Parton Distributions at the LHC: Challenges and Opportunities

Juan Rojo-Chacón

LPTHE, Universités Paris VI-VII

LPNHE, Exp-Theo Joint Meeting

Outline

- Parton distributions and global analyses
- The relevance of PDFs for LHC phenomenology
- Constraining PDFs from LHC measurements

PARTON DISTRIBUTIONS AND GLOBAL ANALYSES

QCD factorization

The DIS cross section can be decomposed using kinematics and Lorentz invariance in terms of structure functions $F_i(x, Q^2)$

$$\frac{d^2\sigma_{DIS}}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[\left(1+(1-y)^2\right)F_1(x,Q^2) + \frac{1-y}{x}\left(F_2(x,Q^2) - 2xF_1(x,Q^2)\right)\right]$$

Each structure function, using the QCD factorization theorem can be written as a convolution of a hard-scattering coefficient $C_i(x, \alpha_s(Q^2))$ and non-perturbative parton distributions $q_j(x, Q^2)$,

$$F_i(x,Q^2) = \int_x^1 \frac{dy}{y} C_{ij}(y, \alpha_s(Q^2)) q_j(x/y, Q^2)$$

where the PDFs satisfy the DGLAP evolution equations:

$$\frac{d q_i(x,Q^2)}{d \ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi}\int_x^1 \frac{dy}{y} P_{ij}\left(y, \alpha_s(Q^2)\right) q_j\left(\frac{x}{y}, Q^2\right)$$

Need to determine $q_i(x, Q_0^2)$ from experimental data.

QCD factorization

The DIS cross section can be decomposed using kinematics and Lorentz invariance in terms of structure functions $F_i(x, Q^2)$

$$\frac{d^2\sigma_{DIS}}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[\left(1+(1-y)^2\right)F_1(x,Q^2) + \frac{1-y}{x}\left(F_2(x,Q^2) - 2xF_1(x,Q^2)\right)\right]$$

Each structure function, using the QCD factorization theorem can be written as a convolution of a hard-scattering coefficient $C_i(x, \alpha_s(Q^2))$ and non-perturbative parton distributions $q_j(x, Q^2)$,

$$F_i(x,Q^2) = \int_x^1 \frac{dy}{y} C_{ij}(y, \alpha_s(Q^2)) q_j(x/y, Q^2)$$

where the PDFs satisfy the DGLAP evolution equations:

$$\frac{d\, q_i(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi}\int_x^1 \frac{dy}{y} P_{ij}\left(y, \alpha_s(Q^2)\right) q_j\left(\frac{x}{y}, Q^2\right)$$

Need to determine $q_i(x, Q_0^2)$ from experimental data.

QCD factorization

The DIS cross section can be decomposed using kinematics and Lorentz invariance in terms of structure functions $F_i(x, Q^2)$

$$\frac{d^2\sigma_{DIS}}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[\left(1+(1-y)^2\right)F_1(x,Q^2) + \frac{1-y}{x}\left(F_2(x,Q^2) - 2xF_1(x,Q^2)\right)\right]$$

Each structure function, using the QCD factorization theorem can be written as a convolution of a hard-scattering coefficient $C_i(x, \alpha_s(Q^2))$ and non-perturbative parton distributions $q_j(x, Q^2)$,

$$F_i(x,Q^2) = \int_x^1 \frac{dy}{y} C_{ij}(y, \alpha_s(Q^2)) q_j(x/y, Q^2)$$

where the PDFs satisfy the DGLAP evolution equations:

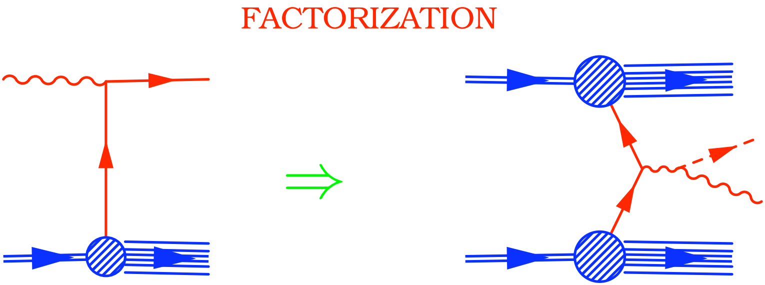
$$\frac{d\,q_i(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi}\int_x^1 \frac{dy}{y} P_{ij}\left(y, \alpha_s(Q^2)\right) q_j\left(\frac{x}{y}, Q^2\right)$$

Need to determine $q_i(x, Q_0^2)$ from experimental data.

QCD factorization

Same PDFs used to predict hadronic collisions:

$$\sigma_{\mathrm{LHC,i}} = \sum_j C_{ij}\left(x, \alpha_s(Q^2)\right) \otimes q_i(x, Q^2) \otimes q_j(x, Q^2)$$



The standard approach

First determine the **best-fit** pdf in an iterative way:

1. Parametrize PDFs at low scale Q_0^2 with a functional form

$$q_i\left(x, Q_0^2, \{A_i, b_i, \ldots\}\right) = A_i x^{b_i}(1-x)^{c_i}\left(1 + d_i x + e_i x^2 + \ldots\right)$$

Large x: Counting rules, small x: Regge theory (not from QCD!)

2. Evolve each PDF (DGLAP equations) to the scale Q^2 of experimental data + add hard scattering coefficients:

$$F_i^{(QCD)}(x, Q^2, \{A_i, b_i, \ldots\}1) = C_{ij}(x, \alpha(Q^2)) \otimes q_j\left(x, Q^2, \{A_i, b_i, \ldots\}\right)$$

3. Minimize a statistical estimator:

$$\chi^2\left(\{A_i, b_i, \ldots\}\right) = \frac{1}{N_{dat}} \sum_{i,j=1}^{N_{dat}} \left(F_i^{(exp)} - F_i^{(QCD)}\right)\left(\mathrm{cov}_{ij}^{-1}\right)\left(F_j^{(exp)} - F_j^{(QCD)}\right)$$

The standard approach

First determine the **best-fit** pdf in an iterative way:

1. Parametrize PDFs at low scale Q_0^2 with a functional form

$$q_i\left(x, Q_0^2, \{A_i, b_i, \ldots\}\right) = A_i x^{b_i} (1-x)^{c_i} \left(1 + d_i x + e_i x^2 + \ldots\right)$$

   Large x: Counting rules, small x: Regge theory (not from QCD!)

2. Evolve each PDF (DGLAP equations) to the scale Q^2 of experimental data + add hard scattering coefficients:

$$F_i^{(QCD)}(x, Q^2, \{A_i, b_i, \ldots\}1) = C_{ij}(x, \alpha(Q^2)) \otimes q_j\left(x, Q^2, \{A_i, b_i, \ldots\}\right)$$

3. Minimize a statistical estimator:

$$\chi^2\left(\{A_i, b_i, \ldots\}\right) = \frac{1}{N_{dat}} \sum_{i,j=1}^{N_{dat}} \left(F_i^{(exp)} - F_i^{(QCD)}\right)\left(\mathrm{cov}_{ij}^{-1}\right)\left(F_j^{(exp)} - F_j^{(QCD)}\right)$$

The standard approach

First determine the **best-fit** pdf in an iterative way:

1. Parametrize PDFs at low scale Q_0^2 with a functional form

$$q_i\left(x, Q_0^2, \{A_i, b_i, ...\}\right) = A_i x^{b_i}(1-x)^{c_i}\left(1 + d_i x + e_i x^2 + \ldots\right)$$

   Large x: Counting rules, small x: Regge theory (not from QCD!)

2. Evolve each PDF (DGLAP equations) to the scale Q^2 of experimental data + add hard scattering coefficients:

$$F_i^{(QCD)}(x, Q^2, \{A_i, b_i, ...\}1) = C_{ij}(x, \alpha(Q^2)) \otimes q_j\left(x, Q^2, \{A_i, b_i, ...\}\right)$$

3. Minimize a statistical estimator:

$$\chi^2\left(\{A_i, b_i, ...\}\right) = \frac{1}{N_{dat}}\sum_{i,j=1}^{N_{dat}}\left(F_i^{(exp)} - F_i^{(QCD)}\right)\left(\mathrm{cov}_{ij}^{-1}\right)\left(F_j^{(exp)} - F_j^{(QCD)}\right)$$

The standard approach

First determine the **best-fit** pdf in an iterative way:

1. Parametrize PDFs at low scale Q_0^2 with a functional form

$$q_i\left(x, Q_0^2, \{A_i, b_i, \ldots\}\right) = A_i x^{b_i} (1-x)^{c_i} \left(1 + d_i x + e_i x^2 + \ldots\right)$$

   Large x: Counting rules, small x: Regge theory (not from QCD!)

2. Evolve each PDF (DGLAP equations) to the scale Q^2 of experimental data + add hard scattering coefficients:

$$F_i^{(QCD)}(x, Q^2, \{A_i, b_i, \ldots\}1) = C_{ij}(x, \alpha(Q^2)) \otimes q_j \left(x, Q^2, \{A_i, b_i, \ldots\}\right)$$

3. Minimize a statistical estimator:

$$\chi^2\left(\{A_i, b_i, \ldots\}\right) = \frac{1}{N_{dat}} \sum_{i,j=1}^{N_{dat}} \left(F_i^{(exp)} - F_i^{(QCD)}\right)\left(\mathrm{cov}_{ij}^{-1}\right)\left(F_j^{(exp)} - F_j^{(QCD)}\right)$$

PDF fitting collaborations

1. Global fits (data sets: DIS + DY + Jets + W asymmetry + ...):
   - CTEQ
   - MRSW (formerly MRST)
2. Reduced data sets:
   - Alekhin
   - NNPDF Collaboration
   - ATLAS-PDFs, ZEUS-PDFs, H1-PDFs, ...

All modern PDF sets available in a common format through the LHAPDF library:

http://projects.hepforge.org/lhapdf/

Data sets in global fits

Data set	CTEQ6.5	MRSW07
HERA DIS NC reduced cross sections	Yes	Yes
HERA DIS CC reduced cross sections	Yes	Yes
HERA DIS F_2^c and F_2^b (heavy flavours)	Yes	Yes
Fixed target NC DIS: BCDMS and NMC	Yes	Yes
Fixed target CC DIS: CCFR	Yes	No
Fixed target CC DIS: NuTeV and CHORUS	No	Yes
Fixed target Drell-Yan: E605, E886	Yes	Yes
CDF W lepton asymmetry	Yes	Yes
CDF/D0 inclusive jet production	Yes	Yes

CTEQ6.5 = hep-ph/0611254

MRSW07 = arXiv:0706.0459 [hep-ph]

Errors in PDFs

Estimate the *"experimental"* uncertainty of the **best-fit** PDF set: Explore parameter space (varying pars. $a_i = (a_i)_{\rm best-fit} + \delta a_i$) of the PDFs allowing

$$\Delta\chi^2 = \chi^2\left(\{a_i\}\right) - \chi^2_{best\ \ fit}\left(\{(a_i)_{best-fit}\}\right)$$

Then the sets of PDFs satisfying $\Delta\chi^2$ condition estimate the PDF uncertainties.

Problems of this approach:

- Other sources of uncertainties not accounted for: Parametrization bias due to the use of fixed functional forms: *Ex.* why at small x $q_i(x) \sim x^{b_i}$?
- Restricted to parameter space of PDFs functional forms.
- Uncertainties are not faithfully estimated: the introduction of arbitrary tolerance criteria $\Delta\chi^2$ (50-100!!) makes the estimation of errors arbitrary.
- Lack of general error propagation: use linear error propagation.
- Problems of incompatibility between experiments.

Errors in PDFs

Estimate the *"experimental"* uncertainty of the **best-fit** PDF set: Explore parameter space (varying pars. $a_i = (a_i)_{\mathrm{best-fit}} + \delta a_i$) of the PDFs allowing

$$\Delta\chi^2 = \chi^2\left(\{a_i\}\right) - \chi^2_{best\ \ fit}\left(\{(a_i)_{best-fit}\}\right)$$

Then the sets of PDFs satisfying $\Delta\chi^2$ condition estimate the PDF uncertainties.
Problems of this approach:

- Other sources of uncertainties not accounted for: Parametrization bias due to the use of fixed functional forms: *Ex.* why at small x $q_i(x) \sim x^{b_i}$?
- Restricted to parameter space of PDFs functional forms.
- Uncertainties are not faithfully estimated: the introduction of arbitrary tolerance criteria $\Delta\chi^2$ (50-100!!) makes the estimation of errors arbitrary.
- Lack of general error propagation: use linear error propagation.
- Problems of incompatibility between experiments.

Errors in PDFs

Estimate the *"experimental"* uncertainty of the **best-fit** PDF set: Explore parameter space (varying pars. $a_i = (a_i)_{\mathrm{best-fit}} + \delta a_i$) of the PDFs allowing

$$\Delta\chi^2 = \chi^2\left(\{a_i\}\right) - \chi^2_{best\ \ fit}\left(\{(a_i)_{best-fit}\}\right)$$

Then the sets of PDFs satisfying $\Delta\chi^2$ condition estimate the PDF uncertainties.
Problems of this approach:

- Other sources of uncertainties not accounted for: Parametrization bias due to the use of fixed functional forms: *Ex.* why at small x $q_i(x) \sim x^{b_i}$?
- Restricted to parameter space of PDFs functional forms.
- Uncertainties are not faithfully estimated: the introduction of arbitrary tolerance criteria $\Delta\chi^2$ (50-100!!) makes the estimation of errors arbitrary.
- Lack of general error propagation: use linear error propagation.
- Problems of incompatibility between experiments.

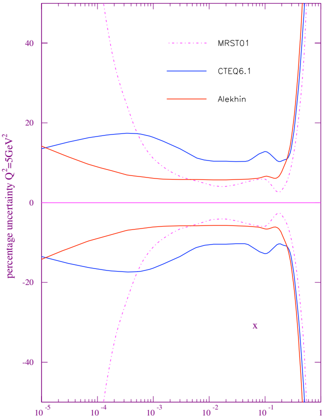
Theoretical error in input parametrization

R. Thorne talk at PDF4LHC

MRST uncertainty blows up for very small x, whereas Alekhin (and ZEUS and H1) gets slowly bigger, and CTEQ saturates (or even decreases).

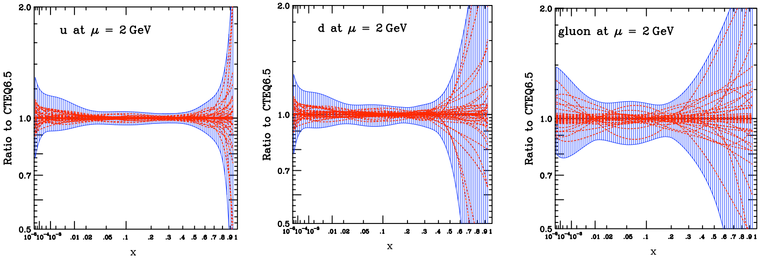
Related to input forms and scales.

(*Neck* in MRST gluon cured in MSTW).



PDF uncertainties

PDF eigenvectors and total uncertainty from CTEQ6.5



Uncertainties larger at small-x and large-x (lack of exp. data)

Gluon pdf rather unconstrained (in DIS only through scaling violations).

S. Forte, CERN 02-08

WHERE DO WE STAND NOW?

WHAT WE HAVE LEARNT

- LIGHT QUARK STRUCTURE IN "VALENCE" REGION $0.1 \lesssim x \lesssim 0.5$ (old fixed target dis data)
- SINGLET AND GLUON AT SMALL $x < 10^{-2}$ (HERA)
- SEA ASYMMETRY AT MEDIUM $x \sim 0.1 \div 0.2$ (Drell-Yan)
- HINTS ON STRANGENESS (neutrinos)

WHAT WE ARE STILL MISSING

- GLUONS AT LARGE x (cfr large E_T jet problem)
- NONSINGLET & VALENCE AT SMALL x
- DETAILED INFO ON STRANGENESS (cfr NuTeV problem)
- INFO ON HEAVY QUARKS (cfr small x W xsect problem)

News from global PDF analysis

- Inclusion of heavy quark masses in PDF fits (ACOT scheme) by CTEQ6.5 $\rightarrow$ Increases the W, Z cross-sections at LHC by $\sim 8\%$.
- Full NNLO global analysis will be the default in the LHC era (CTEQ7 expected before summer, MRSW already NNLO).
- Final set of combined HERA data to be presented soon $\rightarrow$ Backbone of PDF analysis for many years.
- Recent measurement of F_L at HERA will help in constraining low-x gluon.

News from global PDF analysis

- Inclusion of heavy quark masses in PDF fits (ACOT scheme) by CTEQ6.5 $\rightarrow$ Increases the W, Z cross-sections at LHC by $\sim 8\%$.
- Full NNLO global analysis will be the default in the LHC era (CTEQ7 expected before summer, MRSW already NNLO).
- Final set of combined HERA data to be presented soon $\rightarrow$ Backbone of PDF analysis for many years.
- Recent measurement of F_L at HERA will help in constraining low-x gluon.

News from global PDF analysis

- Inclusion of heavy quark masses in PDF fits (ACOT scheme) by CTEQ6.5 $\rightarrow$ Increases the W, Z cross-sections at LHC by $\sim 8\%$.
- Full NNLO global analysis will be the default in the LHC era (CTEQ7 expected before summer, MRSW already NNLO).
- Final set of combined HERA data to be presented soon $\rightarrow$ Backbone of PDF analysis for many years.
- Recent measurement of F_L at HERA will help in constraining low-x gluon.

News from global PDF analysis

- Inclusion of heavy quark masses in PDF fits (ACOT scheme) by CTEQ6.5 $\rightarrow$ Increases the W, Z cross-sections at LHC by $\sim 8\%$.
- Full NNLO global analysis will be the default in the LHC era (CTEQ7 expected before summer, MRSW already NNLO).
- Final set of combined HERA data to be presented soon $\rightarrow$ Backbone of PDF analysis for many years.
- Recent measurement of F_L at HERA will help in constraining low-x gluon.

HERA is still alive!

S. Glazov talk at PDF4LHC

Experimental data still to come

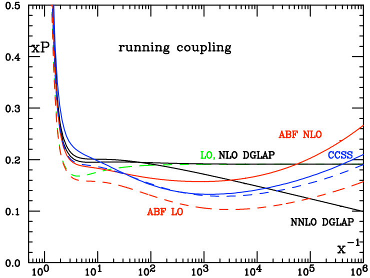
- Final analysis of F_2 structure function at low $Q^2 < 100$ GeV^2 and low x (H1).
- Analysis of σ_r at high Q^2 and high x using HERA-II data.
- Measurement of F_L structure function.
- HERA-II analysis of $F_2^{c\bar{c}}$ and $F_2^{b\bar{b}}$.
- Combination of all HERA data.
- PDF extraction based on the combined HERA data.
- Tevatron W asymmetry and Z rapidity with complete statistics.

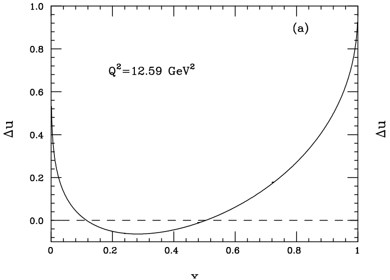
Tools for PDF evolution

- Parton evolution codes
  1. `HOPPET`: a x-space Higher Order Perturbative Parton Evolution Toolkit (2001-2008 G. Salam, (+2007 J.R. for doc.))
  2. `QCDNUM`: x-space evolution, used by HERA collaborations and in `LHAPDF` (M. Botje)
  3. `PEGASUS`: N-space evolution, only suitable for *mellinizable* PDFs (A. Vogt).
- Grid techniques to incorporate hadronic processes to PDF fits (much better than K-factors!)
  1. `FastNLO`: NLO jet cross sections (Rabbertz, Wobisch, ...).
  2. `NLOgrid, APPLgrid` (Carli, Clements, Salam, ...).
- LHAPDF library.

Improving the theory

The standard for global fits is NNLO DGLAP parton evolution with NNLO physical observables (only jets missing).
Extend QCD theory with small-x and large-x resummation (near future)





Left: ABF: Altarelli-Ball-Forte, CCSS: Ciafaloni-Colferai-Salam-Staso

Right: Corcella-Magnea.

THE NEURAL NETWORK APPROACH TO PARTON DISTRIBUTIONS

The NNPDF Collaboration:
Luigi Del Debbio, Stefano Forte, José I. Latorre,
Andrea Piccione and Juan Rojo,
(+2007) Richard D. Ball, Alberto Guffanti and Maria Ubiali.
http://sophia.ecm.ub.es/nnpdf

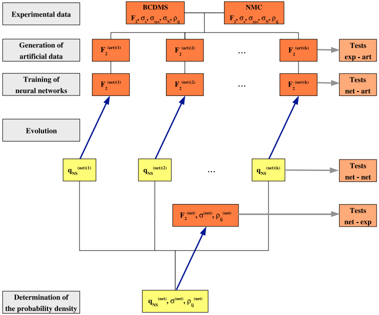
Non-Singlet fit: JHEP 0703:039,2007
Full DIS fit: around the corner ...

The NNPDF approach

Basic Idea: Monte Carlo sampling coupled to Neural Network interpolation

- Generate a set of Monte Carlo replicas $\sigma^{(k)}(p_i)$ of the original data set $\sigma^{(\mathrm{data})}(p_i)$, representation of $\mathcal{P}[\sigma(p_i)]$ at discrete set of points p_i
- Train a neural net for each pdf on each replica, obtaining a representation of the pdfs $q_i^{(\mathrm{net})(k)}$
- The set of neural nets is a representation of the probability density:

$$\langle \sigma\left[q_i\right] \rangle = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} \sigma\left[q_i^{(\mathrm{net})(k)}\right]$$

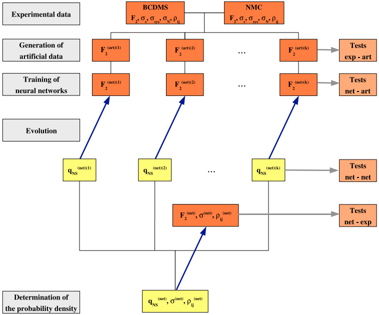


The NNPDF approach

Basic Idea: Monte Carlo sampling coupled to Neural Network interpolation

- Generate a set of Monte Carlo replicas $\sigma^{(k)}(p_i)$ of the original data set $\sigma^{(\mathrm{data})}(p_i)$, representation of $\mathcal{P}[\sigma(p_i)]$ at discrete set of points p_i
- Train a neural net for each pdf on each replica, obtaining a representation of the pdfs $q_i^{(\mathrm{net})(k)}$
- The set of neural nets is a representation of the probability density:

$$\langle \sigma\left[q_i\right] \rangle = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} \sigma\left[q_i^{(\mathrm{net})(k)}\right]$$

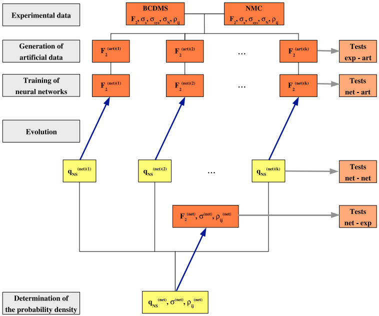


The NNPDF approach

Basic Idea: Monte Carlo sampling coupled to Neural Network interpolation

- Generate a set of Monte Carlo replicas $\sigma^{(k)}(p_i)$ of the original data set $\sigma^{(\mathrm{data})}(p_i)$, representation of $\mathcal{P}[\sigma(p_i)]$ at discrete set of points p_i
- Train a neural net for each pdf on each replica, obtaining a representation of the pdfs $q_i^{(\mathrm{net})(k)}$
- The set of neural nets is a representation of the probability density:

$$\langle \sigma\left[q_i\right] \rangle = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} \sigma\left[q_i^{(\mathrm{net})(k)} \right]$$

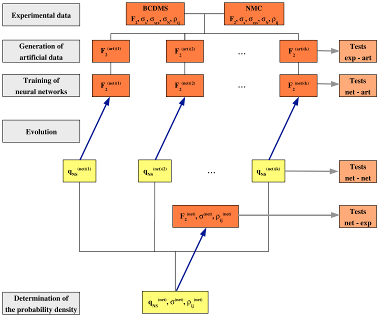


The NNPDF approach

Basic Idea: Monte Carlo sampling coupled to Neural Network interpolation

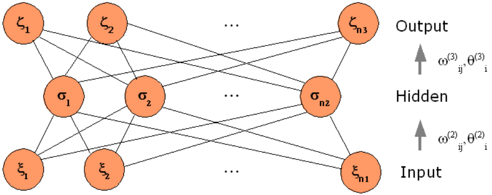
- Generate a set of Monte Carlo replicas $\sigma^{(k)}(p_i)$ of the original data set $\sigma^{(\mathrm{data})}(p_i)$, representation of $\mathcal{P}[\sigma(p_i)]$ at discrete set of points p_i
- Train a neural net for each pdf on each replica, obtaining a representation of the pdfs $q_i^{(\mathrm{net})(k)}$
- The set of neural nets is a representation of the probability density:

$$\langle \sigma\left[q_i\right] \rangle = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} \sigma\left[{q_i}^{(\mathrm{net})(k)} \right]$$



Neural networks

Use neural networks (robust, universal and unbiased approximants) to parametrize $q_i(x, Q_0^2)$ instead of simple functional forms.

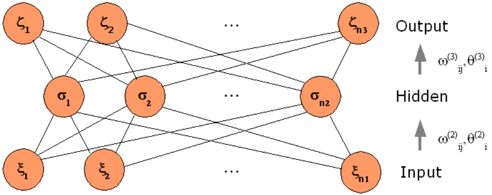


Neural networks common in HEP applications (event recognition, event classification ...)

$$\xi_i^{(l+1)} = g\left(\sum_{j=1}^{n(l)} \omega_{ij}^{(l)} \xi_j^l\right), \quad g(x) = \frac{1}{1+e^x}, \quad l = 2, \ldots, L$$

Neural networks

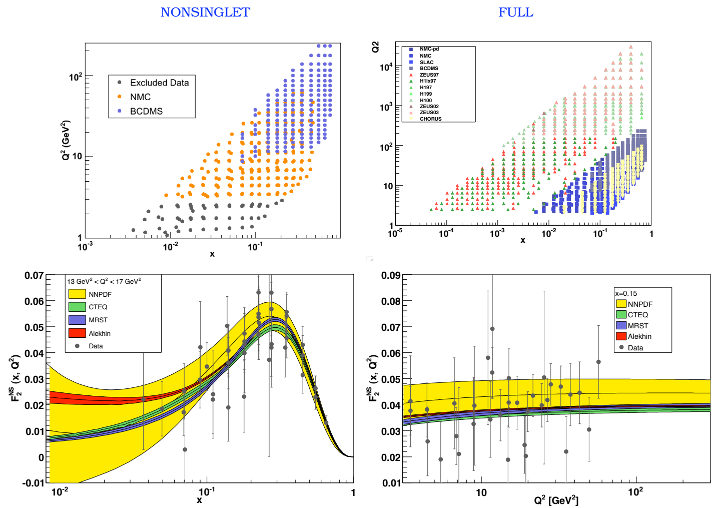
Use neural networks (robust, universal and unbiased approximants) to parametrize $q_i(x, Q_0^2)$ instead of simple functional forms.



Neural networks common in HEP applications (event recognition, event classification ...)

$$\xi_i^{(l+1)} = g\left(\sum_{j=1}^{n(l)} \omega_{ij}^{(l)} \xi_j^l\right), \quad g(x) = \frac{1}{1+e^x}, \quad l = 2, \ldots, L$$

The NNPDF approach - Results

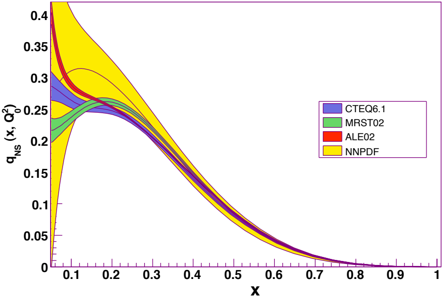


The NNPDF approach - the NS case

B. Webber intro talk at PDF4LHC

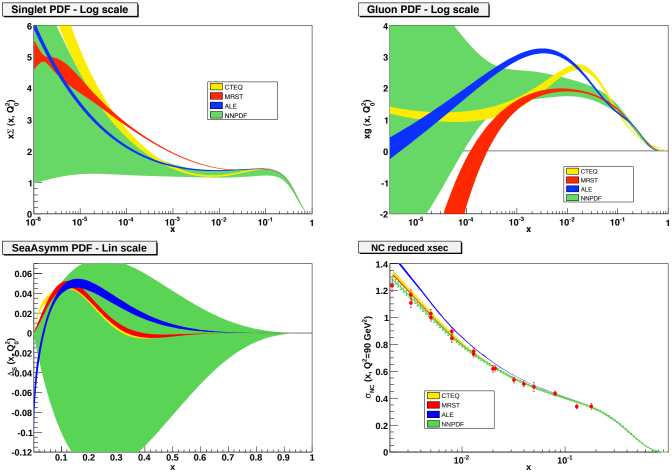
Parametrization Uncertainties

- Neural network approach doesn't constrain form of PDFs so much as fixed (Regge?) parametrizations



L Del Debbio et al., JHEP0703, 039

The NNPDF approach - Preliminary results



The NNPDF approach - Summary

R. Thorne (MRSW collaboration) at PHYSTAT for LHC 07

Statistical approach used by **Neural Network** group (Del Debbio *et al.*) construct a set of Monte Carlo replicas $\sigma^k(p_i)$ of the original data set $\sigma^{data}(p_i)$. Representation of $P[\sigma(p_i)]$ at points p_i.

Train a neural network for the parton distribution function on each replica, obtaining a representation of the pdfs $q_i^{(net)(k)}$.

The set of neural nets is a representation of the probability density – mean μ_O and deviation σ_O of observable O then given by

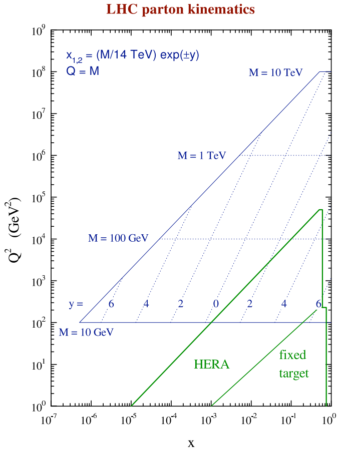
$$\mu_O = \frac{1}{N_{rep}} \sum_1^{N_{rep}} O[q_i^{(net)(k)}], \quad \sigma_O^2 = \frac{1}{N_{rep}} \sum_1^{N_{rep}} (O[q_i^{(net)(k)}] - \mu_O)^2.$$

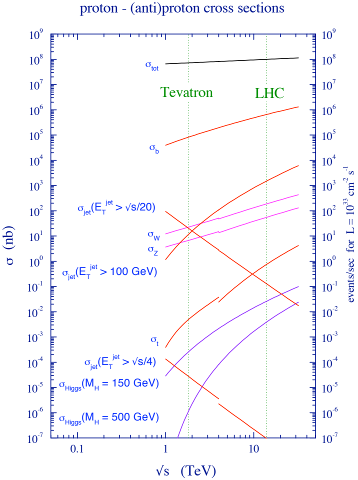
Can incorporate full information about measurements and their error correlations in the distribution of $\sigma^{data}(p_i)$.

This is statistically correct, and does not rely on the approximation of linear propagation of errors in calculating observables, but is more complicated and time intensive.

PDFs AND LHC PHENOMENOLOGY

From HERA to LHC





PDF impact for discovery at LHC

Measurements not much affected by PDF uncertainties

1. SM Higgs production
2. Z' production (or any other process whose signal is heavy mass di-lepton pairs)

Other more affected:

1. BSM signals (like contact interactions) in the high - E_T jet cross section $\rightarrow$ Errors in large-x gluon
2. Size of compact extra dimensions decreases from 6 TeV to 2 TeV due to PDF uncertainties.

Full phenomenological study of the impact of PDFs at the LHC still missing.

PDF impact for discovery at LHC

Measurements not much affected by PDF uncertainties

1. SM Higgs production
2. Z' production (or any other process whose signal is heavy mass di-lepton pairs)

Other more affected:

1. BSM signals (like contact interactions) in the high - E_T jet cross section $\rightarrow$ Errors in large-x gluon
2. Size of compact extra dimensions decreases from 6 TeV to 2 TeV due to PDF uncertainties.

Full phenomenological study of the impact of PDFs at the LHC still missing.

PDF impact for discovery at LHC

Measurements not much affected by PDF uncertainties

1. SM Higgs production
2. Z' production (or any other process whose signal is heavy mass di-lepton pairs)

Other more affected:

1. BSM signals (like contact interactions) in the high - E_T jet cross section $\rightarrow$ Errors in large-x gluon
2. Size of compact extra dimensions decreases from 6 TeV to 2 TeV due to PDF uncertainties.

Full phenomenological study of the impact of PDFs at the LHC still missing.

PDF impact for measurements at LHC

Electroweak boson production →Standard candles at the LHC?
As of HERA-LHC workshop 2005 (hep-ph-0511119):

PDF Set	$\sigma(W^+).B(W^+ \to l^+\nu_l)$	$\sigma(W^-).B(W^- \to l^-\bar{\nu}_l)$	$\sigma(Z).B(Z \to l^+l^-)$
ZEUS-S no HERA	10.63 ± 1.73 nb	7.80 ± 1.18 nb	1.69 ± 0.23 nb
ZEUS-S	12.07 ± 0.41 nb	8.76 ± 0.30 nb	1.89 ± 0.06 nb
CTEQ6.1	11.66 ± 0.56 nb	8.58 ± 0.43 nb	1.92 ± 0.08 nb
MRST01	11.72 ± 0.23 nb	8.72 ± 0.16 nb	1.96 ± 0.03 nb

So pdfs errors were thought to be $\Delta\sigma_W/\sigma_W \leq 5\%$
But central prediction by CTEQ6.5 has increased by $\Delta\sigma_W/\sigma_W \sim 8\%$ (larger small-x quarks).

PDF impact for measurements at LHC

Electroweak boson production →Standard candles at the LHC?
As of HERA-LHC workshop 2005 (hep-ph-0511119):

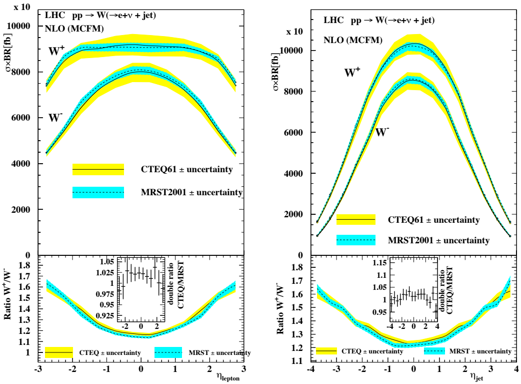
PDF Set	$\sigma(W^+).B(W^+ \to l^+\nu_l)$	$\sigma(W^-).B(W^- \to l^-\bar{\nu}_l)$	$\sigma(Z).B(Z \to l^+l^-)$
ZEUS-S no HERA	10.63 ± 1.73 nb	7.80 ± 1.18 nb	1.69 ± 0.23 nb
ZEUS-S	12.07 ± 0.41 nb	8.76 ± 0.30 nb	1.89 ± 0.06 nb
CTEQ6.1	11.66 ± 0.56 nb	8.58 ± 0.43 nb	1.92 ± 0.08 nb
MRST01	11.72 ± 0.23 nb	8.72 ± 0.16 nb	1.96 ± 0.03 nb

So pdfs errors were thought to be $\Delta\sigma_W/\sigma_W \leq 5\%$

But central prediction by CTEQ6.5 has increased by $\Delta\sigma_W/\sigma_W \sim 8\%$ (larger small-x quarks).

W + jet production

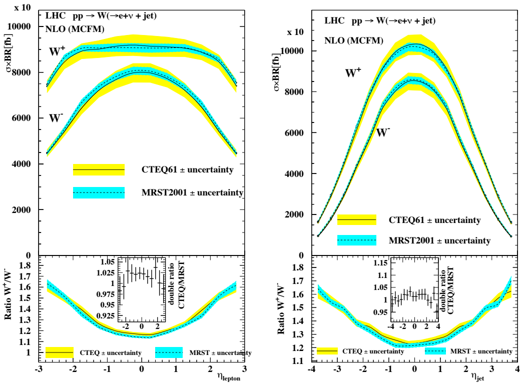
W boson production together with jets (hep-ph-0511119):



Typical size of PDF uncertainties $\sim 10\%$:

W + jet production

W boson production together with jets (hep-ph-0511119):

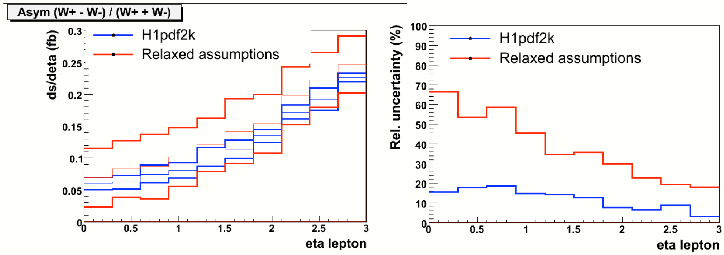


Typical size of PDF uncertainties ∼ 10%:

S. Forte, CERN 02-08

DETERMINING QUARKS AT SMALL x

- W PRODUCTION AT LHC PROBES $x \sim 10^{-2}$
- $W^{\pm}$ ASYMMETRIES SENSITIVE TO $\bar{u}/\bar{d}$
- $\Rightarrow$ IF SMALL x BEHAVIOUR IS NOT AS CURRENTLY ASSUMED ("REGGE"), $W^{\pm}$ ASYMMETRY CHANGES BY UP TO FACTOR 5!



E. Perez (CMS 2006)

Crucial information in first hours of LHC running.

Extra dimensions

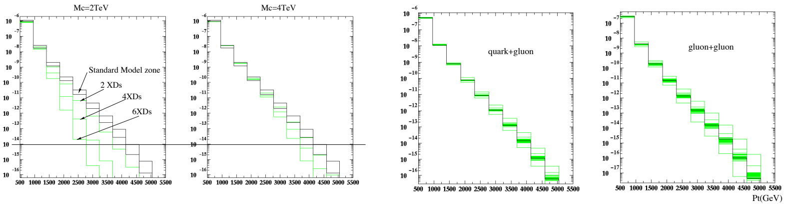
EXAMPLE: LACK OF KNOWLEDGE OF LARGE x GLUON LIMITS DISCOVERY POTENTIAL FOR EXTRA DIMENSIONS

UPPER LIMIT ON COMPACTIFICATION SCALE FROM DIJET CROSS SECTIONS FROM 100 FB^{-1} AT LHC Ferrag (ATLAS, 2006)

	2 extra dimensions	4 extra dimensions	6 extra dimensions
THEORETICALLY	5 TeV	5 TeV	5 TeV
INCLUDING PDF UNCERTAINTIES	$<$ 2 TeV	$<$ 3 TeV	$<$ 4 TeV

CROSS–SECTION IN FIXED p_t BINS

EXTRA DIMENSIONS VS STANDARD MODEL | PDF UNC.: QG CHANNEL | GG CHANNEL



CONSTRAINING PDFS AT THE LHC

Constraining PDFs at LHC

Potential measurements

1. Inclusive jet cross-section $\rightarrow$ Requires $\leq 1\%$ error in jet energy scale (very challenging!).
2. Dijet production.
3. Vector boson production and asymmetries, Drell-Yan pair production $\rightarrow$ Excellent statistics
4. $pp \rightarrow Z\,(\rightarrow l^+ l^-) +$ jets: the dominant sub-process is $qg \rightarrow Z + q \rightarrow$ Clean process, sensitive to the gluon PDF.
5. Prompt photon production. Either $pp \rightarrow \gamma X$ or $pp \rightarrow \gamma$+jet. Sensitive to gluon ($qg \rightarrow q + \gamma$) (But problems with isolation).
6. Heavy flavour production $\rightarrow$ Ask M. Cacciari.

Lots of work still to do!

Constraining PDFs at LHC

Potential measurements

1. Inclusive jet cross-section → Requires $\leq 1\%$ error in jet energy scale (very challenging!).
2. Dijet production.
3. Vector boson production and asymmetries, Drell-Yan pair production → Excellent statistics
4. $pp \to Z\left(\to l^+l^-\right) +$ jets: the dominant sub-process is $qg \to Z + q \to$ Clean process, sensitive to the gluon PDF.
5. Prompt photon production. Either $pp \to \gamma X$ or $pp \to \gamma+$jet. Sensitive to gluon ($qg \to q + \gamma$) (But problems with isolation).
6. Heavy flavour production → Ask M. Cacciari.

Lots of work still to do!

Constraining PDFs at LHC

Potential measurements

1. Inclusive jet cross-section → Requires $\leq 1\%$ error in jet energy scale (very challenging!).
2. Dijet production.
3. Vector boson production and asymmetries, Drell-Yan pair production → Excellent statistics
4. $pp \to Z\left(\to l^+l^-\right) +$ jets: the dominant sub-process is $qg \to Z + q \to$ Clean process, sensitive to the gluon PDF.
5. Prompt photon production. Either $pp \to \gamma X$ or $pp \to \gamma+$jet. Sensitive to gluon ($qg \to q + \gamma$) (But problems with isolation).
6. Heavy flavour production → Ask M. Cacciari.

Lots of work still to do!

Constraining PDFs at LHC

Potential measurements

1. Inclusive jet cross-section → Requires $\leq 1\%$ error in jet energy scale (very challenging!).
2. Dijet production.
3. Vector boson production and asymmetries, Drell-Yan pair production → Excellent statistics
4. $pp \to Z\left(\to l^+ l^-\right) + \text{jets}$: the dominant sub-process is $qg \to Z + q \to$ Clean process, sensitive to the gluon PDF.
5. Prompt photon production. Either $pp \to \gamma X$ or $pp \to \gamma$+jet. Sensitive to gluon ($qg \to q + \gamma$) (But problems with isolation).
6. Heavy flavour production → Ask M. Cacciari.

Lots of work still to do!

Constraining PDFs at LHC

Potential measurements

1. Inclusive jet cross-section $\rightarrow$ Requires $\leq 1\%$ error in jet energy scale (very challenging!).
2. Dijet production.
3. Vector boson production and asymmetries, Drell-Yan pair production $\rightarrow$ Excellent statistics
4. $pp \rightarrow Z \left(\rightarrow l^+ l^-\right) +$ jets: the dominant sub-process is $qg \rightarrow Z + q \rightarrow$ Clean process, sensitive to the gluon PDF.
5. Prompt photon production. Either $pp \rightarrow \gamma X$ or $pp \rightarrow \gamma+$jet. Sensitive to gluon ($qg \rightarrow q + \gamma$) (But problems with isolation).
6. Heavy flavour production $\rightarrow$ Ask M. Cacciari.

Lots of work still to do!

Constraining PDFs at LHC

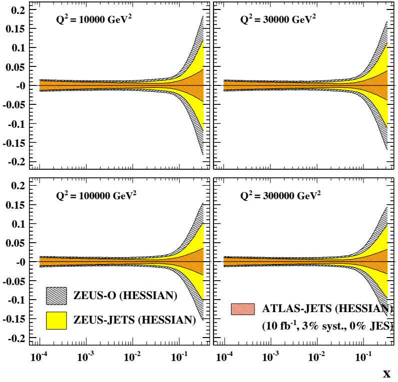
Potential measurements

1. Inclusive jet cross-section → Requires $\leq 1\%$ error in jet energy scale (very challenging!).
2. Dijet production.
3. Vector boson production and asymmetries, Drell-Yan pair production → Excellent statistics
4. $pp \to Z \left(\to l^+ l^-\right) +$ jets: the dominant sub-process is $qg \to Z + q \to$ Clean process, sensitive to the gluon PDF.
5. Prompt photon production. Either $pp \to \gamma X$ or $pp \to \gamma +$jet. Sensitive to gluon ($qg \to q + \gamma$) (But problems with isolation).
6. Heavy flavour production → Ask M. Cacciari.

Lots of work still to do!

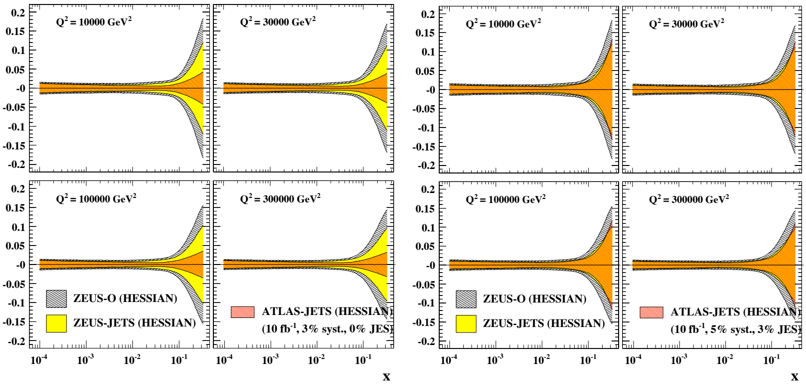
Jet production (ATLAS)

Inclusive jets in ATLAS PDF fit (from Mandy Cooper-Sarkar)



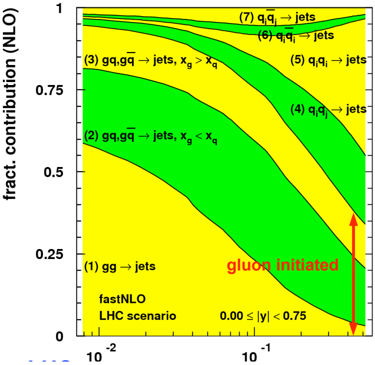
Jet production (ATLAS)

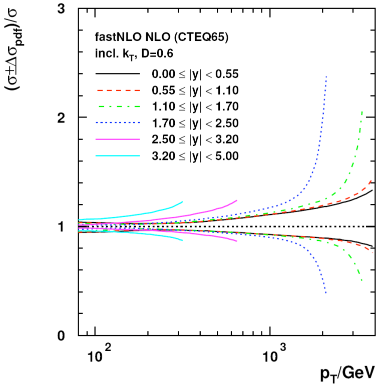
Inclusive jets in ATLAS PDF fit (from Mandy Cooper-Sarkar)



Jet production (CMS)

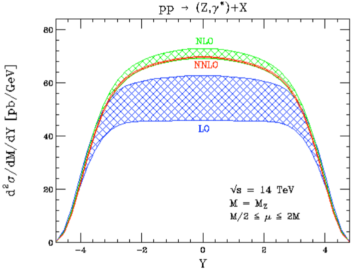
Inclusive jets in CMS (K .Rabbertz) → Poor knowledge of gluon





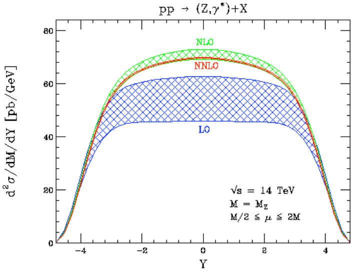
Vector boson production

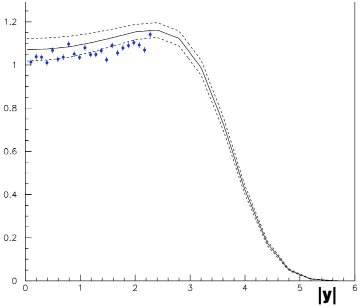
Very good control of theoretical uncertainties.
ATLAS-PDFs fit with simulated data (C. Gwenlan)



Vector boson production

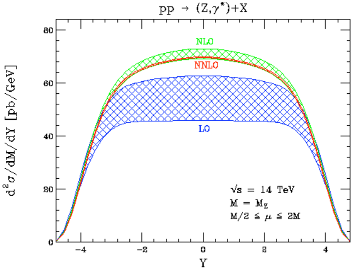
Very good control of theoretical uncertainties.
ATLAS-PDFs fit with simulated data (C. Gwenlan)

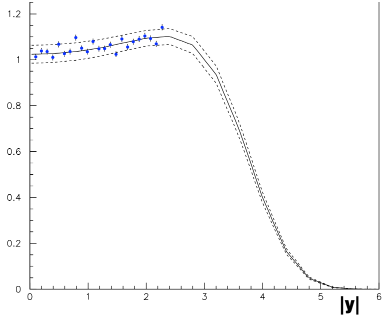




Vector boson production

Very good control of theoretical uncertainties.
ATLAS-PDFs fit with simulated data (C. Gwenlan)





PDFs4LHC workshop

A new series of workshops devoted to parton distributions at the LHC. Main topics:

- To agree on a common procedure for the use of PDFs for all the LHC experiments
- To identify the needs for and input from first data of the LHC to PDF determination, and how to present LHC data to make it maximally useful.
- Determination and evaluation of PDF uncertainties
- PDFs for Monte Carlo generators (Modified LO PDFs - No MSR, NLO α_s?)

Initial workshop: CERN, february 22-23. See slides in:

http://indico.cern.ch/conferenceDisplay.py?confId=27439

PDFs4LHC workshop

A new series of workshops devoted to parton distributions at the LHC.
Main topics:

- To agree on a common procedure for the use of PDFs for all the LHC experiments
- To identify the needs for and input from first data of the LHC to PDF determination, and how to present LHC data to make it maximally useful.
- Determination and evaluation of PDF uncertainties
- PDFs for Monte Carlo generators (Modified LO PDFs - No MSR, NLO α_s?)

Initial workshop: CERN, february 22-23. See slides in:

http://indico.cern.ch/conferenceDisplay.py?confId=27439

PDFs4LHC workshop

A new series of workshops devoted to parton distributions at the LHC.
Main topics:

- To agree on a common procedure for the use of PDFs for all the LHC experiments
- To identify the needs for and input from first data of the LHC to PDF determination, and how to present LHC data to make it maximally useful.
- Determination and evaluation of PDF uncertainties
- PDFs for Monte Carlo generators (Modified LO PDFs - No MSR, NLO α_s?)

Initial workshop: CERN, february 22-23. See slides in:

http://indico.cern.ch/conferenceDisplay.py?confId=27439

PDFs4LHC workshop

A new series of workshops devoted to parton distributions at the LHC.
Main topics:

- To agree on a common procedure for the use of PDFs for all the LHC experiments
- To identify the needs for and input from first data of the LHC to PDF determination, and how to present LHC data to make it maximally useful.
- Determination and evaluation of PDF uncertainties
- PDFs for Monte Carlo generators (Modified LO PDFs - No MSR, NLO α_s?)

Initial workshop: CERN, february 22-23. See slides in:

http://indico.cern.ch/conferenceDisplay.py?confId=27439

PDFs4LHC workshop

A new series of workshops devoted to parton distributions at the LHC.
Main topics:

- To agree on a common procedure for the use of PDFs for all the LHC experiments
- To identify the needs for and input from first data of the LHC to PDF determination, and how to present LHC data to make it maximally useful.
- Determination and evaluation of PDF uncertainties
- PDFs for Monte Carlo generators (Modified LO PDFs - No MSR, NLO α_s?)

Initial workshop: CERN, february 22-23. See slides in:

http://indico.cern.ch/conferenceDisplay.py?confId=27439

PDFs4LHC workshop

A new series of workshops devoted to parton distributions at the LHC.
Main topics:

- To agree on a common procedure for the use of PDFs for all the LHC experiments
- To identify the needs for and input from first data of the LHC to PDF determination, and how to present LHC data to make it maximally useful.
- Determination and evaluation of PDF uncertainties
- PDFs for Monte Carlo generators (Modified LO PDFs - No MSR, NLO α_s?)

Initial workshop: CERN, february 22-23. See slides in:

`http://indico.cern.ch/conferenceDisplay.py?confId=27439`

Summary

- PDF uncertainties are one of the dominant sources of uncertainties for many relevant LHC processes
- Neural network PDFs ready to explore phenomenological implications.
- Precision phenomenology at the LHC requires feedback: use LHC measurements to further constrain PDFs
- Interplay between experimentalists and theorists crucial to improve PDFs at the LHC!

Suggestions welcome!

Summary

- PDF uncertainties are one of the dominant sources of uncertainties for many relevant LHC processes
- Neural network PDFs ready to explore phenomenological implications.
- Precision phenomenology at the LHC requires feedback: use LHC measurements to further constrain PDFs
- Interplay between experimentalists and theorists crucial to improve PDFs at the LHC!

Suggestions welcome!

EXTRA MATERIAL

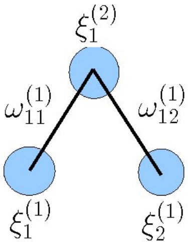
Neural nets

What is a neural net?

$$\xi_i^{(l+1)} = g\left(\sum_{j=1}^{n(l)} \omega_{ij}^{(l)} \xi_j^l\right), \quad g(x) = \frac{1}{1+e^x}, \quad l = 2, \ldots, L$$

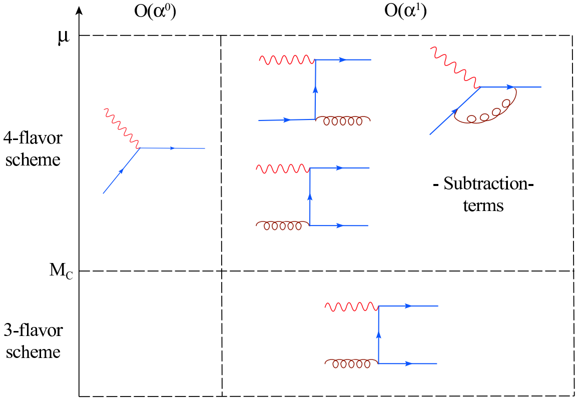
where $\omega_{ij}^{(l)}$ are the *weights* and $\xi_i^{(l+1)}$ the *activation state* of each neuron.
Simplest neural network: Architecture 2-1

$$\xi_1^{(2)} = \left[1 + \exp\left(\omega_{11}^{(1)}\xi_1^{(1)} + \omega_{12}^{(1)}\xi_2^{(1)}\right)\right]^{-1}$$



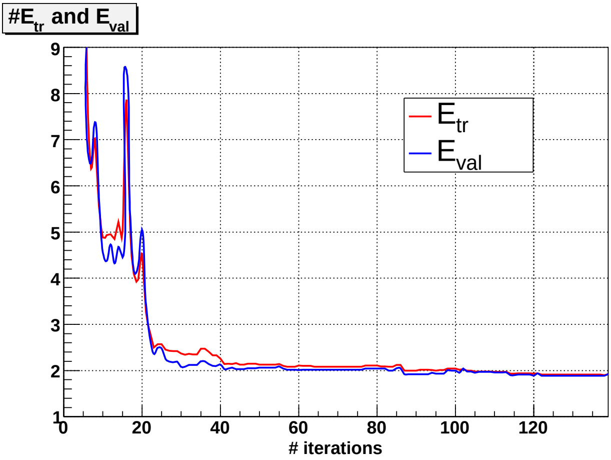
Heavy quark mass effects

Finite m_c, m_b effects important near threshold (from hep-ph/0611254)



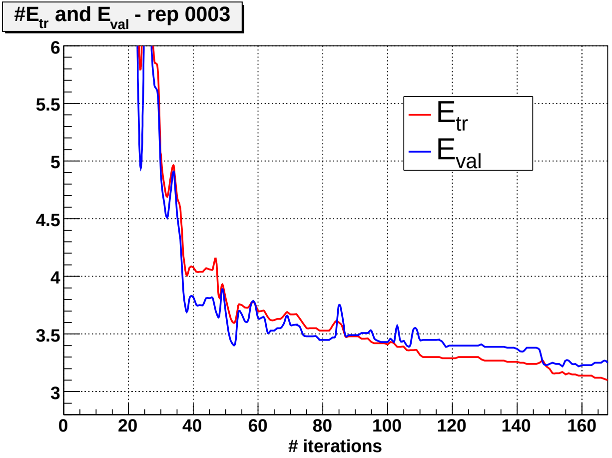
Dynamical stopping

Stop minimization before learning statistical fluctuations (overlearning)



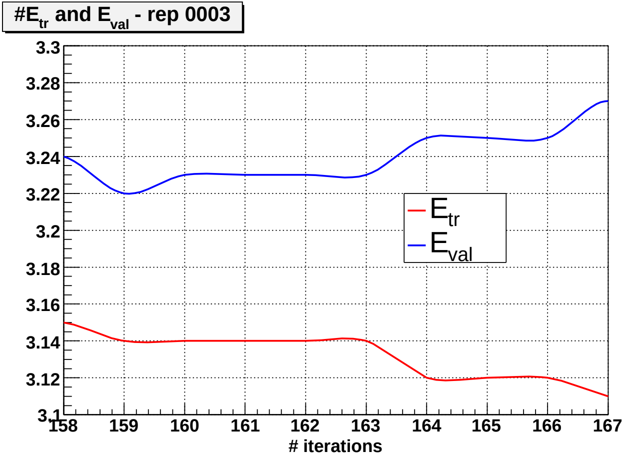
Dynamical stopping

Stop minimization before learning statistical fluctuations (overlearning)

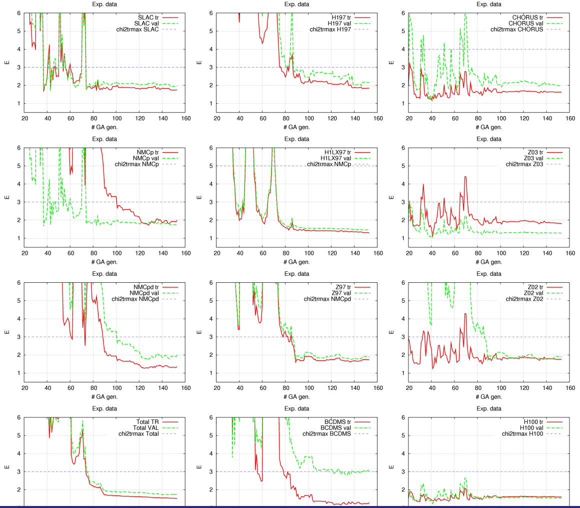


Dynamical stopping

Stop minimization before learning statistical fluctuations (overlearning)

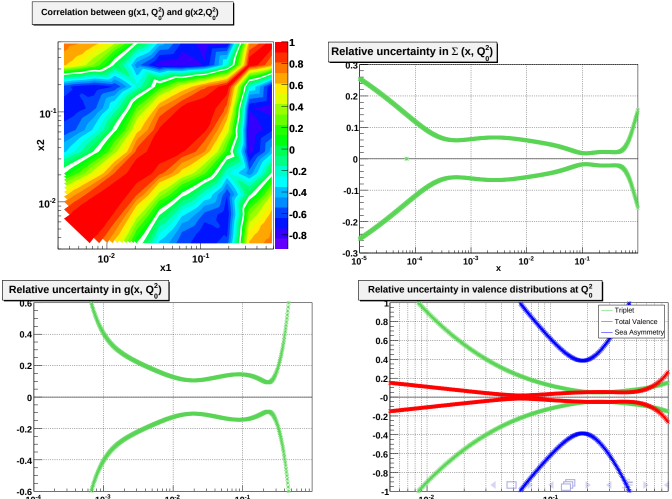


Dynamical stopping



The NNPDF approach - Preliminary results

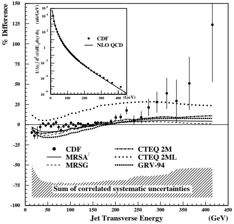
Relative uncertainties and correlations



PDF uncertainties

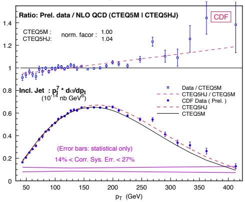
CASE STUDY I: THE CDF LARGE E_T JETS
CDF 1995

- DISCREPANCY BETWEEN QCD CALCULATION AND CDF JET DATA (1995)
- EVIDENCE FOR QUARK COMPOSITENESS?
- BUT NO INFO ON PARTON UNCERTAINTY $\Rightarrow$ RESULT STRONGLY DEPENDS ON GLUON AT $x \gtrsim 0.1$

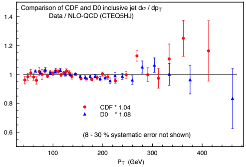


DISCREPANCY REMOVED IF JET DATA INCLUDED IN THE FIT

NEW CTEQ FIT (1996)



FINAL CTEQ FIT (1998)

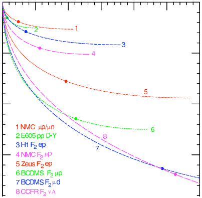


S. Forte, CERN 02-08

STANDARD SOLUTION: CTEQ TOLERANCE CRITERION

- DETERMINE EIGENVECTORS OF χ^2 PARABOLOID
- DETERMINE 90% C.L. FOR EACH EXPT. ALONG EACH EIGENVECTOR
- DETERMINE MOST RESTRICTIVE INTERVAL ABOUT GLOBAL MINIMUM (TOLERANCE)
- $\Delta\chi^2 = 100$

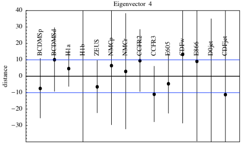
MINIMUM χ^2_i VS GLOBAL χ^2



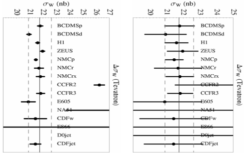
Collins, Pumplin 2001

CCFR, BCDMS INCOMPATIBLE

TOLERANCE PLOT FOR 4TH EIGENVEC.



σ_W: ONE σ VS. TOLERANCE



(CTEQ6, 2002-2007)

S. Forte, CERN 02-08

WHAT'S THE PROBLEM?

- FOR A SINGLE QUANTITY, WE QUOTE 1 SIGMA ERRORS: VALUE± ERROR
- FOR A PAIR OF NUMBERS, WE QUOTE A 1 SIGMA ELLIPSE
- FOR A FUNCTION, WE NEED AN "ERROR BAR" IN A SPACE OF FUNCTIONS

MUST DETERMINE THE PROBABILITY DENSITY (MEASURE) $\mathcal{P}[f_i(x)]$
IN THE SPACE OF PARTON DISTRIBUTION FUNCTIONS $f_i(x)$ (i=quark, antiquark gluon)

EXPECTATION VALUE OF $\sigma\left[f_i(x)\right] \Rightarrow$ FUNCTIONAL INTEGRAL

$$\Big\langle \sigma\left[f_i(x)\right] \Big\rangle = \int \mathcal{D}f_i \, \sigma\left[f_i(x)\right] \, \mathcal{P}[f_i],$$

MUST DETERMINE AN INFINITE–DIMENSIONAL OBJECT FROM A FINITE SET OF DATA POINTS

S. Forte, CERN 02-08

THE BAYESIAN MONTE CARLO (GIELE, KOSOWER, KELLER 2001)

- generate a Monte-Carlo sample of fcts. with "reasonable" prior distn. (e.g. an available parton set) → representation of probability functional $\mathcal{P}[f_i]$
- calculate observables with functional integral
- update probability using Bayesian inference on MC sample: better agreement with data → more functions in sample
- iterate until convergence achieved

PROBLEM IS MADE FINITE-DIMENSIONAL BY THE CHOICE OF PRIOR, BUT RESULT DO NOT DEPEND ON THE CHOICE IF SUFFICIENTLY GENERAL

HARD TO HANDLE "FLAT DIRECTIONS"

(Monte Carlo replicas which lead to same agreement with data);

COMPUTATIONALLY VERY INTENSIVE;

DIFFICULT TO ACHIEVE INDEP. FROM PRIOR

PDF4LHC

A. de Roeck at PDF4LHC

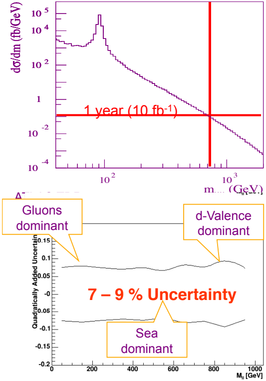
PDF4LHC

- Issues for PDFs (list to be extended/needs perhaps priorities)
  - PDFs both for calculations and Monte Carlos (NNLO/NLO/LO/other?)
    - Maybe one of the most pressing issues to come to a good workable solution, but will take some time before we will switch to drastically different approaches.
  - Data to be included in the PDFs.: Selection of data to be used; new data (F_L and other); combined data (H1/ZEUS); extracting more from the data that we have (Largely done within HERA-LHC context)
  - Discussion on the uncertainties on the PDFs
  - Theoretical uncertainties and regions/processes where they matter
    - Heavy flavour treatment
    - Low-x (High-x)
    - Other PDFs like uPDFs (and GPDs)
    - NLO, NNLO…
  - Other…. ⇒ Your input input is important !!!

HERA is still alive!

Mandy Cooper Sarkar talk at PDF4LHC

But not all discovery physics is strongly compromised: e.g PDF Uncertainty in High-mass Drell-Yan- won't stop us seeing Zprimes



and PDF uncertainties don't affect the Higgs discovery potential too badly