PDF Uncertainties in Top Production at the LHC

Elements for a discussion

Juan Rojo

INFN, Sezione di Milano

CERN Top TH Institute
08/06/2009

MOTIVATION

Issues in standard PDF determinations

- Standard PDF determinations suffer of several drawbacks
  1. Fixed functional forms, $q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1+\ldots)$.
     Are they flexible enough?
  2. Artificially large tolerances $\Delta\chi^2 \gg 1$
     Are they really needed due to incompatible data?
  3. Gaussian linear error propagation
     Is this really enough for all observables?
- Large tolerances $\to$ Error blow-up by a factor $S = \sqrt{\Delta\chi^2/2.7}$
  $\to S_{\rm cteq} \sim 6$, $S_{\rm mstw} \sim 4.5$ both in input data and in PDFs
- Summary $\to$ Both the PDF input parametrization (and flavour assumptions) and the statistical treatment (value of $\Delta\chi^2$) need to be tuned to experimental data ....
- ... which implies that adding more data sometimes results in no error reduction or even an increase of PDF uncertainties (MSTW, CQ09G)

Issues in standard PDF determinations

- Standard PDF determinations suffer of several drawbacks
  1. Fixed functional forms, $q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1 + \ldots)$.
     Are they flexible enough?
  2. Artificially large tolerances $\Delta\chi^2 \gg 1$
     Are they really needed due to incompatible data?
  3. Gaussian linear error propagation
     Is this really enough for all observables?
- Large tolerances $\to$ Error blow-up by a factor $S = \sqrt{\Delta\chi^2/2.7}$
  $\to S_{\rm cteq} \sim 6$, $S_{\rm mstw} \sim 4.5$ both in input data and in PDFs
- Summary $\to$ Both the PDF input parametrization (and flavour assumptions) and the statistical treatment (value of $\Delta\chi^2$) need to be tuned to experimental data ....
- ... which implies that adding more data sometimes results in no error reduction or even an increase of PDF uncertainties (MSTW, CQ09G)

Issues in standard PDF determinations

- Standard PDF determinations suffer of several drawbacks
  1. Fixed functional forms, $q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1+\ldots)$.
     Are they flexible enough?
  2. Artificially large tolerances $\Delta\chi^2 \gg 1$
     Are they really needed due to incompatible data?
  3. Gaussian linear error propagation
     Is this really enough for all observables?
- Large tolerances $\longrightarrow$ Error blow-up by a factor $S = \sqrt{\Delta\chi^2/2.7}$
  $\longrightarrow S_{\rm cteq} \sim 6$, $S_{\rm mstw} \sim 4.5$ both in input data and in PDFs
- Summary $\longrightarrow$ Both the PDF input parametrization (and flavour assumptions) and the statistical treatment (value of $\Delta\chi^2$) need to be tuned to experimental data ....
- ... which implies that adding more data sometimes results in no error reduction or even an increase of PDF uncertainties (MSTW, CQ09G)

Issues in standard PDF determinations

- Standard PDF determinations suffer of several drawbacks
  1. Fixed functional forms, $q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1+\ldots)$.
     Are they flexible enough?
  2. Artificially large tolerances $\Delta\chi^2 \gg 1$
     Are they really needed due to incompatible data?
  3. Gaussian linear error propagation
     Is this really enough for all observables?
- Large tolerances $\rightarrow$ Error blow-up by a factor $S = \sqrt{\Delta\chi^2/2.7}$
  $\rightarrow S_{\rm cteq} \sim 6$, $S_{\rm mstw} \sim 4.5$ both in input data and in PDFs
- Summary $\rightarrow$ Both the PDF input parametrization (and flavour assumptions) and the statistical treatment (value of $\Delta\chi^2$) need to be tuned to experimental data ....
- ... which implies that adding more data sometimes results in no error reduction or even an increase of PDF uncertainties (MSTW, CQ09G)

Issues in standard PDF determinations

- Standard PDF determinations suffer of several drawbacks
  1. Fixed functional forms, $q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1+\ldots)$.
     Are they flexible enough?
  2. Artificially large tolerances $\Delta\chi^2 \gg 1$
     Are they really needed due to incompatible data?
  3. Gaussian linear error propagation
     Is this really enough for all observables?
- Large tolerances $\longrightarrow$ Error blow-up by a factor $S = \sqrt{\Delta\chi^2/2.7}$
  $\longrightarrow S_{\rm cteq} \sim 6$, $S_{\rm mstw} \sim 4.5$ both in input data and in PDFs
- Summary $\longrightarrow$ Both the PDF input parametrization (and flavour assumptions) and the statistical treatment (value of $\Delta\chi^2$) need to be tuned to experimental data ....
- ... which implies that adding more data sometimes results in no error reduction or even an increase of PDF uncertainties (MSTW, CQ09G)

Issues in standard PDF determinations

- Standard PDF determinations suffer of several drawbacks
  1. Fixed functional forms, $q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1+\ldots)$.
     Are they flexible enough?
  2. Artificially large tolerances $\Delta\chi^2 \gg 1$
     Are they really needed due to incompatible data?
  3. Gaussian linear error propagation
     Is this really enough for all observables?
- Large tolerances $\longrightarrow$ Error blow-up by a factor $S = \sqrt{\Delta\chi^2/2.7}$
  $\longrightarrow S_{\rm cteq} \sim 6$, $S_{\rm mstw} \sim 4.5$ both in input data and in PDFs
- Summary $\rightarrow$ Both the PDF input parametrization (and flavour assumptions) and the statistical treatment (value of $\Delta\chi^2$) need to be tuned to experimental data ....
- ... which implies that adding more data sometimes results in no error reduction or even an increase of PDF uncertainties (MSTW, CQ09G)

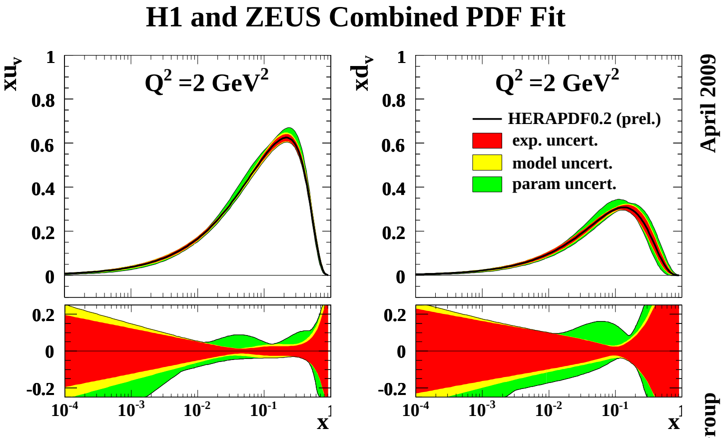
Issues in standard PDF determinations

- Standard PDF determinations suffer of several drawbacks
  1. Fixed functional forms, $q_i(x, Q_0^2) = A_i x^{b_i} (1-x)^{c_i} (1 + \ldots)$.
     Are they flexible enough?
  2. Artificially large tolerances $\Delta\chi^2 \gg 1$
     Are they really needed due to incompatible data?
  3. Gaussian linear error propagation
     Is this really enough for all observables?
- Large tolerances $\longrightarrow$ Error blow-up by a factor $S = \sqrt{\Delta\chi^2/2.7}$
  $\longrightarrow S_{\rm cteq} \sim 6$, $S_{\rm mstw} \sim 4.5$ both in input data and in PDFs
- Summary $\longrightarrow$ Both the PDF input parametrization (and flavour assumptions) and the statistical treatment (value of $\Delta\chi^2$) need to be tuned to experimental data ....
- ... which implies that adding more data sometimes results in no error reduction or even an increase of PDF uncertainties (MSTW, CQ09G)

Issues in standard PDF determinations (II)

1. Recent analysys (Pumplin, arXiv:0904.2425) → No tension between datasets in the CTEQ analysis found which requires $\Delta\chi^2 = 100$ (reescaling experimental data by a factor 6)
2. The latest HERA0.2 PDF analysis (with $\Delta\chi^2 = 1$) leads to errors consistent with CTEQ (with $\Delta\chi^2 = 100$) only after accounting for (sizable) parametrization uncertainties.

Parametrization bias is a very important problem for a meaningful interpretation of PDF uncertainties!



THE NNPDF APPROACH

NNPDF collaboration

R.D.Ball[1], L.Del Debbio[1], S.Forte[2], A.Guffanti[3], J.I.Latorre[4], A. Piccione[2], J. R.[2], M. Ubiali[1]

[1] PPT Group, School of Physics, University of Edinburgh
[2] Dipartimento di Fisica, Università di Milano
[3] Physikalisches Institut, Albert-Ludwigs-Universität Freiburg
[4] Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona

The NNPDF approach

- Generate $N_{\rm rep}$ Monte Carlo replicas $F_i^{(\rm art)(k)}$ of the original data $F_i^{(\rm exp)}$
  Avoid gaussian/linearized assumptions

$$F_i^{(\rm art)(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\rm exp)} + \sum_{p=1}^{N_{\rm sys}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

- Evolve each PDF parametrized with Neural Networks $\to$ Unbiased parametrization

$$F_i^{(\rm net)(k)}(x,Q^2) = C_{i\alpha}(x,\alpha(Q^2)) \otimes q_\alpha^{(\rm net)(k)}\left(x,Q^2\right)$$

- Minimization of χ^2 with Genetic Algorithms. + Dynamical Stopping:

$$\chi^{2(k)} = \frac{1}{N_{\rm dat}} \sum_{i,j=1}^{N_{\rm dat}} \left(F_i^{(\rm art)(k)} - F_i^{(\rm net)(k)}\right)\left({\rm cov}_{ij}^{-1}\right)\left(F_j^{(\rm art)(k)} - F_j^{(\rm net)(k)}\right)$$

The NNPDF approach

- Generate N_{rep} Monte Carlo replicas $F_i^{(\mathrm{art})(k)}$ of the original data $F_i^{(\mathrm{exp})}$
  Avoid gaussian/linearized assumptions

$$F_i^{(\mathrm{art})(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\mathrm{exp})} + \sum_{p=1}^{N_{\mathrm{sys}}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

- Evolve each PDF parametrized with Neural Networks → Unbiased parametrization

$$F_i^{(\mathrm{net})(\mathrm{k})}(x,Q^2) = C_{i\alpha}(x,\alpha(Q^2)) \otimes q_{\alpha}^{(\mathrm{net})(k)}\left(x,Q^2\right)$$

- Minimization of χ^2 with Genetic Algorithms. + Dynamical Stopping:

$$\chi^{2(k)} = \frac{1}{N_{\mathrm{dat}}}\sum_{i,j=1}^{N_{\mathrm{dat}}}\left(F_i^{(\mathrm{art})(k)} - F_i^{(\mathrm{net})(k)}\right)\left(\mathrm{cov}_{ij}^{-1}\right)\left(F_j^{(\mathrm{art})(k)} - F_j^{(\mathrm{net})(k)}\right)$$

The NNPDF approach

- Generate N_{rep} Monte Carlo replicas $F_i^{(\mathrm{art})(k)}$ of the original data $F_i^{(\mathrm{exp})}$
  Avoid gaussian/linearized assumptions

$$F_i^{(\mathrm{art})(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\mathrm{exp})} + \sum_{p=1}^{N_{\mathrm{sys}}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

- Evolve each PDF parametrized with Neural Networks $\rightarrow$ Unbiased parametrization

$$F_i^{(\mathrm{net})(\mathrm{k})}(x, Q^2) = C_{i\alpha}(x, \alpha(Q^2)) \otimes q_\alpha^{(\mathrm{net})(k)}\left(x, Q^2\right)$$

- Minimization of χ^2 with Genetic Algorithms. + Dynamical Stopping:

$$\chi^{2(k)} = \frac{1}{N_{\mathrm{dat}}}\sum_{i,j=1}^{N_{\mathrm{dat}}}\left(F_i^{(\mathrm{art})(k)} - F_i^{(\mathrm{net})(k)}\right)\left(\mathrm{cov}_{ij}^{-1}\right)\left(F_j^{(\mathrm{art})(k)} - F_j^{(\mathrm{net})(k)}\right)$$

The NNPDF approach

- Generate $N_{\rm rep}$ Monte Carlo replicas $F_i^{(\rm art)(k)}$ of the original data $F_i^{(\rm exp)}$
  Avoid gaussian/linearized assumptions

$$F_i^{(\rm art)(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\rm exp)} + \sum_{p=1}^{N_{\rm sys}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

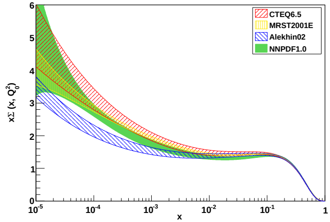
- Evolve each PDF parametrized with Neural Networks → Unbiased parametrization

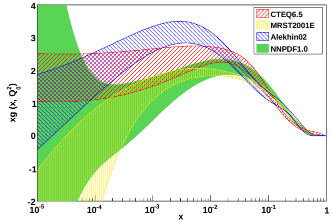
$$F_i^{(\rm net)(k)}(x,Q^2) = C_{i\alpha}(x,\alpha(Q^2)) \otimes q_\alpha^{(\rm net)(k)}\left(x,Q^2\right)$$

- Minimization of χ^2 with Genetic Algorithms. + Dynamical Stopping:

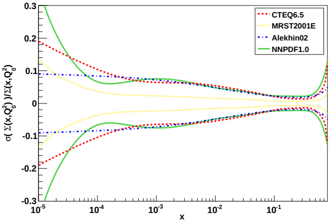
$$\chi^{2(k)} = \frac{1}{N_{\rm dat}}\sum_{i,j=1}^{N_{\rm dat}}\left(F_i^{(\rm art)(k)} - F_i^{(\rm net)(k)}\right)\left({\rm cov}_{ij}^{-1}\right)\left(F_j^{(\rm art)(k)} - F_j^{(\rm net)(k)}\right)$$

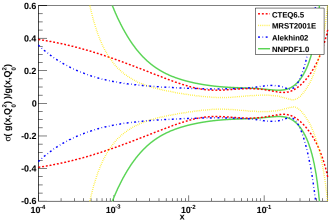
NNPDF1.0: First NNPDF parton set





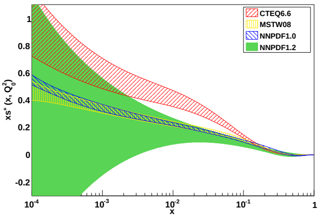
Note blow-up of uncertaities in extrapolation regions

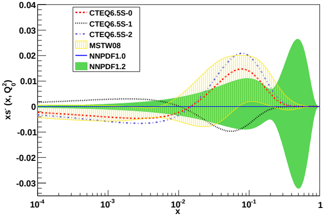


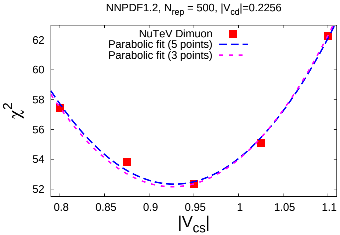


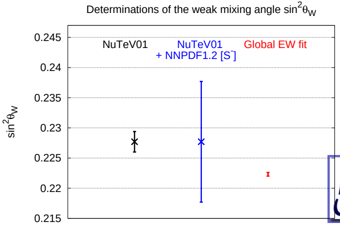
NNPDF1.2: Precision strange & EW parameters

Unbiased determination of the strange PDFs from NuTeV dimuons
Most precise ever direct determination of $|V_{cs}|$ + Nutev anomaly disapears from large uncertainties in $[S^-]$



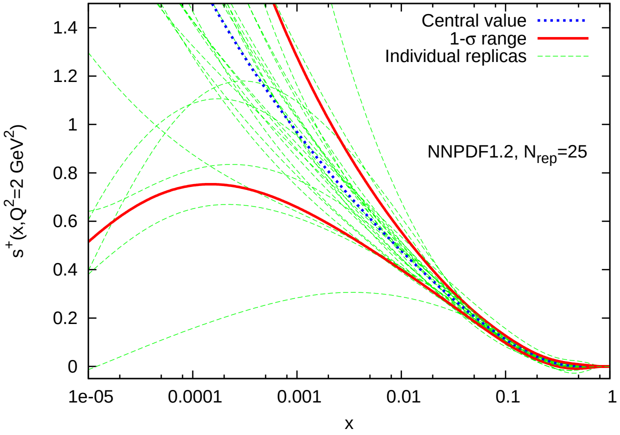






Neural networks: free from parametrization bias

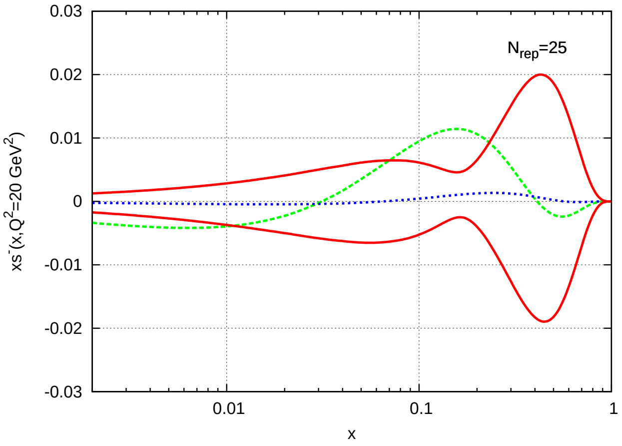
Artificial neural network follow experimental data trends without any ad-hoc theoretical bias
Moderate spread in data region, blow up in extrapolation



Neural networks: free from parametrization bias

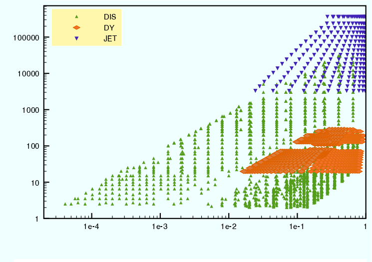
Artificial neural network follow experimental data trends without any ad-hoc theoretical bias

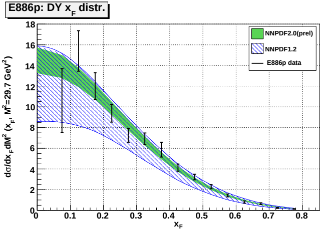
For S^- multiple nodes allowed if preferred by data (very difficult in the standard approach)

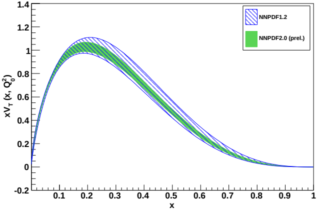


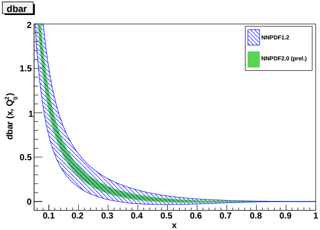
NNPDF2.0: Towards a global PDF analysis (Preliminary)!

PDF uncertainties reduced as expected from statistics without any modification of PDF parametrizations







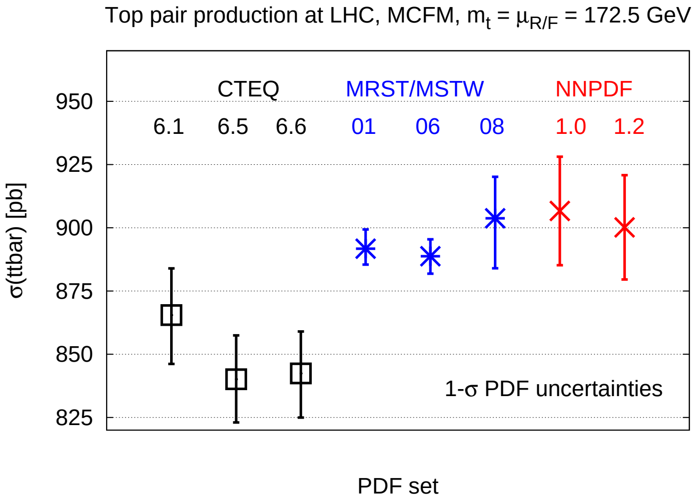


TOP QUARK PRODUCTION AND PDF UNCERTAINTIES

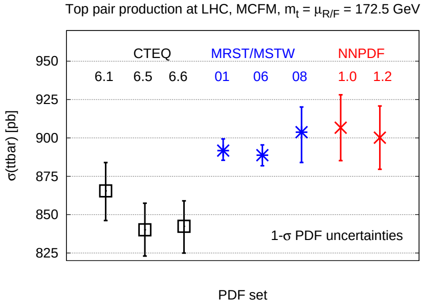
Beware: very preliminary results!!!

$\sigma^{\text{tot}}(t\bar{t})$ at the LHC

Compare results from different PDF sets (errors: 68% confidence levels)



$\sigma^{\mathrm{tot}}(t\bar{t})$ at the LHC



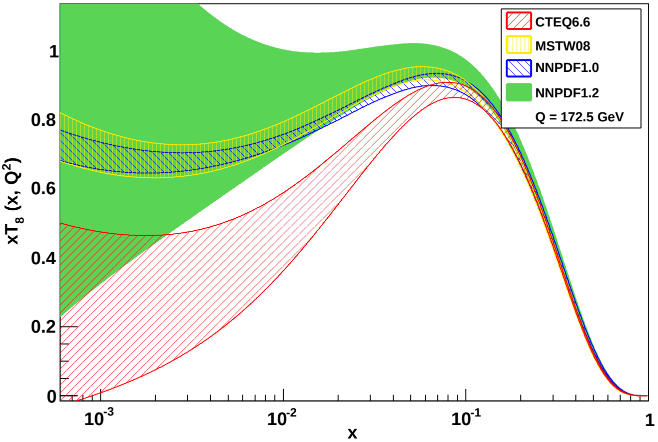
- Predictions from MTW08 and CTEQ6.6 still inconsistent
- NNPDF1.2 consistent with CTEQ6.1 (with ZM-VFN schemes for HQ)

1. Systematic downwards shift by $1-\sigma$ in CTEQ from 6.1 to 6.5/6.6 $\longrightarrow$ General Mass scheme for heavy quarks
2. Stability of central value MRST sets $\longrightarrow$ Despite change in HQ scheme which shifted $\sim 7\%$ $\sigma_{W/Z}$ in 2006!
3. Note sizable increase in PDF errors in MSTW08, despite using much more data $\longrightarrow$ Example of parametrization bias!

$\sigma^{\text{tot}}\,(t\bar{t})$ at the LHC

Some PDF combinations differ between MSTW08 and CTEQ6.6 (strangeness)

$$xT_8(x, m_t^2) = x\left(u + \bar{u} + d + \bar{d} - 2(s + \bar{s})\right)$$



Heavy quark production in hadronic collisions

The heavy quark pair production can be written

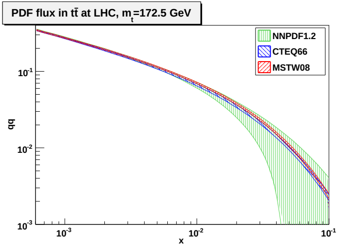
$$\sigma\left(S, m_H^2\right) = \frac{\alpha_s^2\left(\mu^2\right)}{m_H^2} \sum_{ij} \int_{\rho^H}^1 \frac{d\tau}{\tau} \Phi_{ij}\left(\tau, \mu\right) \widetilde{\sigma}_{ij}\left(\frac{\rho^H}{\tau}, \frac{\mu^2}{m_H^2}\right) \ , \rho^H = \frac{4m_H^2}{S}$$

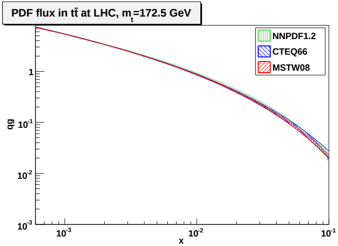
in terms of partonic fluxes

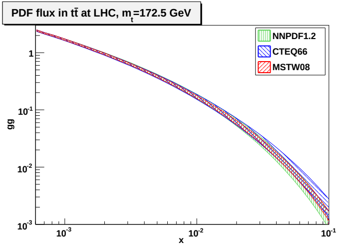
$$\Phi_{ij}\left(\tau, \mu\right) = \int_0^1 dx_1 \int_0^1 dx_2 q_i^{(A)}(x_1, \mu) q_j^{(B)}(x_2, \mu) \delta\left(x_1 x_2 - \tau\right)$$

at the LHC, for $m_t = 172.5$ GeV $\rightarrow \rho^H = 6 \cdot 10^{-4}$

PDF fluxes

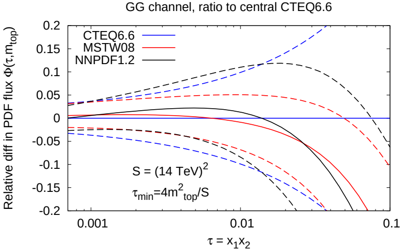


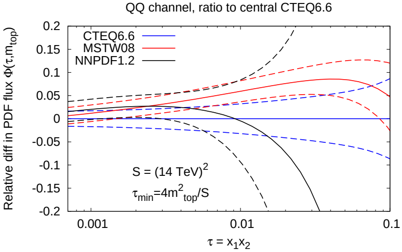


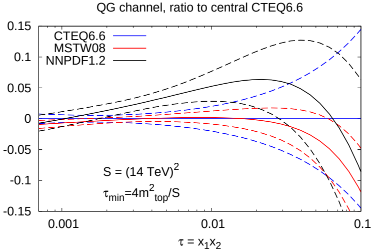


Dominance of the QG channel

PDF fluxes (II)







Elements for a discussion

1. Do we understand systematic difference between CTEQ6.6 and MSTW08, despite using similar data and same HQ scheme?
2. Difficult to identify the source of the discrepancy at the PDF level ...
3. HQ scheme shifts cross-section by $\sim$ 1-σ in CTEQ. Is this result independent of the HQ prescription adopted?
4. Expect soon results from a NNPDF global fit with GM scheme $\longrightarrow$ Strong cross-check of the stability of the prediction for $\sigma\left(t\bar{t}\right)$ and its uncertainties.

Thanks for your attention!

Elements for a discussion

1. Do we understand systematic difference between CTEQ6.6 and MSTW08, despite using similar data and same HQ scheme?
2. Difficult to identify the source of the discrepancy at the PDF level ...
3. HQ scheme shifts cross-section by $\sim$ 1-σ in CTEQ. Is this result independent of the HQ prescription adopted?
4. Expect soon results from a NNPDF global fit with GM scheme $\longrightarrow$ Strong cross-check of the stability of the prediction for $\sigma\left(t\bar{t}\right)$ and its uncertainties.

Thanks for your attention!

Elements for a discussion

1. Do we understand systematic difference between CTEQ6.6 and MSTW08, despite using similar data and same HQ scheme?
2. Difficult to identify the source of the discrepancy at the PDF level ...
3. HQ scheme shifts cross-section by $\sim$ 1-σ in CTEQ. Is this result independent of the HQ prescription adopted?
4. Expect soon results from a NNPDF global fit with GM scheme $\longrightarrow$ Strong cross-check of the stability of the prediction for $\sigma\,(t\bar{t})$ and its uncertainties.

Thanks for your attention!

Elements for a discussion

1. Do we understand systematic difference between CTEQ6.6 and MSTW08, despite using similar data and same HQ scheme?
2. Difficult to identify the source of the discrepancy at the PDF level ...
3. HQ scheme shifts cross-section by $\sim$ 1-σ in CTEQ. Is this result independent of the HQ prescription adopted?
4. Expect soon results from a NNPDF global fit with GM scheme $\longrightarrow$ Strong cross-check of the stability of the prediction for $\sigma\left(t\bar{t}\right)$ and its uncertainties.

Thanks for your attention!

Elements for a discussion

1. Do we understand systematic difference between CTEQ6.6 and MSTW08, despite using similar data and same HQ scheme?
2. Difficult to identify the source of the discrepancy at the PDF level ...
3. HQ scheme shifts cross-section by $\sim$ 1-σ in CTEQ. Is this result independent of the HQ prescription adopted?
4. Expect soon results from a NNPDF global fit with GM scheme $\longrightarrow$ Strong cross-check of the stability of the prediction for $\sigma\,(t\bar{t})$ and its uncertainties.

Thanks for your attention!

Elements for a discussion

1. Do we understand systematic difference between CTEQ6.6 and MSTW08, despite using similar data and same HQ scheme?
2. Difficult to identify the source of the discrepancy at the PDF level ...
3. HQ scheme shifts cross-section by $\sim$ 1-σ in CTEQ. Is this result independent of the HQ prescription adopted?
4. Expect soon results from a NNPDF global fit with GM scheme $\longrightarrow$ Strong cross-check of the stability of the prediction for $\sigma\,(t\bar{t})$ and its uncertainties.

Thanks for your attention!