The (un)polarized strangeness from NNPDF

2^{nd} Workshop on Probing Strangeness in Hard Processes

Emanuele R. Nocera

Università degli Studi di Milano & INFN, sezione di Milano

LNF, Frascati - November, 11 2013

NNPDF

Strange distributions and fragmentation functions
Nucleon tomography (GPDs)
Quark orbital angular momentum (TMDs)
Quark hadronization in the nuclear medium
Hadron spectroscopy and search for exotic mesons

Outline

1. The NNPDF methodology
   - Monte Carlo sampling, Neural Network parametrization and reweighting
2. Unpolarized PDFs: NNPDF2.3
   - with emphasis on the strangeness content of the proton
3. Polarized PDFs: NNPDFpol1.0
   - including recent developments
4. Conclusions

DISCLAIMER

Not a systematic review of recent developments in PDF analyses

[arXiv:1301.6754] [arXiv:1211.5142] [arXiv:1306.6515]

Rather, partial view on PDFs from the NNPDF side

focusing on strangeness

1. The NNPDF methodology

The NNPDF methodology in a nutshell

1. Generate $N_{\rm rep}$ Monte Carlo replicas of the experimental data
   - multi-Gaussian probability distribution
   - take into account all experimental correlations
2. Fit a set of Parton Distribution Functions (PDFs) parametrized at initial scale
   - with Neural Networks
   - to each replica

Resulting PDF replicas are equally probable members of a statistical ensemble which samples the probability density $\mathcal{P}[f_i]$ in the space of PDFs

$$\langle \mathcal{O} \rangle = \int \mathcal{D}f_i \mathcal{P}[f_i]\mathcal{O}[f_i]$$

Expectation values for observables are Monte Carlo integrals

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

and corresponding formulae for the estimators of Monte Carlo samples are used to compute uncertainties, correlations, etc.

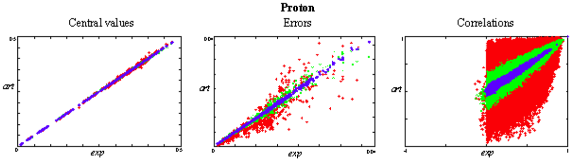
The NNPDF methodology: Monte Carlo sampling

- Monte Carlo replicas are generated according to the distribution

$$\mathcal{O}_i^{\mathrm{art}(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left[\mathcal{O}_i^{(\mathrm{exp})} + \sum_{p=1}^{N_{\mathrm{sys}}} r_p^{(k)}\sigma_{i,p} + r_{i,s}^{(k)}\sigma_{i,s}\right]$$

where $i = 1, N_{\mathrm{dat}}$, $k = 1, N_{\mathrm{rep}}$ and $r^{(k)}$ are (Gaussianly distributed) random numbers

- Monte Carlo replicas are validated against experimental data

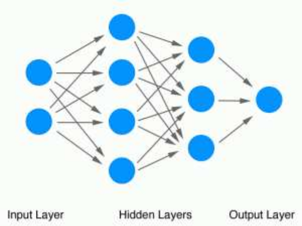


$N_{\mathrm{rep}} = 10$: red
$N_{\mathrm{rep}} = 100$: green
$N_{\mathrm{rep}} = 1000$: blue

$N_{\mathrm{rep}} \sim 100$ to reproduce central values and errors, $N_{\mathrm{rep}} \sim 1000$ to reproduce correlations to percent accuracy

No need to rely on **linear propagation** of errors
Possibility to test the impact of **non-Gaussianly** distributed **uncertainties**
Possibility to test for **non-Gaussian behaviour** of uncertainties of fitted **PDFs**

- Neural Networks provide us with a redundant parametrization for PDFs
- Only require smoothness of the fitted function
- In the end, they are just an other basis of functions



$$\xi_i^{(l)} = g\left(\sum_j^{n_l - 1} \omega_{ij}^{(l-1)} \xi_j^{(l-1)} - \theta_i^{(l)}\right)$$

$$g(x) = \frac{1}{1+e^{-x}}$$

- Use genetic algorithms for minimization owing to high dimensional parameter space
- Avoid to fit statistical fluctuations (overlearning) with proper stopping criterion

Neural Networks **reduce** the theoretical **bias** due to the parametrization
Minimization and **stopping** algorithms ensure **proper** PDF **fitting**

The NNPDF methodology: reweighting [arXiv:1012.0836] [arXiv:1108.1758]

- We would like to assess the impact of including a new data set $\{y\} = \{y_1, \ldots, y_n\}$ (delivered with σ_{ij}) in a prior ensemble of PDF replicas $\{f_k\}$, $k = 1, \ldots, N_{\rm rep}$
- We can apply Bayes theorem to determine the conditional probability of PDF upon inclusion of the new data and update the probability density in the space of PDFs

$$\mathcal{P}_{\rm new} = \mathcal{N}_\chi \mathcal{P}(\chi^2_k|\{f_k\})\mathcal{P}_{\rm old}(\{f_k\}) \quad \mathcal{P}(\chi^2_k|\{f_k\}) = [\chi^2_k(\{y\},\{f_k\}]^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi^2_k(\{y\},\{f_k\})}$$

$$\chi^2_k(\{y\},\{f_k\}) = \sum_{i,j}^{n} \{y_i - y_i[f_k]\}\,\sigma_{ij}\,\{y_j - y_j[f_k]\}$$

- Replicas are no longer equally probable. Expectation values are given by

$$\langle \mathcal{O}[f_i(x,Q^2]\rangle_{\rm new} = \sum_{k=1}^{N_{\rm rep}} w_k \mathcal{O}[f_i^{(k)}(x,Q^2)]$$

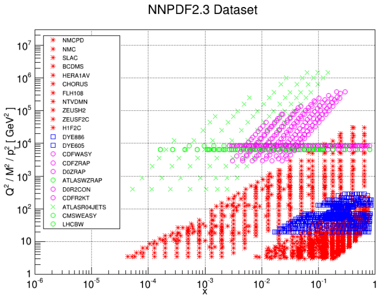
$$w_k \propto [\chi^2_k(\{y\},\{f_k\})]^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi^2_k(\{y\},\{f_k\})} \quad \text{with} \quad N_{\rm rep} = \sum_{k=1}^{N_{\rm rep}} w_k$$

Reweighting allows to incorporate new datasets **without** need of **refitting**

2. Unpolarized PDFs: NNPDF2.3

[arXiv:1207.1303]

Experimental data set



Experiment	$N_{\rm dat}$
Fixed-target DIS	926
Fixed-target neutrino DIS	1026
HERA DIS	834
Fixed-target DY	318
Tevatron W/Z	70
Tevatron Jets	186
LHC W/Z	56
LHC Jets	90

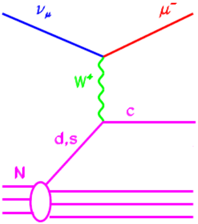
3506 data points

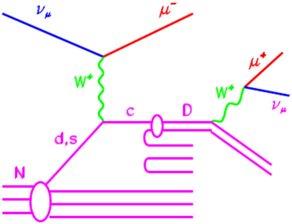
- New data in 2.3
  - **ATLAS 2010** Inclusive Jets, 36 pb^{-1} [arXiv:1112.6297]
  - **ATLAS 2010** $W^{\pm}/Z$ distributions, 36 pb^{-1} [arXiv:1109.5141]
  - **LHCb 2010** $W^{\pm}$ rapidity distributions, 36 pb^{-1} [arXiv:1204.1620]
  - **CMS 2011** W lepton asymmetry, 840 pb^{-1} [arXiv:1206.2598]
- NNPDF2.3 family
  - **NNPDF2.3** - global data set (including LHC data)
  - **NNPDF2.3 Collider** - HERA, Tevatron and LHC data sets only
  - **NNPDF2.3 QED** - also includes QED corrections and extraction of photon PDF
  - ...

Which experimental data sets can constrain strangeness?

Data included in NNPDF2.3

1. Inclusive neutrino DIS (CHORUS [Phys. Lett. B632 (2006) 65])
   $\nu_\mu(\bar{\nu}_\mu)N \to \mu^-(\mu^+)X$
2. Dimuon production in DIS (NuTev [arXiv:hep-ex/0102049])
   $\nu_\mu(\bar{\nu}_\mu)N \to \mu^-\mu^+(\mu^+\mu^-)X$





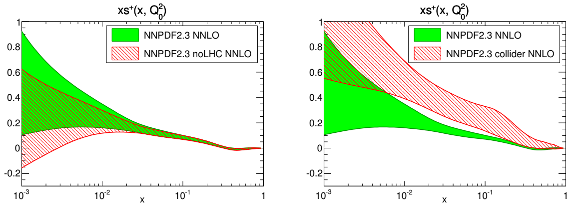
Data not included in NNPDF2.3

1. W boson production in association with charm at LHC (CMS [arXiv:1310.1138])
   $pp \to W + c + X$
2. Kaon electroproduction data from SIDIS on deuteron (HERMES [arXiv:0803.2993])
   see talk by K. Rith

The strange content of the proton

Both total $s^+ = s + \bar{s}$ and valence $s^- = s - \bar{s}$ strangeness are extracted from data
Only MSTW08 [arXiv:0901.0002] also provides independent parametrization of s^-

$s^+ = s + \bar{s}$



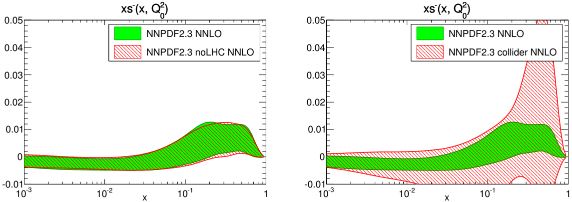
- Moderate impact of LHC data on strangeness (both s^+ and s^-)
- **NNPDF2.3 Collider** prefers larger values for total strangeness, with larger uncertainties (lack of neutrino DIS data which provide in turn more constraint on strangeness)
- No experimental constraints below $x \sim 10^{-2}$
- Theoretical constraint from **valence sum rule** $\int_0^1 dx s^-(x) = 0$ (does not imply symmetric strangeness $s(x) = \bar{s}(x)$)

A **global fit** to all available data **is needed** to determine both s^+ and s^-

The strange content of the proton

Both total $s^+ = s + \bar{s}$ and valence $s^- = s - \bar{s}$ strangeness are extracted from data
Only MSTW08 [arXiv:0901.0002] also provides independent parametrization of s^-

$s^- = s - \bar{s}$

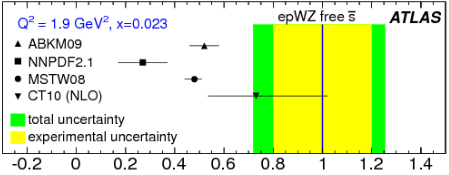


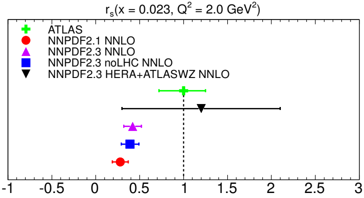
- Moderate impact of LHC data on strangeness (both s^+ and s^-)
- **NNPDF2.3 Collider** prefers larger values for total strangeness, with larger uncertainties (lack of neutrino DIS data which provide in turn more constraint on strangeness)
- No experimental constraints below $x \sim 10^{-2}$
- Theoretical constraint from **valence sum rule** $\int_0^1 dx s^-(x) = 0$ (does not imply symmetric strangeness $s(x) = \bar{s}(x)$)

A **global fit** to all available data **is needed** to determine both s^+ and s^-

The strange content of the proton

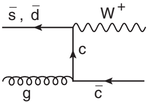
$$r_s(x, Q^2) = \frac{s(x, Q^2) + \bar{s}(x, Q^2)}{2\bar{d}(x, Q^2)}$$

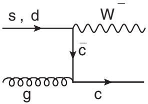




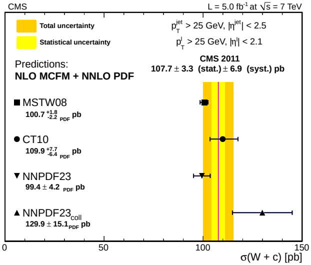
- Evidence for strangeness larger than hitherto thought from ATLAS [arXiv:1203.4051]
- This result, based on HERA+ATLAS data, is inconsistent at the 2σ level with **NNPDF2.1**
- **NNPDF2.3 HERA+ATLASWZ** finds that these data sets cannot determine strangeness with sufficient accuracy to make any reliable conclusion
- The uncertainty of the ATLAS result is likely to be underestimated
- Strangeness in **NNPDF2.3** is accurately determined, though LHC data have slight impact

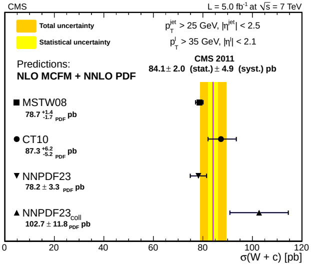
The strange content of the proton





- $W + c$ measurements are directly sensitive to separated s and $\bar{s}$: CMS [arXiv:1310.1138]
- Few percent contribution from (Cabibbo-suppressed) $dg \to W^- + c$ and $\bar{d}g \to W^+ + \bar{c}$



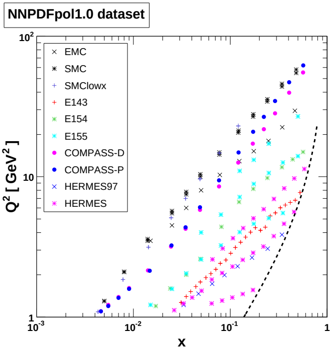


- Nice consistency between measured total cross sections and theoretical expectations
- Larger uncertainties from **NNPDF2.3 Collider** arise from the lack of direct constraints on strangeness in a collider-only fit (no neutrino DIS charm production data are included)

3. Polarized PDFs: NNPDFpol1.0

[arXiv:1303.7236]

Experimental data set



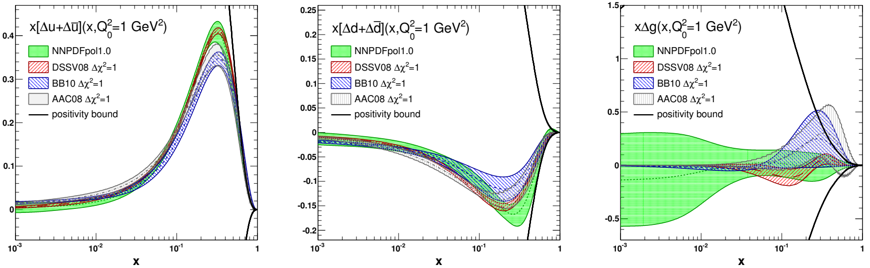
- Include all available data on inclusive DIS structure function $g_1^{p,d,n}$
- Kinematical cut imposed to remove sensitivity to dynamical higher-twist constributions $W^2 = Q^2(1-x)/x \geq 6.25$ GeV^2 no JLab data included (see [arXiv:1310.3734])
- Initial scale $Q_0^2 = 1$ GeV^2, data included down to $Q^2 > 1$ GeV^2

245 data points

Inclusive (NC) DIS does not allow us to disentangle the contributions from q and $\bar{q}$
Choose a basis of four polarized PDFs (gluon + linear combinations of light quarks)
e.g. $\{\Delta\Sigma; \Delta T_3; \Delta T_8; \Delta g\}$ or $\{\Delta u + \Delta\bar{u}; \Delta d + \Delta\bar{d}; \Delta s + \Delta\bar{s}; \Delta g\}$

Positivity and integrability play a substantial role due to loose constraints from data

The NNPDFpol1.0 parton set at $Q_0^2 = 1$ GeV^2



$\Delta u + \Delta \bar{u}$ and $\Delta d + \Delta \bar{d}$

- Central values in reasonable agreement with those of other parton sets
- Uncertainties slightly larger for NNPDF than for other sets (notice that DSSV08 fit is based on a much wider dataset)
- Where no data/theoretical constraints available, uncertainties are larger

Δg

- Central value compatible with zero
- Uncertainty much larger than any other set, especially in the low-x region

DSSV08
[arXiv:0904.3821]
DIS+SIDIS+pp (π^0/jet)

BB10
[arXiv:1005.3113]
DIS only

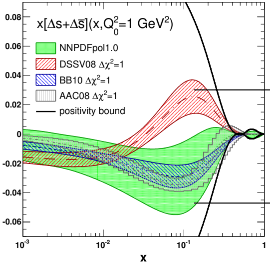
AAC08
[arXiv:0808.0413]
DIS+pp (π^0)

NNPDFpol1.0 [arXiv:1303.7236]
$\int_0^1 dx[\Delta s + \Delta\bar{s}] = -0.13 \pm 0.09$

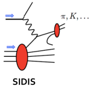
Lattice [arXiv:1112.3354]
$\int_0^1 dx[\Delta s + \Delta\bar{s}] = -0.020(10)(1)$

First moment constrained by
$a_3 = \int_0^1 dx[\Delta u^+ - \Delta d^+] = 1.2701 \pm 0.0025$

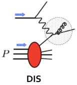
$a_8 = \int_0^1 dx[\Delta u^+ + \Delta d^+ - 2\Delta s^+] = 0.585 \pm 0.025$



directly from SIDIS Kaon data



indirectly from DIS scaling violations



$\Delta s + \Delta\bar{s}$

- Good agreement with BB10 and AAC08, but larger uncertainty for almost all x values
- Inconsistency at 2σ with DSSV08 ($\Delta\chi^2 = 1$) in the medium-small x region (SIDIS data)

Is there mounting tension between DIS and SIDIS data?

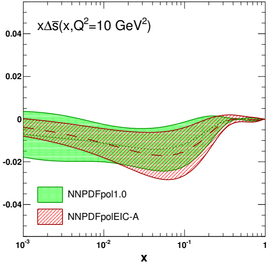
- Inclusion of SIDIS data requires the knowledge of the fragmentation $s \to K$
  $\longrightarrow$ how well do we know the kaon fragmentation function?
- Baryon octet decay constants fix the first moments of non-singlet PDF combinations
  $\longrightarrow$ conservative 30% uncertainty estimation on a_8 to allow for $SU(3)$ symmetry violation
  $\longrightarrow$ lattice finds large $SU(3)$ symmetry violation [arXiv:1112.3354]

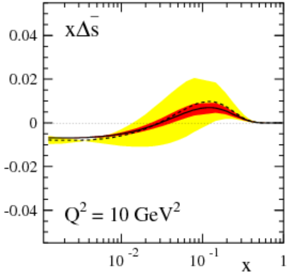
May an EIC help to constrain strangeness?

We would like to assess the impact of EIC data on PDFs

- Include DIS pseudodata at an EIC in our QCD analysis
- Pseudodata generated from DSSV+ global fit
- RHIC option of an EIC:
  $E_p = 100 - 200$ GeV, $E_e = 5$ GeV, $\mathcal{L} = 10$ fb^{-1}, 70% beam polarization

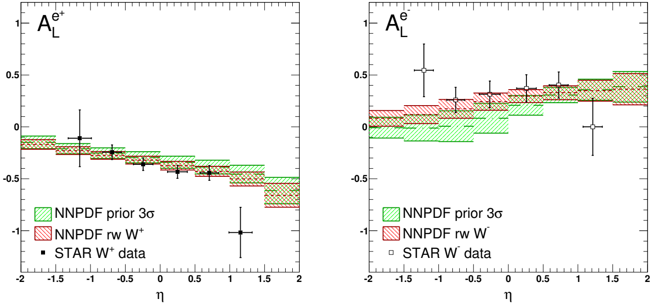
1. NNPDF [arXiv:1310.0461]
   - Only inclusive DIS was considered
   - The shape of the strangeness does not change
   - Slight reduction of the uncertainty at low-x values
   - Other PDFs show more significant improvement
     particularly the gluon, which acquires shape dictated by pseudodata
2. DSSV [arXiv:1206.6014]
   - Semi-inclusive DIS was also considered
   - Substantial reduction of the uncertainty
     all uncertainties are shown for $\Delta\chi^2 = 9$
   - The shape is dictated by SIDIS kaon pseudodata
     important role of parton to Kaon fragmentation
     see talks on Wednesday morning





Sea quarks from W boson production at RHIC

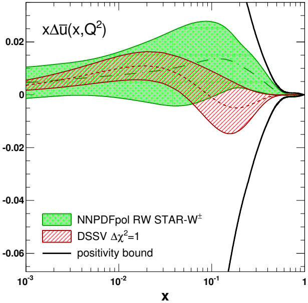
$$A_L^{W^+} \sim \frac{-\Delta u(x_1)\bar{d}(x_2) + \Delta\bar{d}(x_1)u(x_2)}{u(x_1)\bar{d}(x_2) + \bar{d}(x_1)u(x_2)} \qquad A_L^{W^-} \sim \frac{-\Delta d(x_1)\bar{u}(x_2) + \Delta\bar{u}(x_1)d(x_2)}{d(x_1)\bar{u}(x_2) + \bar{u}(x_1)d(x_2)}$$

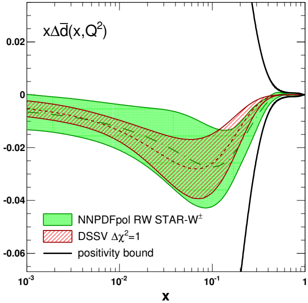


- Provide an handle on Δu, $\Delta\bar{u}$, Δd, $\Delta\bar{d}$ separation, idependent from SIDIS (purely weak process coupling q_L with $\bar{q}_R$ at partonic level, $u_L\bar{d}_R \to W^+$ or $d_L\bar{u}_R \to W^-$)
- Preliminary results from STAR [arXiv:1302.6639] included using Bayesian reweighting

Sea quarks from W boson production at RHIC

$$A_L^{W^+} \sim \frac{-\Delta u(x_1)\bar{d}(x_2) + \Delta\bar{d}(x_1)u(x_2)}{u(x_1)\bar{d}(x_2) + \bar{d}(x_1)u(x_2)} \qquad A_L^{W^-} \sim \frac{-\Delta d(x_1)\bar{u}(x_2) + \Delta\bar{u}(x_1)d(x_2)}{d(x_1)\bar{u}(x_2) + \bar{u}(x_1)d(x_2)}$$





- $W^\pm$ data are effective in constraining light sea quark distributions $\Delta\bar{u}$ and $\Delta\bar{d}$
- Reweighting with separate W^+ and W^- datasets shows that $\Delta\bar{u}$ and $\Delta\bar{d}$ behaviour is driven by experimental information from W^- and W^+ respectively
- All other PDFs are almost unaffected (including strangeness)

4. Conclusions

Conclusions

1. Unpolarized PDFs:
   - Strangeness is mainly constrained by neutrino DIS data (inclusive and dimuon)
   - LHC $W + c$ data will allow for a direct extraction of separate s and $\bar{s}$ distributions
   - Ongoing analysis towards NNPDF3.0 including new data sets from LHC
2. Polarized PDFs:
   - Inclusive DIS data only provide indirect constraint on strangeness
   - Further constraints are provided by SIDIS
     (but we need an accurate determination of kaon FF)
3. Most updated NNPDF ensembles available within LHAPDF interface or at

   `https://nnpdf.hepforge.org/`

   together with stand-alone Fortran/C++/Mathematica code for handling NNPDFs

Conclusions

1. Unpolarized PDFs:
   - Strangeness is mainly constrained by neutrino DIS data (inclusive and dimuon)
   - LHC $W + c$ data will allow for a direct extraction of separate s and $\bar{s}$ distributions
   - Ongoing analysis towards NNPDF3.0 including new data sets from LHC
2. Polarized PDFs:
   - Inclusive DIS data only provide indirect constraint on strangeness
   - Further constraints are provided by SIDIS (but we need an accurate determination of kaon FF)
3. Most updated NNPDF ensembles available within LHAPDF interface or at

   `https://nnpdf.hepforge.org/`

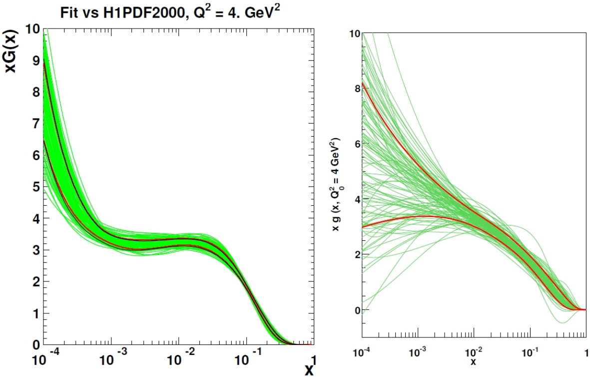
   together with stand-alone Fortran/C++/Mathematica code for handling NNPDFs

Thank you for your attention!

5. Backup

Simple functional forms vs Neural Networks

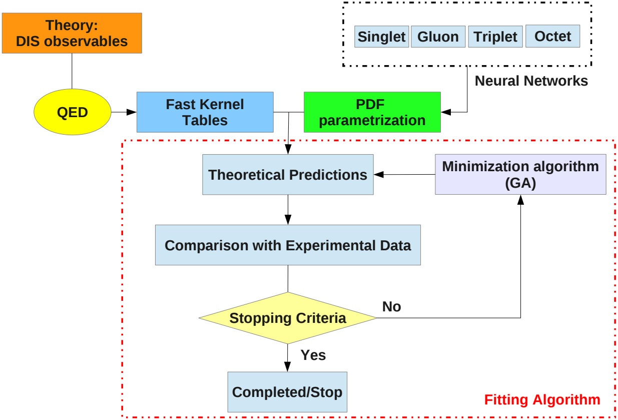
HERA-LHC 2009 PDF benchmarks



simple functional forms — Neural Networks

- Simple functional forms $\Delta q(x) = Ax^b(1-x)^c P(x)$
  $\longrightarrow$ systematic underestimation of uncertainties $\Rightarrow$ tolerance
- Artificial Neural Networks as universal interpolants
  $\longrightarrow$ reduce theoretical bias from choice of PDF functional form

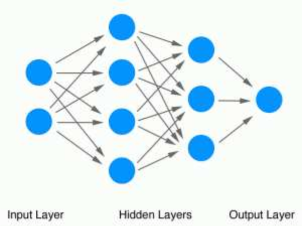
A general overview on the methodology



Ingredients:
Monte Carlo sampling and Neural Networks

Neural Networks

A convenient functional form
providing redundant and flexible parametrization
used as a generator of random functions in the PDF space

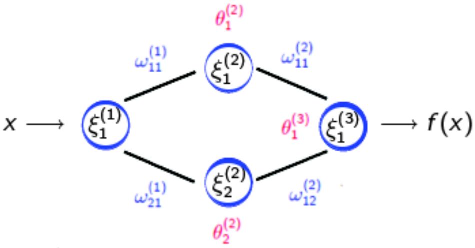


$$\xi_i^{(l)} = g\left(\sum_j^{n_l - 1} \omega_{ij}^{(l-1)} \xi_j^{(l-1)} - \theta_i^{(l)}\right)$$

$$g(x) = \frac{1}{1 + e^{-x}}$$

- made of neurons grouped into layers (define the architecture)
- each neuron receives input from neurons in preceding layer (feed-forward NN)
- activation determined by parameters (weights and thresholds)
- activation determined according to a non-linear function (except the last layer)

EXAMPLE: THE SIMPLEST 1-2-1 NN



$$f(x) \equiv \xi_1^{(3)} = \left\{ 1 + \exp\left[\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1 + e^{\theta_1^{(2)} - x\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1 + e^{\theta_2^{(2)} - x\omega_{21}^{(1)}}} \right] \right\}^{-1}$$

Recall: $\xi_i^{(l)} = g\left(\sum_j^{n_l - 1} \omega_{ij}^{(l-1)} \xi_j^{(l-1)} - \theta_i^{(l)} \right)$; $g(x) = \frac{1}{1 + e^{-x}}$

One more ingredient: minimization and stopping

GENETIC ALGORITHM

Standard minimization unefficient owing to the large parameter space
and non-local x-dependence of the observables
Genetic algorithm provides better exploration of the whole parameter space

- Set Neural Network parameters randomly
- Make clones of the parameter vector and mutate them
- Define a figure of merit or error function for the k-th replica

$$E^{(k)} = \frac{1}{N_{\mathrm{rep}}} \sum_{i,j=1}^{N_{\mathrm{rep}}} \left(g_{1,i}^{(\mathrm{art})(k)} - g_{1,i}^{(\mathrm{net})(k)} \right) \left((\mathrm{cov})^{-1} \right)_{ij} \left(g_{1,j}^{(\mathrm{art})(k)} - g_{1,j}^{(\mathrm{net})(k)} \right)$$

$g_{1,i}^{(\mathrm{art})(k)}$: generated from Monte Carlo sampling
$g_{1,i}^{(\mathrm{net})(k)}$: computed from Neural Network PDFs

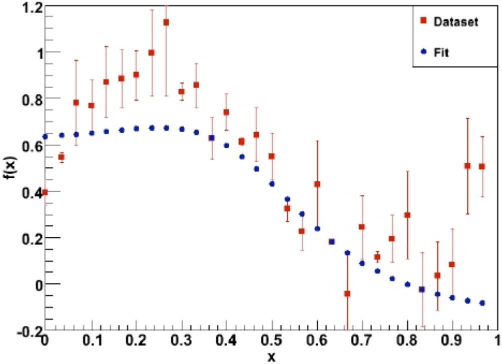
- Select the best set of parameters and perform other manipulations (crossing, mutating, ...) until stability is reached.

Minimization and stopping

DRAWBACK

- NN can learn fluctuations owing to their flexibility

UNDERLEARNING

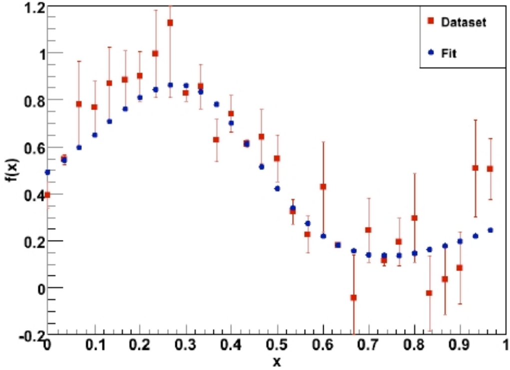


Minimization and stopping

DRAWBACK

- NN can learn fluctuations owing to their flexibility

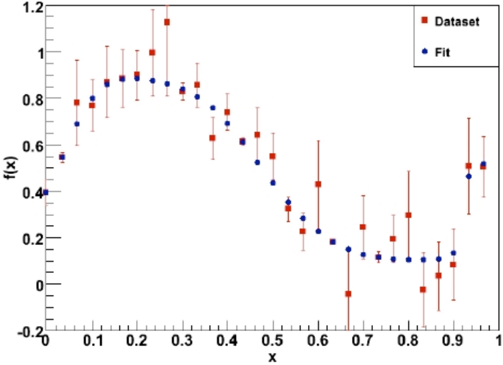
PROPER LEARNING



DRAWBACK

- NN can learn fluctuations owing to their flexibility

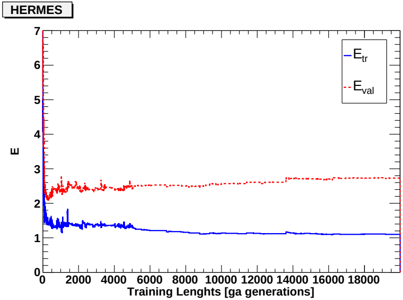
OVERLEARNING



Minimization and stopping

CROSS-VALIDATION METHOD

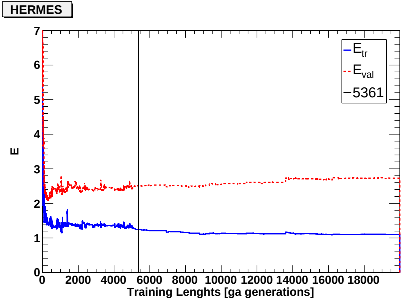
- divide data into two subsets (training & validation)
- train the NN on training subset and compute χ^2 for each subset
- stop when χ^2 of validation subset no longer decreases (NN are learning fluctuations!)



Minimization and stopping

CROSS-VALIDATION METHOD

- divide data into two subsets (training & validation)
- train the NN on training subset and compute χ^2 for each subset
- stop when χ^2 of validation subset no longer decreases (NN are learning fluctuations!)



The best fit does not coincide with the χ^2 absolute minimum

NNPDF timeline

	2008		2009	2010	2012		
	NNPDF1.0	NNPDF1.2	NNPDF2.0	NNPDF2.1	NNPDF2.3	MSTW08	CT10
DIS	✔	✔	✔	✔	✔	✔	✔
Drell-Yan	✘	✘	✔	✔	✔	✔	✔
Jet	✘	✘	✔	✔	✔	✔	✔
LHC	✘	✘	✘	✘	✔	✘	✘
strange	✘	✔	✔	✔	✔	✔	✔
Heavy Quark mass	✘	✘	✘	✔	✔	✔	✔
NNLO	✘	✘	✘	✔	✔	✔	✔

Input unpolarized PDF basis

Seven polarized PDFs (gluon + singlet and non-singlet quark combinations)

1. singlet $\Sigma(x) \equiv \Sigma_{i=1}^{n_f} q_i(x)$
2. gluon $g(x)$
3. total valence $V(x) \equiv u(x) - \bar{u}(x) + d(x) - \bar{d}(x)$
4. triplet $T_3(x) \equiv u(x) - d(x)$
5. sea asymmetry $\Delta_S(x) \equiv \bar{d}(x) - \bar{u}(x)$
6. total strangeness $s^+(x) = s(x) + \bar{s}(x)$
7. valence strangeness $s^-(x) = s(x) - \bar{s}(x)$

- At initial scale $Q_0^2 = 2 GeV^2$

Input polarized PDF basis

Four polarized PDFs (gluon + linear combinations of light quarks)

1. singlet $\Delta\Sigma(x) \equiv \Sigma_{i=1}^{n_f} \Delta q_i(x)$
2. gluon $\Delta g(x)$
3. triplet $\Delta T_3(x) \equiv \Delta u(x) - \Delta d(x)$
4. octet $\Delta T_8(x) \equiv \Delta u(x) + \Delta d(x) - 2\Delta s(x)$

$$\Delta q_i(x, Q^2) = q_i^{\uparrow\uparrow}(x, Q^2) + \bar{q}_i^{\uparrow\uparrow}(x, Q^2) - q_i^{\uparrow\downarrow}(x, Q^2) + \bar{q}_i^{\uparrow\downarrow}(x, Q^2)$$

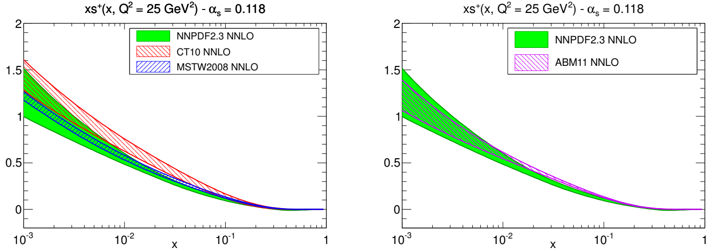
$$\Delta g(x, Q^2) \equiv g^{\uparrow\uparrow}(x, Q^2) - g^{\uparrow\downarrow}(x, Q^2)$$

Inclusive neutral-current DIS data do not allow disentangling the contributions from q and $\overline{q}$. In our notation, Δq takes into account flavor plus anti-flavor contributions.

- At initial scale $Q_0^2 = 1 GeV^2$
- Assume all heavy quarks are generated radiatively
- Adopt $\alpha_s(M_Z^2) = 0.119$, $m_c = 1.4$ GeV, $m_b = 4.75$ GeV

Comparison with other parton sets [arXiv:1211.5142]

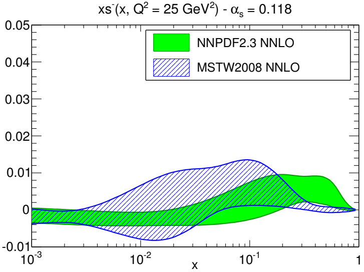
$s^+ = s + \bar{s}$



- HERAPDF1.5 does not provide independent strangeness, hence it is not shown (HERA data alone do not allow for disentangling of the strange contribution)
- The CT10 strange distribution is somewhat higher than that of other groups
  (different non-perturbative parametrization of the PDFs)
  (differences in the heavy quark treatment of neutrino dimuon data)
- Recall that NNPDF2.3 also includes LHC data
  (however, they were shown to have slight impact on strangeness)

Comparison with other parton sets [arXiv:1211.5142]

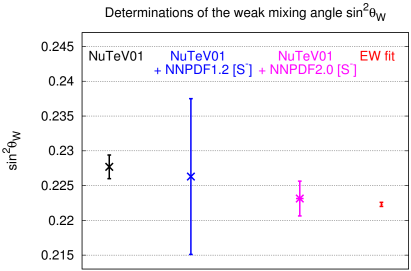
$$s^- = s - \bar{s}$$



- Only MSTW08 and NNPDF2.3 provide independent parametrization of s^- (though the MSTW08 involves only a few parameters)
- Extractions from both parton sets are in reasonable agreement
- Other parton sets are delivered assuming symmetric strangeness $s = \bar{s}$
- Recall that NNPDF2.3 also includes LHC data (however, they were shown to have slight impact on strangeness)

Impact on the NuTev anomaly

- The accurate determination of s^- has important phenomenological implications
- The NuTev anomaly: discrepancy (at 3σ level) between indirect (global fit) and direct (NuTev neutrino scattering) determinations of $sin^2\theta_W$
- The magnitude of the strangeness asymmetry s^- is large enough to remove the NuTev anomaly



- EW fit
  $sin^2\theta_W = 0.2223 \pm 0.0003$
- NuTeV (assumes $[S^-] = 0$)
  $sin^2\theta_W = 0.2277 \pm 0.0017$
- NuTeV + NNPDF1.2 $[S^-]$
  $sin^2\theta_W =$
  $0.2263 \pm 0.0017^{\rm exp} \pm 0.0107^{\rm PDFs}$
- NuTeV + NNPDF2.0 $[S^-]$
  $sin^2\theta_W =$
  $0.22314 \pm 0.0017^{\rm exp} \pm 0.00251^{\rm PDFs}$

$W^{\pm}$ asymmetry and the s^{-} distribution [arXiv:1203.1290]

- The $W^{\pm}$ asymmetry is driven mainly by the difference between $u\bar{d}$ and $\bar{u}d$ fusion, but also has some sensitivity to the strange quark asymmetry $s^{-} = s - \bar{s}$
- The strange s and $\bar{s}$ quarks make significant contributions to W^{+} and W^{-} production at hadron colliders, particularly the LHC, through $c\bar{s}$ and $\bar{c}s$ partonic fusion

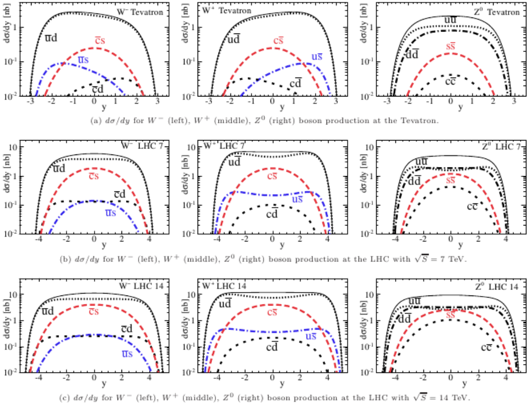
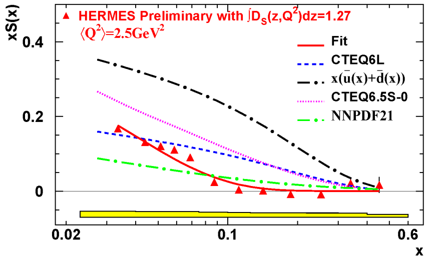
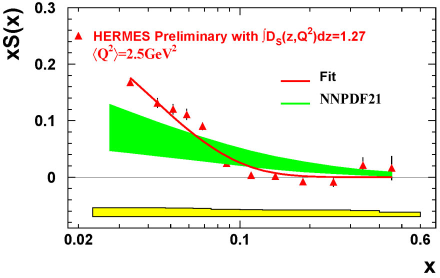


Figure 7: Partonic contributions to the differential cross section of on-shell $W^{\pm}/Z$ boson production at LO as a function of the vector boson rapidity. Partonic contributions containing a strange or anti-strange quark are denoted by (red) dashed and (blue) dot-dashed lines. The solid lines show the total contribution.

Strangeness from charged-kaon production in DIS

- Consider inclusive and semi-inclusive-charged-kaon spin asymmetries for a longitudinally polarized deuteron target
- Extract the LO parton distributions of the strange sea in the proton
- Find that the momentum densities are softer than previously assumed
- Helicity densities are consistent with zero
- The partial moment of the octet axial combination is observed to be substantially less than the axial charge extracted from hyperon decays under the assumption of $SU(3)$ symmetry





Polarized PDF Parametrization

$$\Delta\Sigma(x, Q_0^2) = (1-x)^{m_{\Delta\Sigma}} x^{-n_{\Delta\Sigma}} NN_{\Delta\Sigma}(x)$$

$$\Delta g(x, Q_0^2) = (1-x)^{m_{\Delta g}} x^{-n_{\Delta g}} NN_{\Delta g}(x)$$

$$\Delta T_3(x, Q_0^2) = A_{\Delta T_3}(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)$$

$$\Delta T_8(X, Q_0^2) = A_{\Delta T_8}(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)$$

1. Each polarized PDF parametrized with a multi-layer feed-forward NN (2-5-3-1)
2. Parametrization supplemented with a preprocessing polynomial:
   → exponents m and n randomly chosen in fixed intervals;
   → intervals must be sufficient large not to introduce a bias on the fit
   → check *a posteriori* by studying asymptotic exponents
3. Overall normalization constant factored out for triplet and octet.

$$A_{\Delta T_3} = \frac{a_3}{\int_0^1 dx[(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)]}$$

$$A_{\Delta T_8} = \frac{a_8}{\int_0^1 dx[(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)]}$$

Polarized PDF Parametrization

$$\Delta\Sigma(x, Q_0^2) = (1-x)^{m_{\Delta\Sigma}} x^{-n_{\Delta\Sigma}} NN_{\Delta\Sigma}(x)$$

$$\Delta g(x, Q_0^2) = (1-x)^{m_{\Delta g}} x^{-n_{\Delta g}} NN_{\Delta g}(x)$$

$$\Delta T_3(x, Q_0^2) = A_{\Delta T_3}(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)$$

$$\Delta T_8(X, Q_0^2) = A_{\Delta T_8}(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)$$

1. Each polarized PDF parametrized with a multi-layer feed-forward NN (2-5-3-1)
2. Parametrization supplemented with a preprocessing polynomial:
   → exponents m and n randomly chosen in fixed intervals;
   → intervals must be sufficient large not to introduce a bias on the fit
   → check *a posteriori* by studying asymptotic exponents
3. Overall normalization constant factored out for triplet and octet.

$$A_{\Delta T_3} = \frac{a_3}{\int_0^1 dx[(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)]}$$

$$A_{\Delta T_8} = \frac{a_8}{\int_0^1 dx[(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)]}$$

Polarized PDF Parametrization

$$\Delta\Sigma(x, Q_0^2) = (1-x)^{m_{\Delta\Sigma}} x^{-n_{\Delta\Sigma}} NN_{\Delta\Sigma}(x)$$

$$\Delta g(x, Q_0^2) = (1-x)^{m_{\Delta g}} x^{-n_{\Delta g}} NN_{\Delta g}(x)$$

$$\Delta T_3(x, Q_0^2) = A_{\Delta T_3}(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)$$

$$\Delta T_8(X, Q_0^2) = A_{\Delta T_8}(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)$$

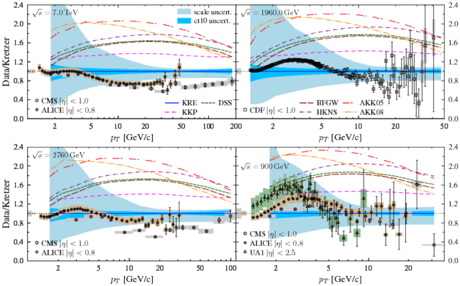
1. Each polarized PDF parametrized with a multi-layer feed-forward NN (2-5-3-1)
2. Parametrization supplemented with a preprocessing polynomial:
   → exponents m and n randomly chosen in fixed intervals;
   → intervals must be sufficient large not to introduce a bias on the fit
   → check *a posteriori* by studying asymptotic exponents
3. Overall normalization constant factored out for triplet and octet.

$$A_{\Delta T_3} = \frac{a_3}{\int_0^1 dx[(1-x)^{m_{\Delta T_3}} x^{-n_{\Delta T_3}} NN_{\Delta T_3}(x)]}$$

$$A_{\Delta T_8} = \frac{a_8}{\int_0^1 dx[(1-x)^{m_{\Delta T_8}} x^{-n_{\Delta T_8}} NN_{\Delta T_8}(x)]}$$

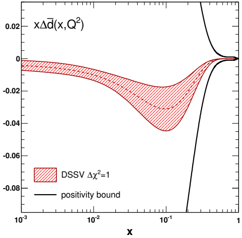
How well do we know fragmentation functions? [arXiv:1311.1415]

- Examine LHC, Tevatron, Sp$\bar{p}$S and RHIC data for inclusive unidentified charged-hadron production at $\sqrt{s} = 0.2 - 7$ TeV against NLO pQCD calculations with differet sets of FFs
- Quantify the systematics associated with the scale and PDF uncertainties
- Find large spread among the predictions with different FFs essentially due to sizable mutual differences in the gluon-to-hadron FFs
- None of the existing FF sets can reproduce the experimental results optimally, overshooting the data by up to a factor two for p_T10 GeV (modest dependence on scale)
- The best agreement is reached by Kretzer FF (based on SIA data only)
- Need for a global refit including LHC data



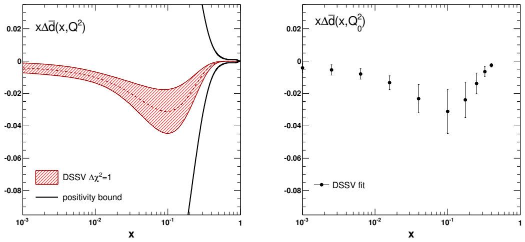
Construction of a suitable prior PDF ensemble

1. Take a polarized parton set wich provides separated $\Delta q - \Delta\bar{q}$ PDFs (from SIDIS)
   $\rightarrow$ we choose DSSV
2. Suppose they reproduce the *true* physical underlying behaviour
3. Sample their $\Delta\bar{u}$ and $\Delta\bar{d}$ distributions at a given reference scale
   $\rightarrow$ we take 10 points in the range $10^{-3} \lesssim x \lesssim 0.4$ at $Q_0^2 = 1$ GeV^2 (half logaritmically, half linearly spaced)
4. Perform a NN fit to these pseudodata
5. Supplement each replica in `NNPDFpol1.0` with $\Delta\bar{u}$ and $\Delta\bar{d}$ obtained in this way
6. Reweight PDFs and check that observables are properly reproduced
7. Check that reweighted results are stable upon the choice of different PDF priors
   $\rightarrow$ relax item 2, for example by increasing the nominal PDF uncertainty, until independence from the prior is reached
   $\rightarrow$ we have considered four different PDF priors: 1σ, 2σ, 3σ, 4σ



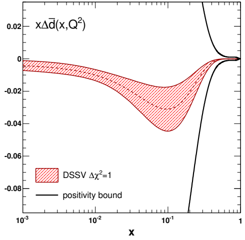
Construction of a suitable prior PDF ensemble

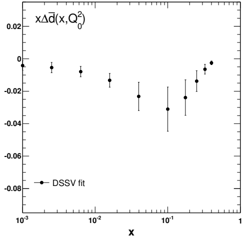
1. Take a polarized parton set wich provides separated $\Delta q - \Delta\bar{q}$ PDFs (from SIDIS)
   $\rightarrow$ we choose DSSV
2. Suppose they reproduce the *true* physical underlying behaviour
3. Sample their $\Delta\bar{u}$ and $\Delta\bar{d}$ distributions at a given reference scale
   $\rightarrow$ we take 10 points in the range $10^{-3} \lesssim x \lesssim 0.4$ at $Q_0^2 = 1$ GeV^2 (half logaritmically, half linearly spaced)
4. Perform a NN fit to these pseudodata
5. Supplement each replica in `NNPDFpol1.0` with $\Delta\bar{u}$ and $\Delta\bar{d}$ obtained in this way
6. Reweight PDFs and check that observables are properly reproduced
7. Check that reweighted results are stable upon the choice of different PDF priors
   $\rightarrow$ relax item 2, for example by increasing the nominal PDF uncertainty, until independence from the prior is reached
   $\rightarrow$ we have considered four different PDF priors: 1σ, 2σ, 3σ, 4σ

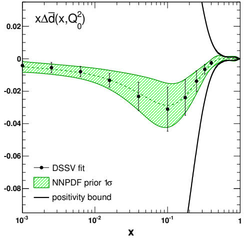


Construction of a suitable prior PDF ensemble

1. Take a polarized parton set wich provides separated $\Delta q - \Delta\bar{q}$ PDFs (from SIDIS)
   $\rightarrow$ we choose DSSV
2. Suppose they reproduce the *true* physical underlying behaviour
3. Sample their $\Delta\bar{u}$ and $\Delta\bar{d}$ distributions at a given reference scale
   $\rightarrow$ we take 10 points in the range $10^{-3} \lesssim x \lesssim 0.4$ at $Q_0^2 = 1$ GeV^2 (half logaritmically, half linearly spaced)
4. Perform a NN fit to these pseudodata
5. Supplement each replica in `NNPDFpol1.0` with $\Delta\bar{u}$ and $\Delta\bar{d}$ obtained in this way
6. Reweight PDFs and check that observables are properly reproduced
7. Check that reweighted results are stable upon the choice of different PDF priors
   $\rightarrow$ relax item 2, for example by increasing the nominal PDF uncertainty, until independence from the prior is reached
   $\rightarrow$ we have considered four different PDF priors: 1σ, 2σ, 3σ, 4σ







Reweighting with $W^{\pm}$ production at RHIC

New dataset on longitudinal single-spin asymmetry for $W^{\pm}$ boson production

$$A_L^{W^+} \sim \frac{-\Delta u(x_1)\bar{d}(x_2) + \Delta\bar{d}(x_1)u(x_2)}{u(x_1)\bar{d}(x_2) + \bar{d}(x_1)u(x_2)} \qquad A_L^{W^-} \sim \frac{-\Delta d(x_1)\bar{u}(x_2) + \Delta\bar{u}(x_1)d(x_2)}{d(x_1)\bar{u}(x_2) + \bar{u}(x_1)d(x_2)}$$

FEATURES

- sensitive to individual quark and antiquark flavours $(\Delta u, \Delta\bar{u}, \Delta d, \Delta\bar{d})$
  (purely weak process coupling q_L with $\bar{q}_R$ at partonic level, $u_L\bar{d}_R \to W^+$ or $d_L\bar{u}_R \to W^-$)
- no need of fragmentation functions (instead of SIDIS)

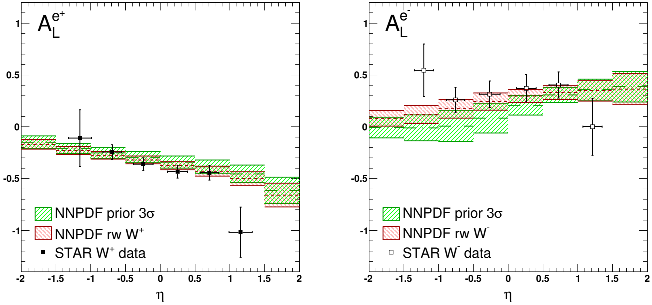
EXPERIMENTAL MEASUREMENT

- STAR and PHENIX at RHIC [arXiv:1009.0326] [arXiv:1009.0505]
  (only preliminary measurements from STAR (2012) [arXiv:1302.6639] will be considered here)

THEORETICAL PREDICTION

- `NNPDFpol1.0` itself is not good since q and $\bar{q}$ contributions are not disentangled
  $\to$ if new data bring in sufficient new information, results will be independent from prior
- the asymmetry is computed at NLO using CHE code [arXiv:1003.4533]

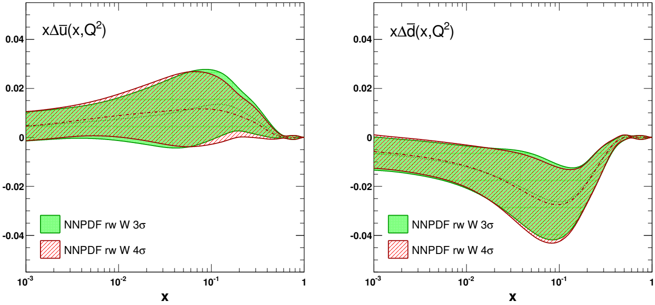
Sea quarks from W boson production at RHIC



Experiment	Set	$N_{\rm dat}$	$\chi^2/N_{\rm dat}$				$\chi^2_{\rm rw}/N_{\rm dat}$			
			1σ	2σ	3σ	4σ	1σ	2σ	3σ	4σ
STAR		12	2.02	2.08	**2.09**	1.94	1.48	1.19	**1.07**	1.07
	STAR-W^+	6	1.37	1.37	**1.54**	1.49	1.21	1.12	**1.10**	1.10
	STAR-W^-	6	2.67	2.79	**2.64**	2.39	1.75	1.25	**1.05**	1.04

Overall good description of both data sets after they are included in the fit via reweighting
Stability of reweighted results reached starting from 3σ or 4σ prior indifferently

Sea quarks from W boson production at RHIC



- $W^{\pm}$ data are effective in constraining light sea quark distributions $\Delta\bar{u}$ and $\Delta\bar{d}$
- Reweighting with separate W^+ and W^- datasets shows that $\Delta\bar{u}$ and $\Delta\bar{d}$ behaviour is driven by experimental information from W^- and W^+ respectively
- All other PDFs are almost unaffected (including strangeness)
- Independence of the reweighted results from the choice of the prior