NNPDF and The LHC:
the way towards a reliable collider-only PDF fit

Alberto Guffanti

Niels Bohr International Academy & Discovery Center
Niels Bohr Institute - Copenhagen

On behalf of

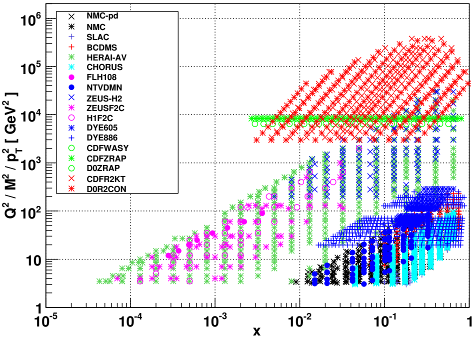
The NNPDF Collaboration
R. D. Ball, L. Del Debbio (Edinburgh), F. Cerutti, J. I. Latorre (Barcelona),
S. Forte, J. Rojo (Milano), V. Bertone(Freiburg), M. Ubiali(Aachen)

LPCC Summer Institute on LHC Physics
CERN

NNPDF2.1

A family of global fits ...

NNPDF2.1 dataset



3415 data points (NLO fit)
(3408 - LO and 3473 - NNLO)

[R. D. Ball et. al, arXiv:1101.1300] - **NLO**
[R. D. Ball et. al, arXiv:1107.2652] - **LO/NNLO**

OBS	*Data set*
Deep Inelastic Scattering	
F_2^d/F_2^p	NMC-pd
F_2^p	NMC, SLAC, BCDMS
F_2^d	SLAC, BCDMS
$\sigma_{NC}^{\pm}$	HERA-I, ZEUS (HERA-II)
$\sigma_{CC}^{\pm}$	HERA-I, ZEUS (HERA-II)
F_L	H1
$\sigma_\nu, \sigma_{\bar\nu}$	CHORUS
dimuon prod.	NuTeV
F_2^c	ZEUS, H1

Drell-Yan & Vector Boson prod.	
$d\sigma^{\rm DY}/dM^2dy$	E605
$d\sigma^{\rm DY}/dM^2dx_F$	E866
W asymm.	CDF
Z rap. distr.	D0/CDF
Inclusive jet prod.	
Incl. $\sigma^{\rm (jet)}$	CDF (k_T) - Run II
Incl. $\sigma^{\rm (jet)}$	D0 (cone) - Run II

NNPDF2.1

... based on the NNPDF Methodology

- **Monte Carlo** determination of errors
  - No need to rely on linear propagation of errors
  - Possibility to test for the impact of non gaussianly distributed errors
  - Possibility to test for non-gaussian behaviour in fitted PDFs ($1-\sigma$ vs. 68% CL)
- **Neural Networks**
  - Provide an **unbiased** parametrization
- **Stopping based on Cross-Validation**
  - Ensures proper fitting avoiding overlearning

NNPDF2.1

... including Heavy Flavour contributions (FONLL)

- We adopt the **FONLL** General Mass-Variable Flavour Number Scheme

  [M. Cacciari, M. Greco and P. Nason, (1998)]

  [S. Forte, P. Nason E. Laenen and J. Rojo, (2010)]

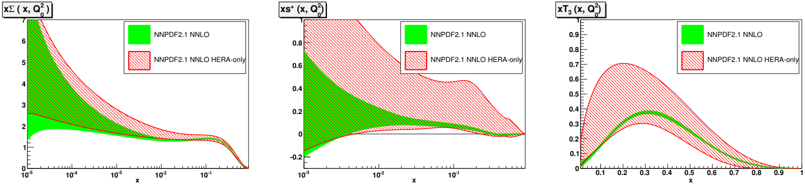
- FONLL gives a prescription to **combine FFN** (Massive) and **ZM-VFN** (Massless) computations, at any given order, **avoiding double counting**.

- With available computations three implementations of FONLL are possibile:
  - **FONLL-A**: $\mathcal{O}(\alpha_s)$ Massless + $\mathcal{O}(\alpha_s)$ Massive - (NLO fit)
  - **FONLL-B**: $\mathcal{O}(\alpha_s)$ Massless + $\mathcal{O}(\alpha_s^2)$ Massive
  - **FONLL-C**: $\mathcal{O}(\alpha_s^2)$ Massless + $\mathcal{O}(\alpha_s^2)$ Massive - (NNLO fit)

- **Fixed Flavour Number Scheme** (3-, 4-, 5-) fits **available**.

NNPDF fits to reduced datasets

- HERAPDF is not the only fit based on **HERA data only**
- HERA-only fit uses the **SAME settings** of global fit
  - No restrictions on **flavour separation**
  - No assumptions on **strangeness** being proportional to non-strange sea
  - Same unbiased **parametrization** (259 parameters)



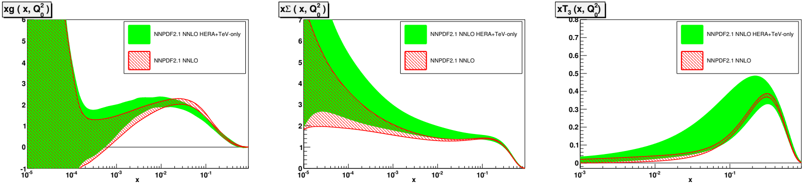
- Large uncertainties due to lack of experimental constraints and no ad-hoc parametrization choices

NNPDF fits to reduced datasets

Collider-only fit

- The fit we would love to have ...
  - Only **high energy data**: minimize effects of higher-twist contribiutions
  - Only **proton data**: no assumptions based on nuclear corrections models

- Based on **HERA and Tevatron** (inclusive jets and W/Z prduction) data



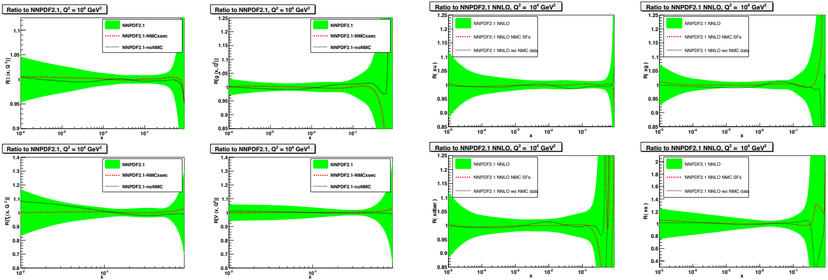
- **LHC** (and HERA-II combined) **data are crucial** in order to improve collider-only fit

Impact of NMC data on PDF determinations

... if you needed one more reason to love a collider-only fit

[R. D. Ball et al, arXiv:1102.3182v2]

- In arXiv:1101.5261, ABM suggested that the **detailed treatment of NMC data** has a **sizable impact on PDF** determinations and the **Higgs cross-section** at the Tevatron and the LHC.
- We addressed the question in the context of the **NNPDF2.1 NLO and NNLO** fits

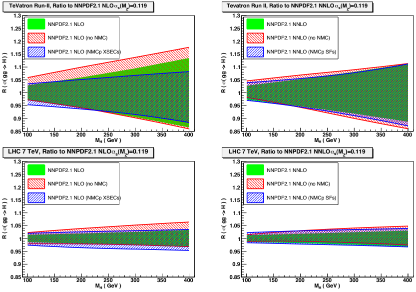


Impact of NMC data on PDF determinations

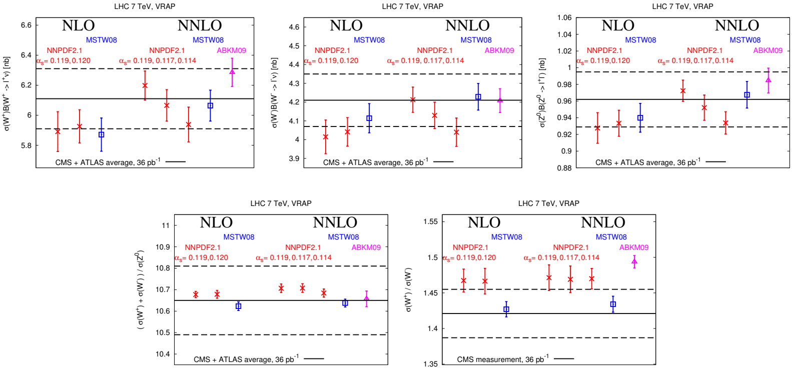
... if you needed one more reason to love a collider-only fit

[R. D. Ball et al, arXiv:1102.3182v2]

- In arXiv:1101.5261, ABM suggested that the **detailed treatment of NMC data** has a **sizable impact on PDF** determinations and the **Higgs cross-section** at the Tevatron and the LHC.
- We addressed the question in the context of the **NNPDF2.1 NLO and NNLO** fits



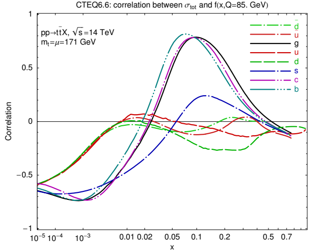
- **Standard Candles** measurements might soon have the ability to **discriminate** among **theory predictions**



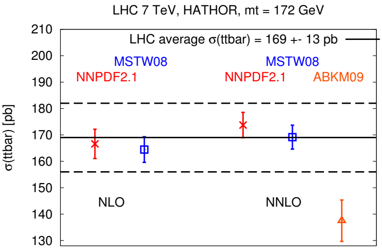
PDF4LHC

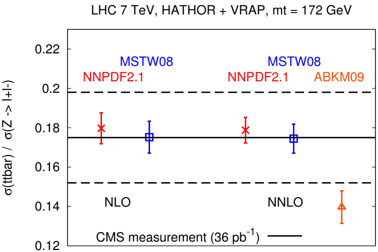
$t\bar{t}$ production

[Nadolsky et al, ArXiv:0802.0007]



- $\sigma_{t\bar{t}}$ correlated to all PDFs
- $\sigma_{t\bar{t}}$ sensitive probe of gluon
- Gluon strongly correlated to α_S





Reweighting NNPDFs

Assessing the impact of new data on PDF fits

[R. D. Ball et al., arXiv:1012.0836]

- The N_{rep} **replicas** of a NNPDF fit give the **probability density** in the space of PDFs

- **Expectation values** for observables computed as

$$\langle \mathcal{F}[f_i(x, Q^2)] \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}\left(f_i^{(net)(k)}(x, Q^2) \right)$$

(... the same is true for errors, correlations, etc.)

- We can **assess the impact** of including **new data** in the fit updating the probability density distribution without refitting.

Reweighting NNPDFs

Assessing the impact of new data on PDF fits

- According to **Bayes Theorem** we have

$$\mathcal{P}_{\rm new}(\{f\}) = \mathcal{N}_\chi \mathcal{P}(\chi^2|\{f\})\mathcal{P}_{\rm init}(\{f\})\,, \quad \mathcal{P}(\chi^2|\{f\}) = [\chi^2(y,\{f\})]^{\frac{n_{dat}-1}{2}} e^{-\frac{\chi^2(y,\{f\})}{2}}$$

- **Averages over the sample** are now **weighted sums**

$$\langle \mathcal{F}[f_i(x,Q^2)]\rangle = \sum_{k=1}^{N_{rep}} w_k \mathcal{F}\left(f_i^{(net)(k)}(x,Q^2)\right)$$

where the **weights** are

$$w_k = \frac{[\chi^2(y,f_k)]^{\frac{n_{dat}-1}{2}} e^{-\frac{\chi^2(y,f_k)}{2}}}{\sum_{i=1}^{N_{\rm rep}}[\chi^2(y,f_i)]^{\frac{n_{dat}-1}{2}} e^{-\frac{\chi^2(y,f_i)}{2}}}$$

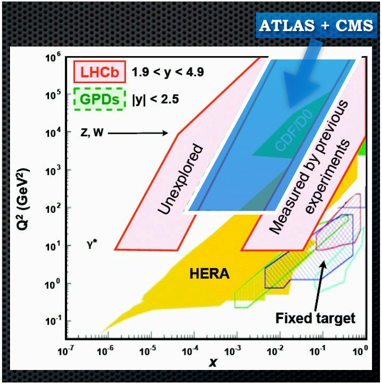
Towards a reliable collider-only fit

W lepton asymmetry data at the LHC

$$A^l_W = \frac{\sigma(pp \to W^+ \to l^+\nu_l) - \sigma(pp \to W^- \to l^-\bar{\nu}_l)}{\sigma(pp \to W^+ \to l^+\nu_l) + \sigma(pp \to W^+ \to l^-\bar{\nu}_l)}$$

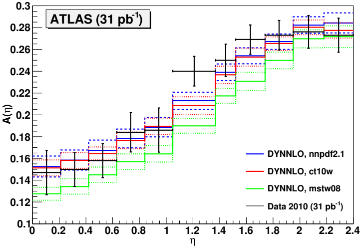
- ATLAS: muon charge asymmetry ($31\mathrm{pb}^{-1}$) [ArXiv:1103:2929]
- CMS: muon charge asymmetry ($36\mathrm{pb}^{-1}$) [ArXiv:1103:3470]
- LHCb: preliminary forward W muon charge asymmetry ($16.5\mathrm{pb}^{-1}$), not corrected for FSR

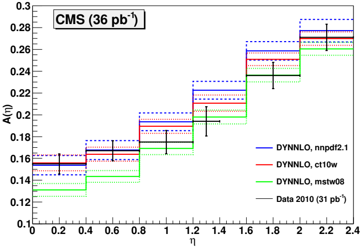
$$A^l_W \sim \frac{u(x_1, M_W^2)\bar{d}(x_2, M_W^2) - d(x_1, M_W^2)\bar{u}(x_2, M_W^2)}{u(x_1, M_W^2)\bar{d}(x_2, M_W^2) + d(x_1, M_W^2)\bar{u}(x_2, M_W^2)}$$



Towards a reliable collider-only fit

The W lepton asymmetry data at LHC





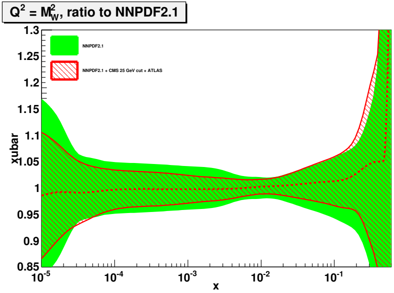
χ^2/d.o.f.	**NNPDF2.1**	**CT10w**	**MSTW08**
ATLAS	0.7	0.8	3.2
CMS e^- $p_T > 25$ GeV	1.9	0.8	2.4
CMS e^- $p_T > 30$ GeV	1.7	1.2	2.5
CMS μ $p_T > 25$ GeV	1.3	0.5	1.1
CMS μ $p_T > 30$ GeV	0.8	0.6	1.3

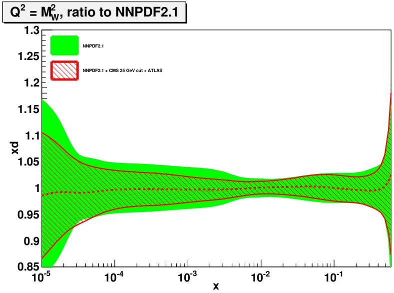
Theory predictions computed using DYNNLO at NLO

[ArXiv:0903.2120]

Towards a reliable collider-only fit

Inclusion of the LHC W lepton asymmetry data (PRELIMINARY)





- ATLAS and CMS data compatible with data included in global analysis
- The provide important constraint to PDFs in the small medium-x region
- Significant uncertainty reduction

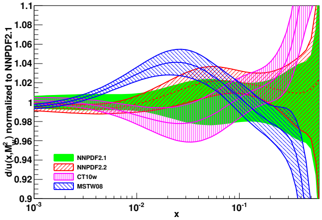
ATLAS
$N_{\mathrm{eff}} = 928$, $\chi^2_{\mathrm{d.o.f.}}$: $0.69 \rightarrow 0.65$
CMS ($p_T^l > 25$GeV)
$N_{\mathrm{eff}} = 554$, $\chi^2_{\mathrm{d.o.f.}}$: $1.41 \rightarrow 0.74$
CMS ($p_T^l > 30$GeV)
$N_{\mathrm{eff}} = 717$, $\chi^2_{\mathrm{d.o.f.}}$: $0.98 \rightarrow 0.72$

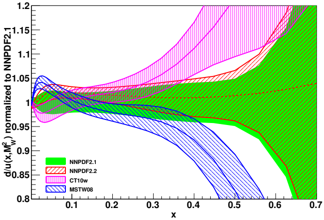
Towards a reliable collider-only fit

Combining Tevatron and LHC data (PRELIMINARY)

LOG scale

LIN scale

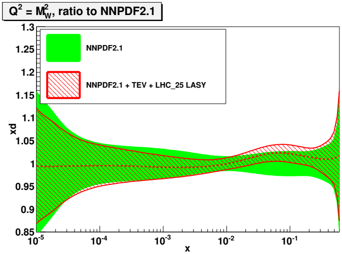




ATLAS+CMS(25)+D0$_\mu$+D0$_e$(20)

$N_{\rm eff}$ = 196, $\chi^2_{\rm d.o.f.}$: 2.18 $\rightarrow$ 0.86

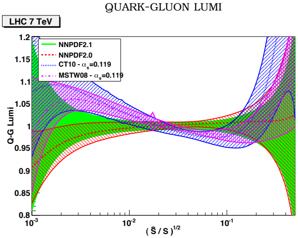
- Uncertainty reduction medium-small x and shift in central value driven by the LHC data
- Uncertainty reduction at medium-large x driven by Tevatron data



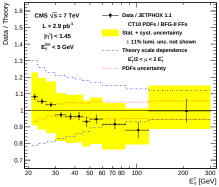
LHC4PDFs

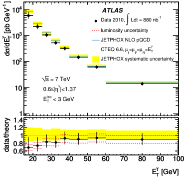
Prompt photons at the LHC

- Direct probe of *qg* luminosity
- Constraint on medium-*x* gluon ... **complementary to inclusive jets**
- Possible “problems” ... but concentrated in the **small-$p_\perp$ region**



- Measurements by ATLAS and CMS should be soon finalized





Theoretical setup: JETPHOX NLO + NNPDF2.1

- JETPHOX 1.3.0 NLO pQCD code [Guillet-Arleo]
- NNPDF2.1 (100 replicas) interfaced via LHAPDF5.8.5
- BFG-II parton-to-photon FFs (but suppressed by isolation cuts).
- All scales set to default: $\mu_R = \mu_F = \mu_{FF} = E_T^\gamma$
- Exp. kinematics+isolation cuts & p_T binnings for 30 systems:

100 replicas direct-γ NLO: ~ 7h CPU / 1M evts (~5 days for 20 Mevts !)
100 replicas frag-γ NLO: ~10h CPU / 1M evts (~1 week for 20 Mevts !)

$\times$**30 !**

- NNPDF2.1 "reweighting technique":

(1) $d\sigma_{NLO}/dp_T$ for 100 (or 1000) replicas: NNPDF21_100.LHgrid

(2) χ^2 analysis $d\sigma_{EXP}/dp_T$ - $d\sigma_{NLO}/dp_T$ for each replica.

(3) Obtain associated "weight" for each replica:

$$w_k = \frac{(\chi_k^2)^{n/2-1} e^{-\frac{1}{2}\chi_k^2}}{\frac{1}{N}\sum_{k=1}^{N} (\chi_k^2)^{n/2-1} e^{-\frac{1}{2}\chi_k^2}}.$$

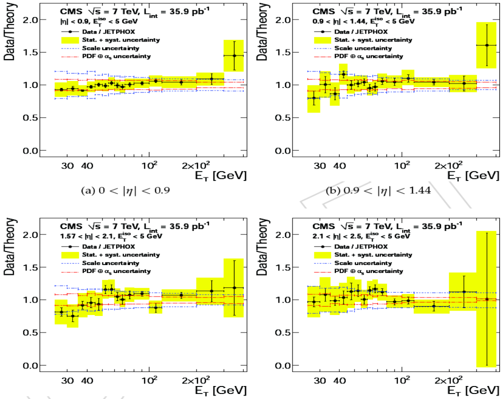
(4) Obtain reweighted PDF replicas:

$$\langle \mathcal{O} \rangle_{\rm new} = \frac{1}{N} \sum_{k=1}^{N} w_k \mathcal{O}[f_k]$$

[R.D.Ball et al. NPB 849 (2011) 112]

LHC isolated-γ vs JETPHOX (example)

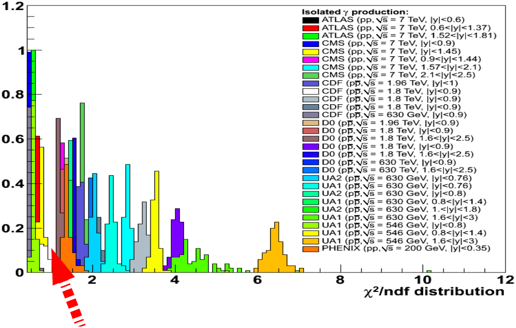
[See talk by Nicholas Chanon (CMS)]



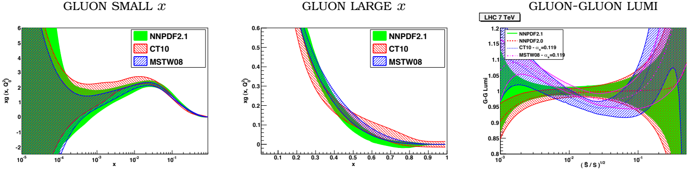
- **Excellent agreement** data-NLO for pp → γ_{isol}+X at 7 TeV

χ^2 world γ-data vs NNPDF replicas

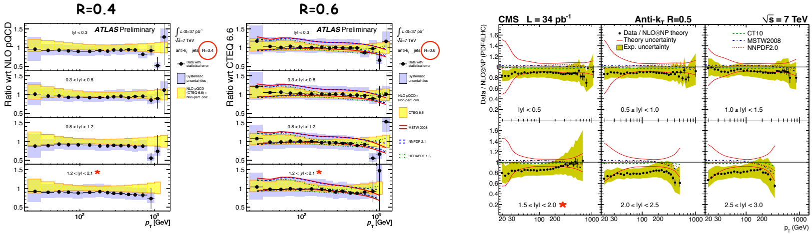
- χ^2/ndf distribution of 100 replicas for each one of 30 systems:
  (syst.+stat. uncertainties in quadrature. Lumi not considered)



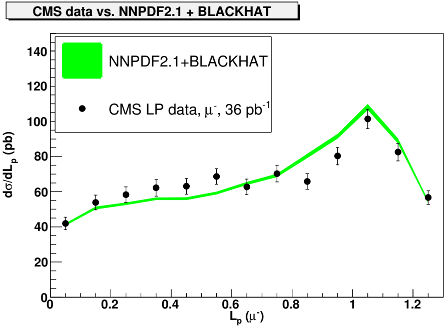
- Larger impact expected on **gluon in medium-x region** $(0.05 \leq x \leq 0.2)$



- Inconsistencies among PDF fits in this region might be resolved by precise data ... uncertainties dominated by systematics



- Can we learn something more looking at *W* **polarization**?
- **CMS measurement** available but still **uncorrected** for: background subtraction, efficiency, resolution.
- NLO calculation implemented in **BLACKHAT** (thanks to Daniel for running the code for us)



LHC4PDFs

... the data we would love to have

- Medium- and large-x **gluon**
  - Prompt photons
  - Inclusive Jets
  - t-quark distributions ($p_\perp$, y) ?

- **Light flavour separation** at medium- & small-x
  - Low-mass Drell-Yan
  - Z rapidity distribution
  - W(+jets) asymmetry

- **Strangeness & Heavy Flavours**
  - $W + c$
  - $Z + c$, $\gamma + c$
  - $Z + b$

... and if we could have them as **APPLGRID or FastNLO grids** ...
that would be really sweet! ;)