et al, Phys. Lett. B67(2008) 1.]

- Unbiased parametrizations for PDFs allow to discriminate variations in $s^+(x)$ from variations in CKM matrix elements
- NNPDF $|V_{cs}|$ determination to be included in next PDG global average

Combined PDF+α_s uncertainties

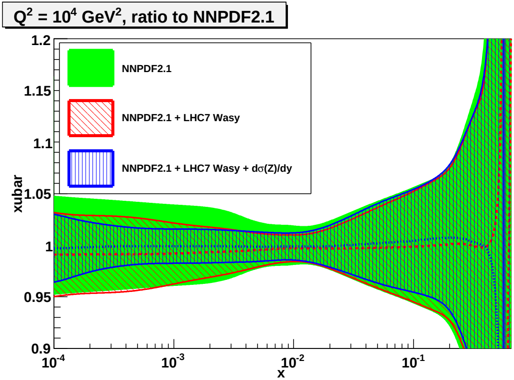




Combined PDF+α_s uncertainties important for processes sensitive to α_s at Born level ($gg \to H$, $t\bar{t}$)

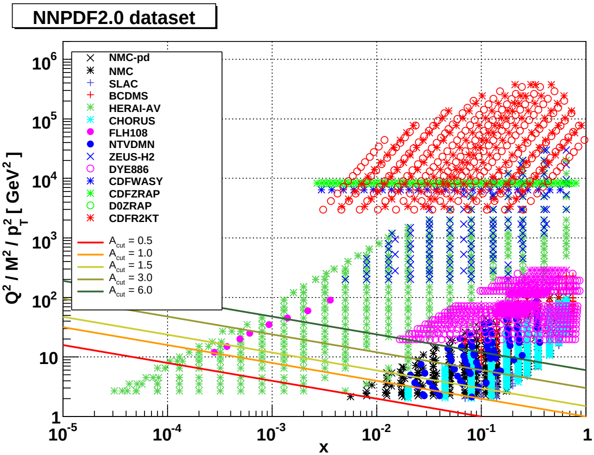
Impact of future LHC data

NNPDF allows to optimize and design dedicated LHC measurements to achieve best performance
Example: Proposal to measure machine lumi with 2% accuracy → Within the NNPDF approach we can quantify impact for precision EWK measurements ....

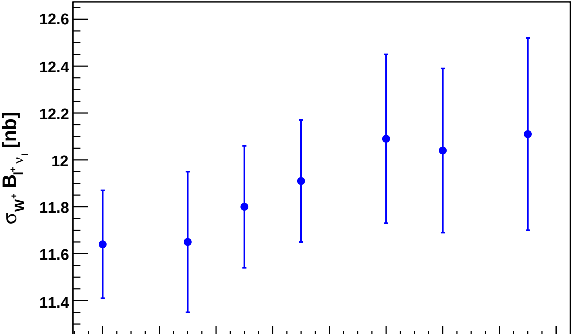


Precision Physics at the LHC, LPNHE Paris 2010

Deviations from NLO DGLAP



Deviations from NLO DGLAP



Deviations from NLO DGLAP