THE NNPDF APPROACH TO PARTON DETERMINATION: IDEAS AND RESULTS

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

Istituto Nazionale di Fisica Nucleare

ROMA

MARCH 28, 2011

PROLOGUE: THE DISCOVERY OF THE W

LAST TIME WE LOOKED FOR "NEW" PHYSICS AT A HADRON COLLIDER

THEORETICAL PREDICTION...

42 *G. Altarelli et al. / Vector boson production*

TABLE 2
Values (in nb) of the total cross sections for $W^\pm$ and Z^0 production

	$W^+ + W^-$	$W^+ + W^-$	$W^+ + W^-$	Z^0	Z^0	Z^0	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$
$\sqrt{S}$ (GeV)	GHR	DO1	DO2	GHR	DO1	DO2	GHR	DO1	DO2
540	4.2	4.3	4.1	1.3	1.3	1.2	3.1	3.4	3.5
700	6.2	6.3	6.1	2.0	1.9	1.8	3.1	3.3	3.4
1000	9.5	9.5	9.6	3.1	3.0	2.9	3.1	3.2	3.3
1300	12.5	12.5	12.9	4.0	3.9	3.9	3.1	3.2	3.3
1600	15.5	15.6	16.5	5.0	4.8	5.0	3.1	3.2	3.3

ALTARELLI, ELLIS, GRECO, MARTINELLI, 1984

...AND EXPERIMENTAL DISCOVERY

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CERN-EP/85-108
11 July 1985

W PRODUCTION PROPERTIES AT THE CERN SPS COLLIDER

UA1 Collaboration, CERN, Geneva, Switzerland

Aachen[1] - Amsterdam (NIKHEF)[2] - Annecy (LAPP)[3] - Birmingham[4] - CERN[5] - Harvard[6] - Helsinki[7] - Kiel[8] - London (Imperial College[9] and Queen Mary College[10]) - Padua[11] - Paris (Coll. de France)[12] - Riverside[13] - Rome[14] - Rutherford Appleton Lab.[15] - Saclay (CEN)[16] - Victoria[17] - Vienna[18] - Wisconsin[19] Collaboration

The corresponding experimental result for the 1984 data at $\sqrt{s}$ = 630 GeV is

$$(\sigma \cdot B)_W = 0.63 \pm 0.05\ (\pm 0.09)\ \text{nb}.$$

This is in agreement with the theoretical expectation [14] of $0.47^{+0.14}_{-0.08}$ nb. We note that the 15%

- AGREEMENT AND UNCERTAINTIES AT 20% CONSIDERED TO BE SATISFACTORY
- RESULTS FROM DIFFERENT PDF SETS DIFFER BY AT LEAST 5%
- NO WAY TO ESTIMATE PDF UNCERTAINTIES

PDFs: THEN AND NOW

PDFs IN 1984

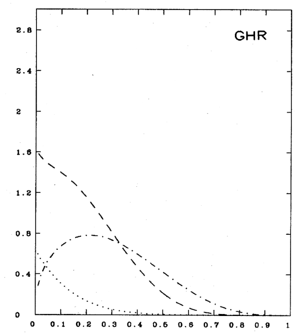


FIG. 25. Parton distributions of Glück, Hoffmann, and Reya (1982), at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and q_v (dotted line).

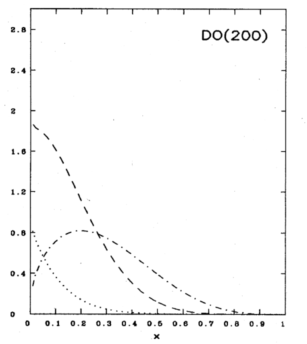


FIG. 27. "Soft-gluon" ($\Lambda=200$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

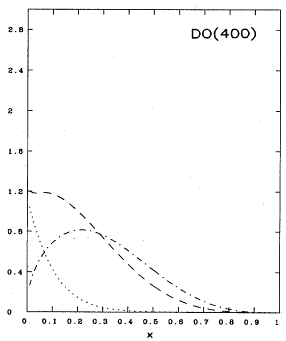
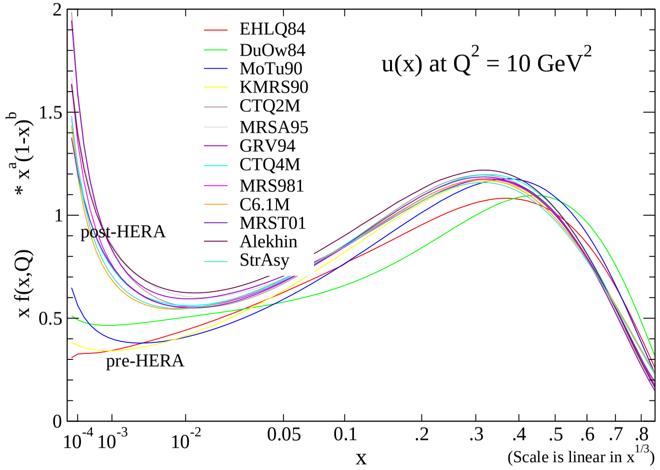


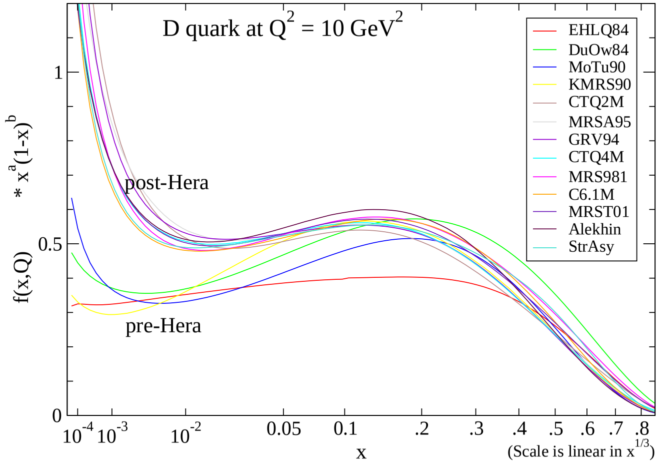
FIG. 26. "Hard-gluon" ($\Lambda=400$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

Rev. Mod. Phys., Vol. 56, No. 4, October 1984

GHR VS DUKE-OWENS

20 YEARS OF VALENCE PDFs





W.-K. TUNG, 2004

- HADRON COLLIDERS CIRCA 1985 $\Rightarrow$ QUALITATIVE QCD: DISCOVERY PHYSICS
- DIS AT NMC AND HERA 1995-2005 $\Rightarrow$ QUANTITATIVE QCD: PRECISION PHYSICS
- HADRON COLLIDERS CIRCA 2010 $\Rightarrow$ PRECISION QCD $\leftrightarrow$ NEW PHYSICS

SUMMARY

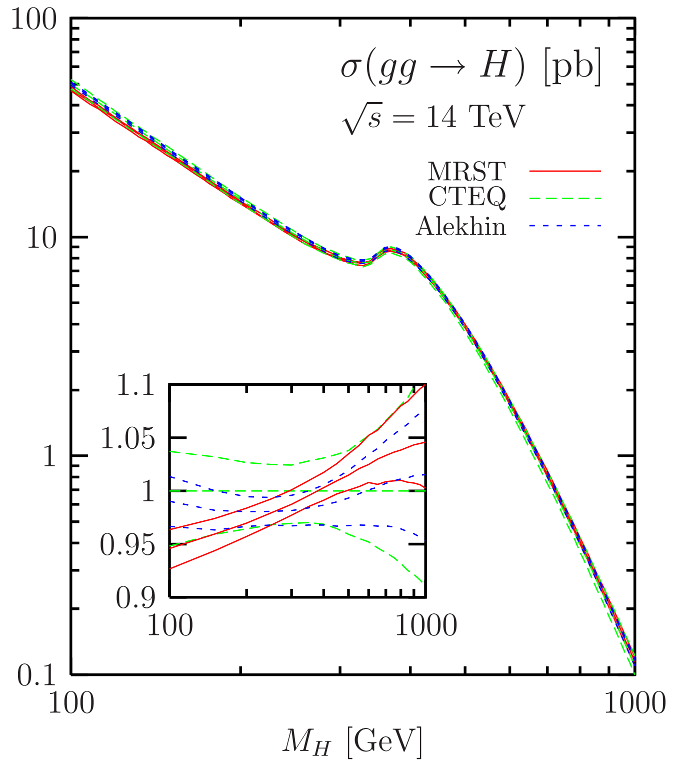
- THE STATE OF THE ART
  PDFS FOR THE LHC
- THE PROBLEM, AND A SOLUTION
  THE NNPDF APPROACH
- RESULTS
  PDFs AND STATISTICS
- PRECISION PHYSICS WITH PDFs
  DETERMINING PHYSICAL PARAMETERS
- PDFs AT THE LHC
  PREDICTIONS FOR THE LHC, PDFS FROM THE LHC
- PROGRESS
  PDFS AT NNLO

THE STATE OF THE ART

PDFS: TOWARDS LHC PHYSICS

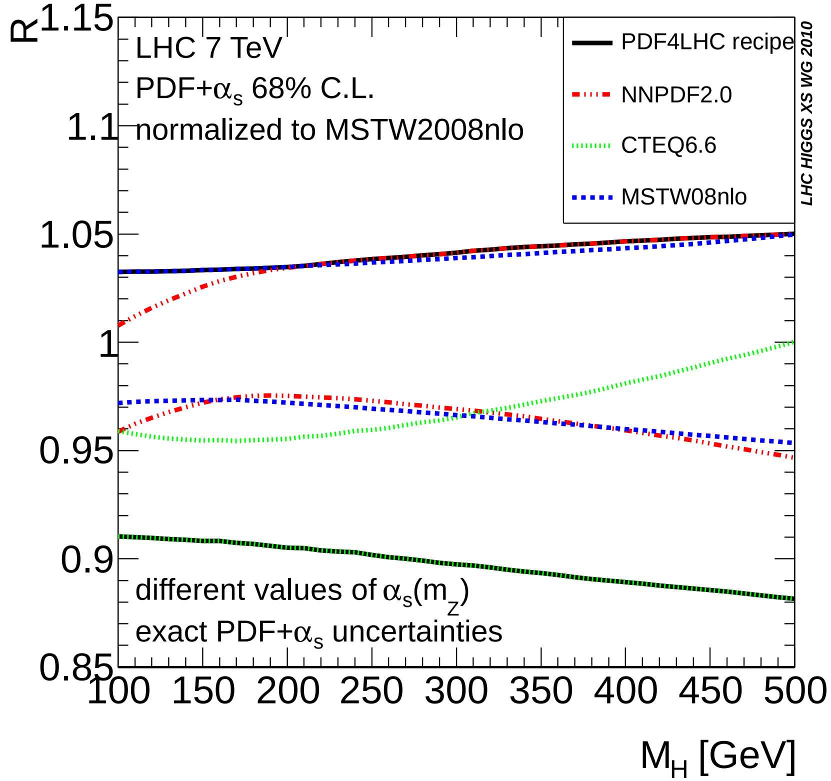
THE HIGGS CROSS SECTION

FIRST PDFs WITH UNCERTAINTIES



(Djouadi, Ferrag, 2004)

CURRENT GLOBAL PDF SETS



(Higgs WG, 2011)

PDFs WITH ERROR
(2002-2003)
CTEQ, MRST (global); Alekhin (DIS)

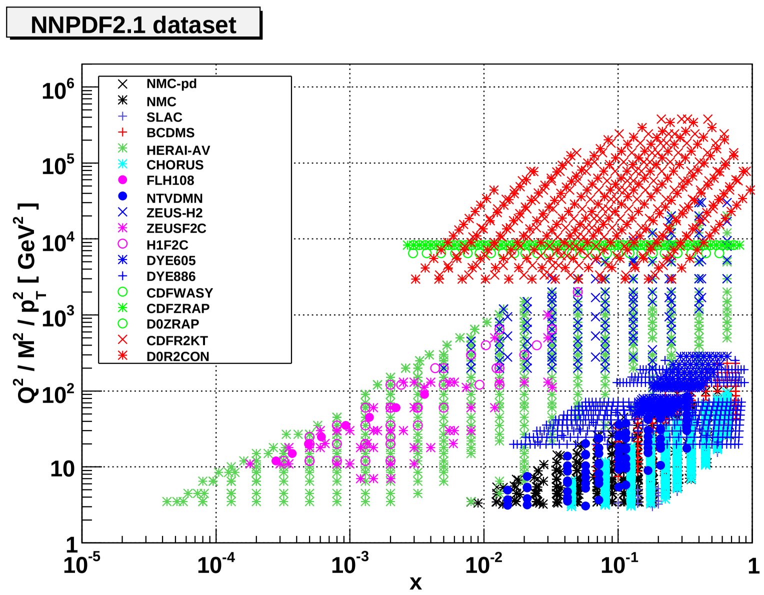
- WIDELY DIFFERENT UNCERTAINTY ESTIMATES
- UNSATISFACTORY AGREEMENT WITHIN UNCERTAINTIES

- THREE GLOBAL (DIS+HADRONIC) PDF SETS AVAILABLE
- GOOD (& IMPROVING) AGREEMENT OF CENTRAL VALUES UNCERTAINTIES

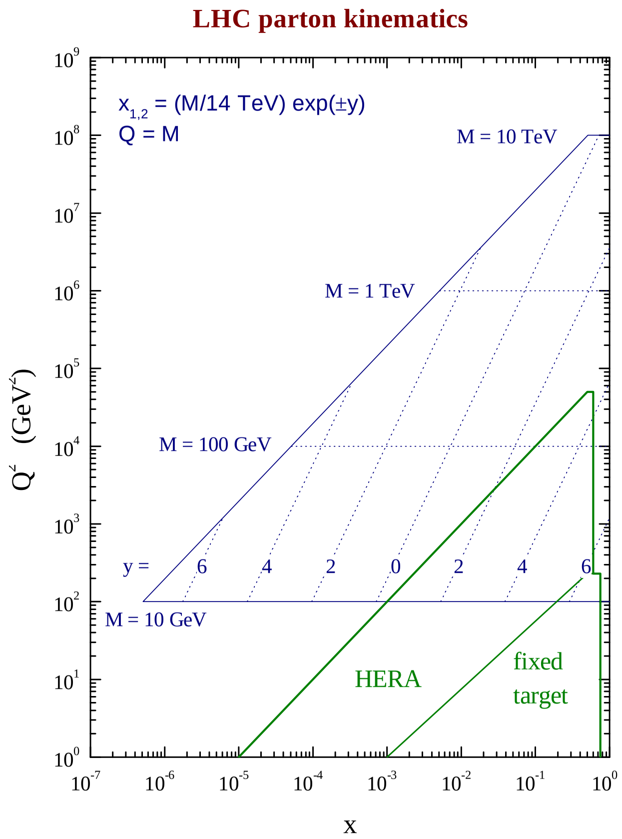
CURRENT PDF SETS

$$\sigma_X(s, M_X^2) = \sum_{a,b} \int_{x_{\min}}^1 dx_1\, dx_2\, f_{a/h_1}(x_1) f_{b/h_2}(x_2) \hat{\sigma}_{q_a q_b \to X}\left(x_1 x_2 s, M_X^2\right)$$

data for a global fit (NNPDF2.1)



LHC kinematics



- CT10: GLOBAL, NLO, SEVERAL α_s VALUES, VFN
- MSTW08: GLOBAL, NLO & NNLO, SEVERAL α_s VALUES, VFN
- NNPDF2.1: GLOBAL, NLO, SEVERAL α_s VALUES, VFN
- ALEKHIN ABKM: DIS+SOME DY, NLO & NNLO, SINGLE α_s VALUE, FFN
- HERAPDF1.5: HERA DIS+JETS, NLO & NNLO, SEVERAL α_s VALS, VFN
- SETS BASED ON MODEL ASSUMPTIONS (GRV/GJR, STATISTICAL PDFS,...)

FEATURES, TRADEOFFS AND CHOICES

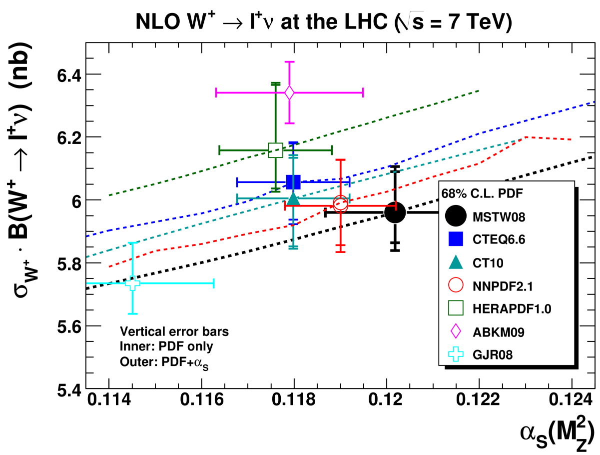
- DATASET:CTEQ, MSTW, NNPDF FIXED TARGET AND COLLIDER, eP AND $\bar{P}P$ DATA; ABKM, GJR GLOBAL DIS+ FIXET-TARGET DY; HERAPDF: HERA ONLY
- STATISTICAL TREATMENT: CTEQ, MSTW HESSIAN WITH DYNAMICAL TOLERANCE; HERAPDF, STANDARD HESSIAN+PARM. ERROR ANALYSIS; GJR, HESSIAN WITH FIXED TOLERANCE; ABKM STANDARD HESSIAN; NNPDF MONTE CARLO (ALSO STUDIED BY HERAPDF)
- PARTON PARAMETRIZATION:CTEQ, MSTW, HERAPDF $x^\alpha(1-x)^\beta\times$ POLYNOMIALS; GJR: DITTO + VALENCELIKE ASSUMPTION; NNPDF NEURAL NETS; CHEBYSHEV POLYNOMIALS STUDIED BY HERAPDF;
- α_s VALUE:CTEQ, NNPDF: EXTERNAL PARAMETER, SEVERAL VALUES AVAILABLE; MSTW: FITTED, BUT ALSO VARIABLE AS EXT.PARAMETER; ABKM, GJR: FITTED, NOT VARIABLE AS EXT. PARAMETER
- PERTURBATIVE ORDER:CTEQ: NLO, BUT DY LO WITH K-FACTORS; MSTW: NNLO, BUT DY LO WITH K-FACTORS; NNPDF FULL NLO, NNLO DIS; ABKM, GJR: FULL NNLO
- HEAVY QUARKS:CTEQ: GM-VFN (SACOT-χ SCHEME); MSTW: GM-VFN (ACOT+TR' SCHEME); NNPDF: GM-VFN (FONLL SCHEME); ABKM: FFN ($N_f = 3, 4$ MATCHED WITH BMSN SCHEME); GJR: FFN ($N_f = 3$)

THE PDF4LHC RECOMMENDATION

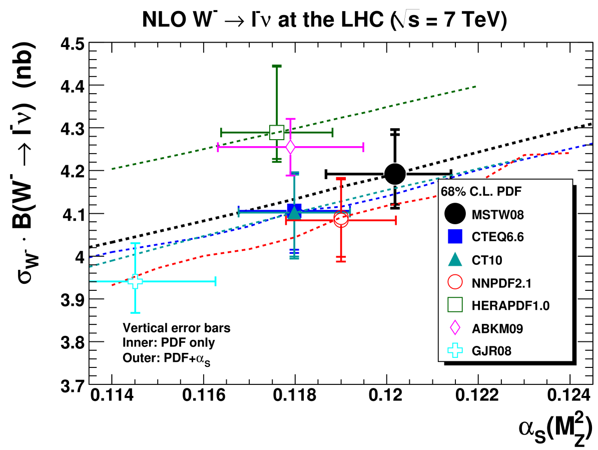
- AT NLO, ENVELOPE OF CTEQ, MSTW, NNPDF
- AT NNLO, MSTW WITH UNCERTAINTY RESCALED MY MSTWNLO/NLO ENVELOPE

WHERE DO WE STAND?
LHC STANDARD CANDLES

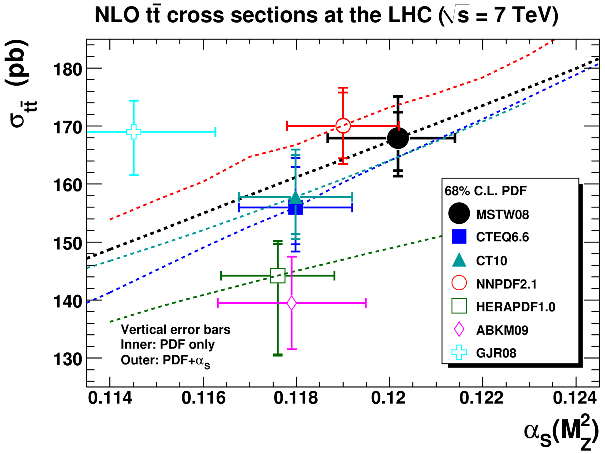
W^+



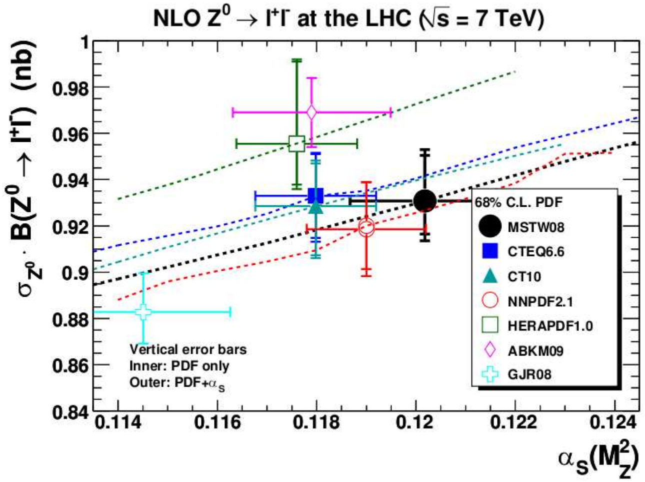
W^-



TOP



Z



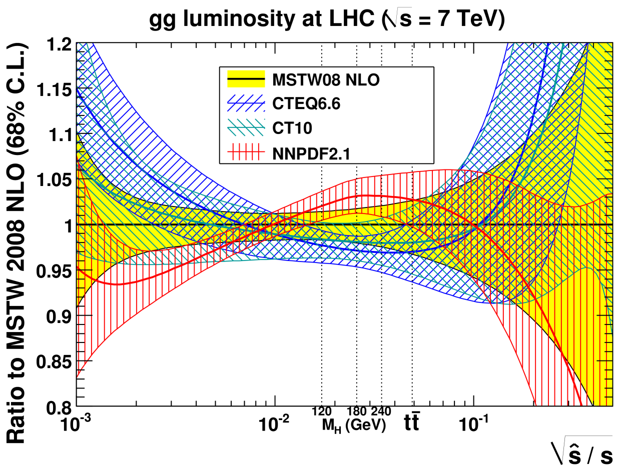
(G. Watt, 2011)

- GLOBAL FITS IN GOOD MUTUAL AGREEMENT
- CHOICE OF α_s VALUE IMPORTANT

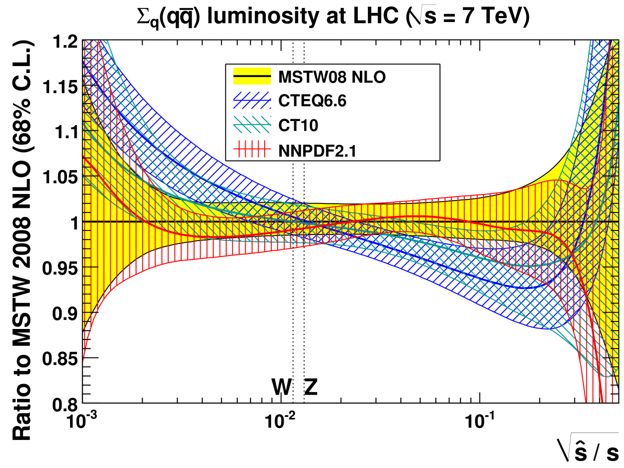
WHERE DO WE STAND?

PARTON LUMINOSITIES

GLUON-GLUON

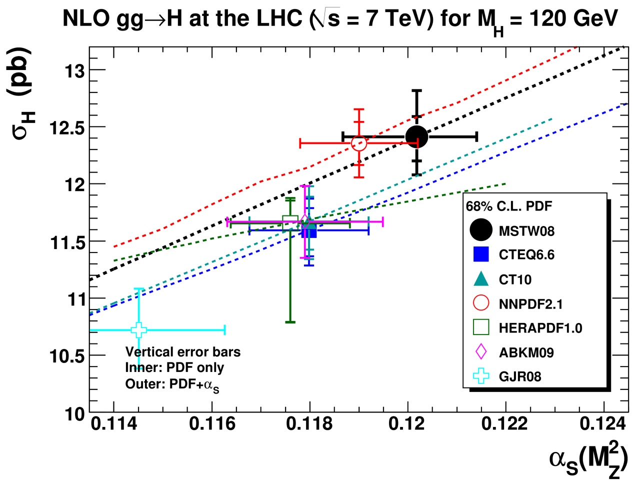


QUARK-QUARK



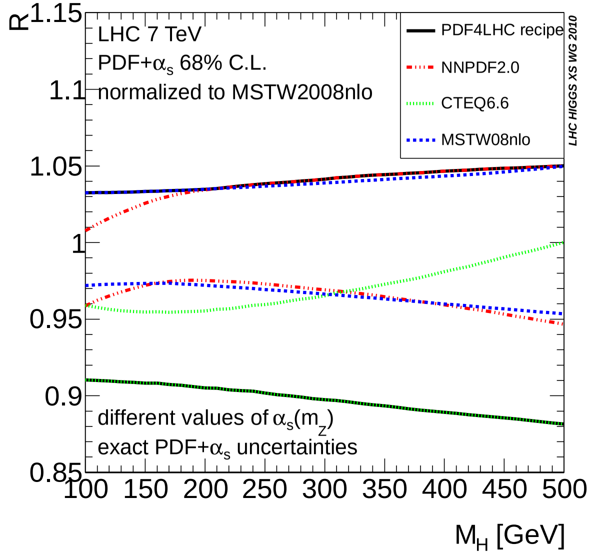
THE HIGGS CROSS SECTION

$m_H = 120$ GeV: DIFFERENT GROUPS



(G. Watt, 2011)

THE PDF4LHC RECIPE



(Higgs WG, 2011)

THE PROBLEM AND THE NNPDF SOLUTION

WHAT'S THE PROBLEM? D. Kosower, 1999

- 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) $\mathcal{P}[F_2]$

IN THE SPACE OF FUNCTIONS $F_2(x, Q^2)$

EXPECTATION VALUE OF $\mathcal{F}\left[F_2(x,Q^2)\right] \Rightarrow$ FUNCTIONAL INTEGRAL

$$\left\langle \mathcal{F}\left[F_2(x,Q^2)\right] \right\rangle = \int \mathcal{D}F_2\, \mathcal{F}\left[F_2(x,Q^2)\right]\, \mathcal{P}[F_2],$$

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

THE STANDARD (HESSIAN) SOLUTION
(MSTW, CTEQ):
FUNCTIONAL PARTON FITTING

- CHOOSE A FIXED FUNCTIONAL FORM:
  - MSTW: 20 PARMS.

    $xq(x,Q_0^2) = A(1-x)^\eta(1+\epsilon x^{0.5}+\gamma x)x^\delta$, (5 indep. fns.); $x[\bar{u}-\bar{d}](x,Q_0^2) = A(1-x)^\eta(1+\gamma x+\delta x^2)x^\delta$;

    $x[s-\bar{s}](x,Q_0^2) = A_-(1-x)^{\eta_s}x^{\delta_-}(1-x/x_0)$; $xg(x,Q_0^2) = A_g(1-x)^{\eta_g}(1+\epsilon_g x^{0.5}+\gamma_g x)x^{\delta_g}+A_{g'}(1-x)$
  - CTEQ/TEA: 22$\longrightarrow$ 26 PARMS.

    $$x\,f(x,Q_0) = a_0\,x^{a_1}\,(1-x)^{a_2}\,\exp\left(a_3x+a_4x^2+a_5\sqrt{x}+a_6x^{-a_7}\right)$$ (7 indep. functions)

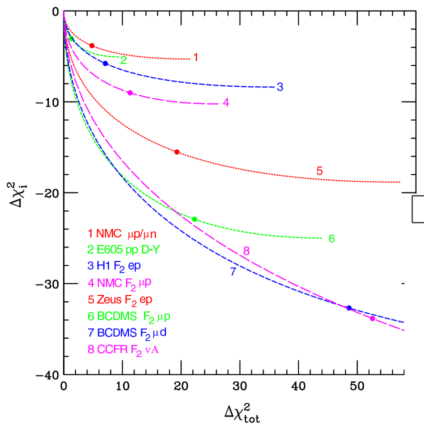
    a_6, a_7 ONLY USED FOR GLUON
  - BASIS FUNCTIONS: $u_v \equiv u-\bar{u}$, $d_v \equiv d-\bar{d}$, $\bar{u}\pm\bar{d}$ (MSTW) OR $\bar{u}$, $\bar{d}$ (CTEQ), $s^\pm \equiv s\pm\bar{s}$ (CTEQ $\bar{s}+s$ ONLY), g.
- EVOLVE TO DESIRED SCALE & COMPUTE PHYSICAL OBSERVABLES
- DETERMINE BEST-FIT VALUES OF PARAMETERS
- DETERMINE ERROR BY PROPAGATION OF ERROR ON PARMS. ('HESSIAN METHOD')
  PARM. SCANS ALSO POSSIBLE ('LAGR. MULTIPLIER METHOD')

THE TOLERANCE PROBLEM

MSTW/CTEQ: ONE–σ IS DEFINED UP TO A "TOLERANCE"

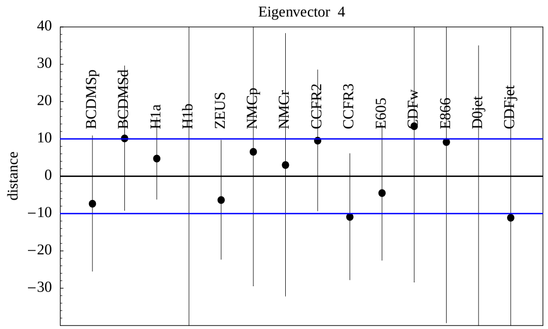
- STANDARD $\Delta\chi^2 = 1$ BANDS TOO NARROW $\Rightarrow$ SPREAD OF χ^2 FOR INDIVIDUAL EXPERIMENTS MUCH TOO LARGE!
- TOLERANCE $\Rightarrow$ ENVELOPE OF UNCERTAINTIES OF EXPERIMENTS
- DYNAMICAL $\Rightarrow$ SEPARATELY DETERMINED FOR EACH HESSIAN EIGENVECTOR

MINIMUM χ_i^2 VS GLOBAL χ^2

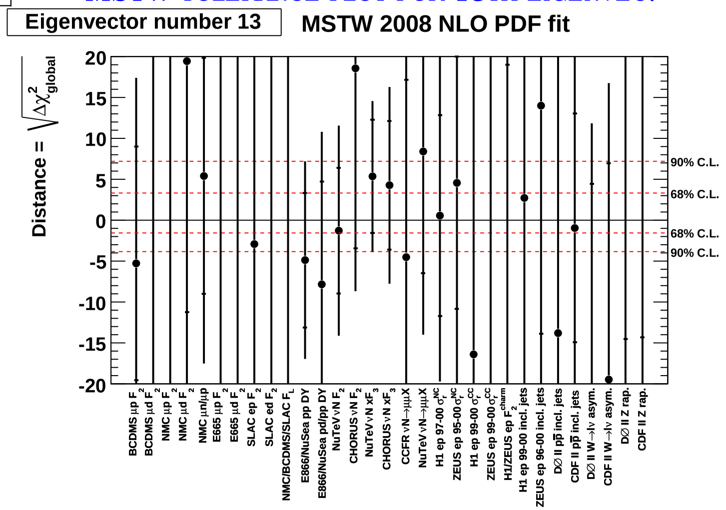


Collins, Pumplin 2001

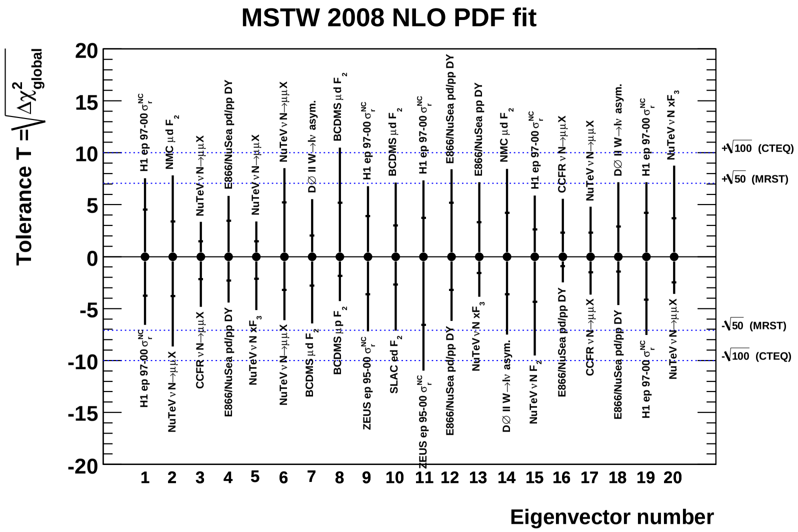
CTEQ TOLERANCE PLOT FOR 4TH EIGENVEC.



MSTW TOLERANCE PLOT FOR 13TH EIGENVEC.



GLOBAL MSTW TOLERANCE

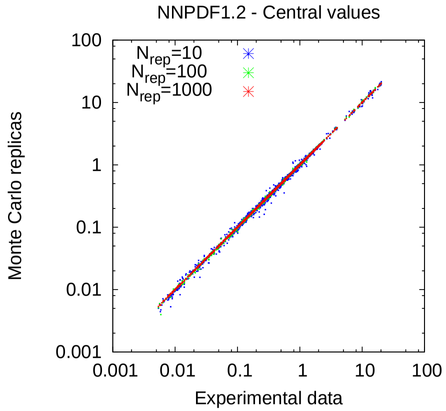


THE NNPDF APPROACH:
THE NEURAL MONTE CARLO

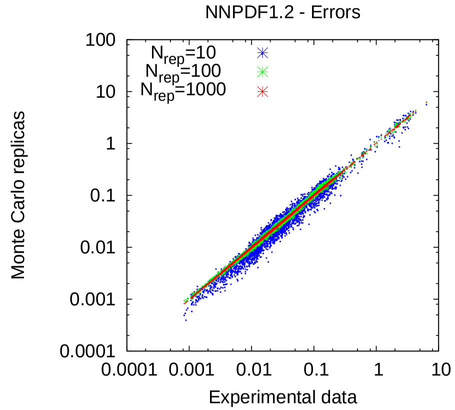
BASIC IDEA: MONTE CARLO SAMPLING OF THE PROBABILITY MEASURE IN THE (FUNCTION) SPACE OF PDFs

- START FROM MONTE CARLO SAMPLING OF DATA SPACE
- SPACE OF FUNCTIONS HUGE
  5 BINS FOR 10 PTS$\times$ 7 FCTNS $\to 5^{70} \sim 10^{49}$ BINS
- IMPORTANCE SAMPLING: DATA TELL US WHICH BINS ARE POPULATED

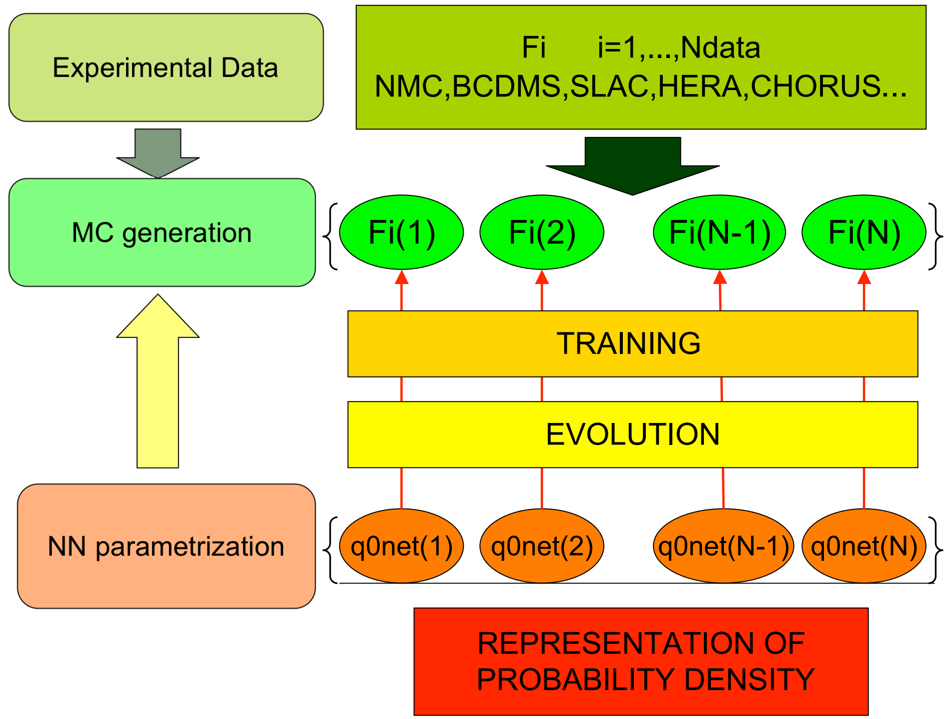
replica averages vs. central values



replica standard dev. vs. uncertainties

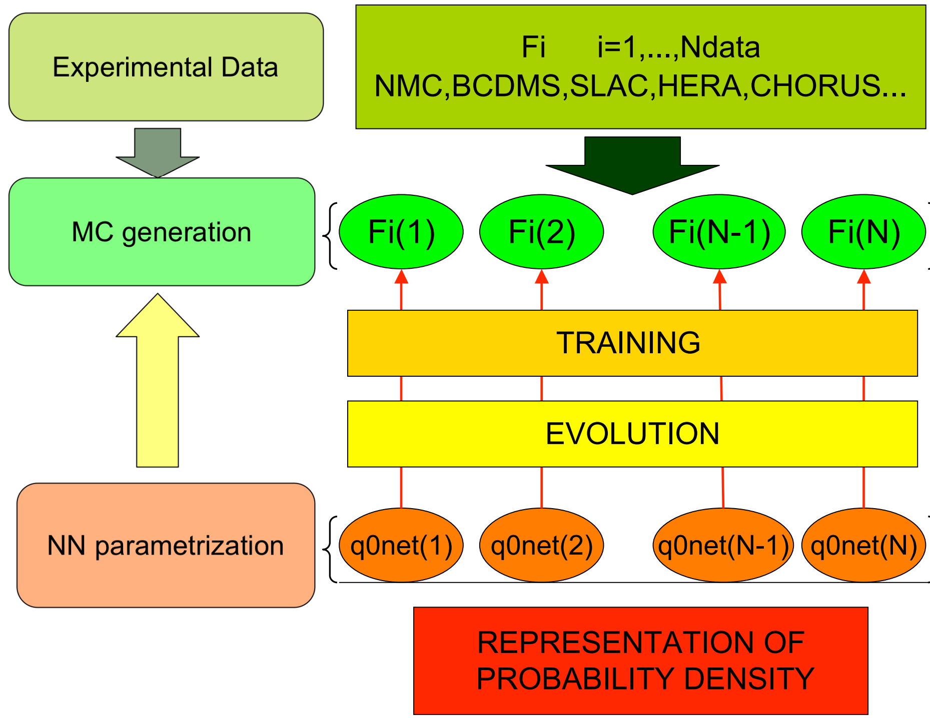


10 REPLICAS ENOUGH FOR CENTRAL VALS, 100 FOR UNCERTAINTIES, 1000 FOR CORRELNS



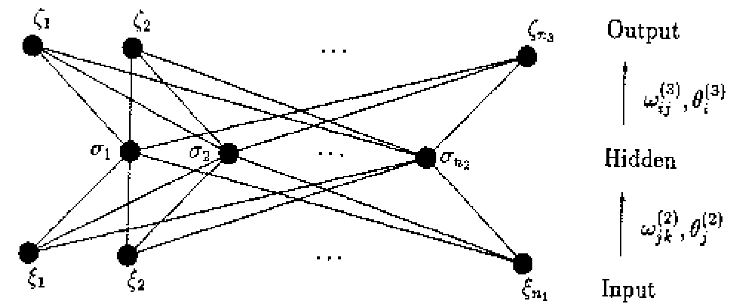
DATA MONTE CARLO ⇒ PDF MONTE CARLO
NEURAL NETWORK PARM+ CROSS-VALIDATION METHOD

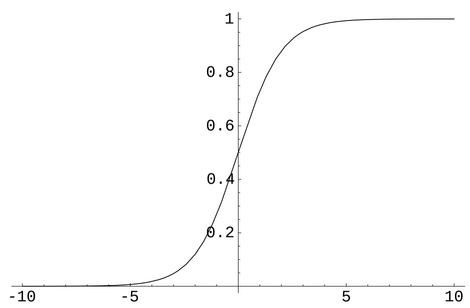
- EACH PDF↔ NEURAL NETWORK PARAMETRIZED BY 37 PARAMETERS
- NNPDF1.X, NNPDF2.X: $37 \times 7 = 259$ PARMS (RECALL MSTW, CTEQ → 20 FREE PARAMETERS) "INFINITE" NUMBER OF PARAMETERS⇒ CAN REPRESENT ANY FUNCTION
- COMPLEX SHAPES (LARGE NO.OF PARAMETERS) REQUIRE LONGER FITTING
- FIT STOPS WHEN QUALITY OF FIT TO RANDOMLY SELECTED "VALIDATION" DATA (NOT FITTED) STOPS IMPROVING
- CAN OBTAIN A FIT WITH χ^2 LOWER THAN BEST FIT ("OVERLEARNING")



NEURAL NETWORKS

A NONLINEAR FUNCTIONAL FORM





MULTILAYER FEED-FORWARD NETWORKS

- Each neuron receives input from neurons in preceding layer and feeds output to neurons in subsequent layer
- Activation determined by weights and thresholds

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

- Sigmoid activation function

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

EXAMPLE: A 1-2-1 NN

$$f(x) = \frac{1}{1+e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - x\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - x\omega_{21}^{(1)}}}}}$$

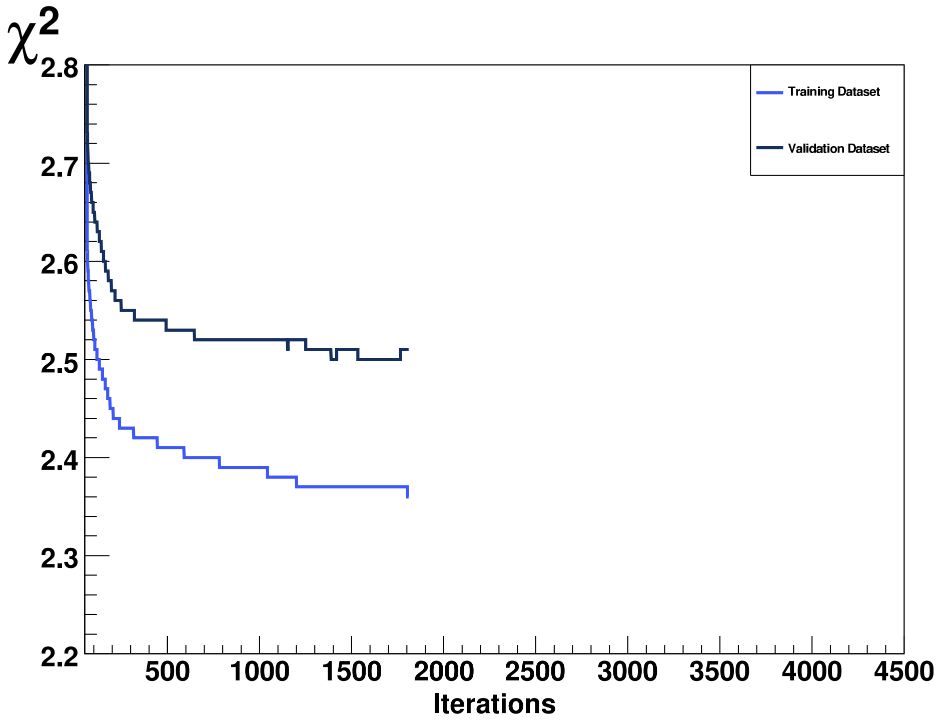
- ANY FUNCTION CAN BE REPRESENTED BY A SUFFICIENTLY BIG NEURAL NETWORK $\Rightarrow$ CAN CHECK INDEPENDECE OF SIZE
- MINIMZATION PERFORMED BY GENETIC ALGORITHM
- INFORMATION CONTAINED ("COARSENESS") IN NN INCREASES DURING MINIMIZATION

DATA MONTE CARLO ⇒ PDF MONTE CARLO
CROSS-VALIDATION

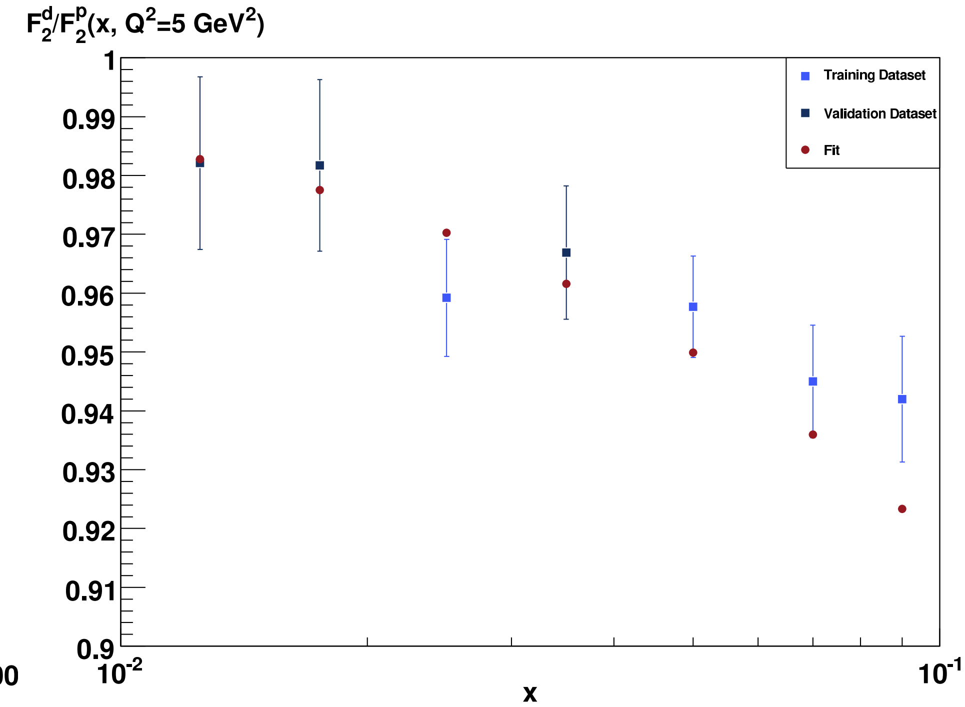
- REPLICAS ARE FITTED TO A DATA SUBSET
- A DIFFERENT SUBSET OF DATA USE FOR EACH REPLICA
- OPTIMAL FIT WHEN FIT TO VALIDATION (CONTROL) DATA STOPS IMPROVING
-

OPTIMAL FITTING

χ^2



FIT TO DATA

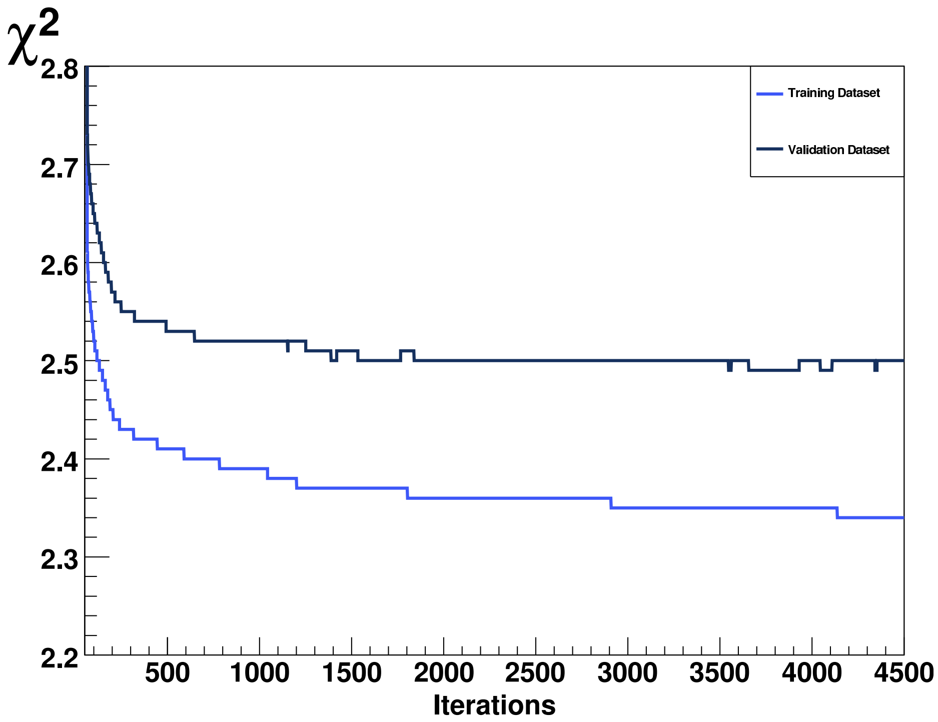


DATA MONTE CARLO $\Rightarrow$ PDF MONTE CARLO
CROSS-VALIDATION

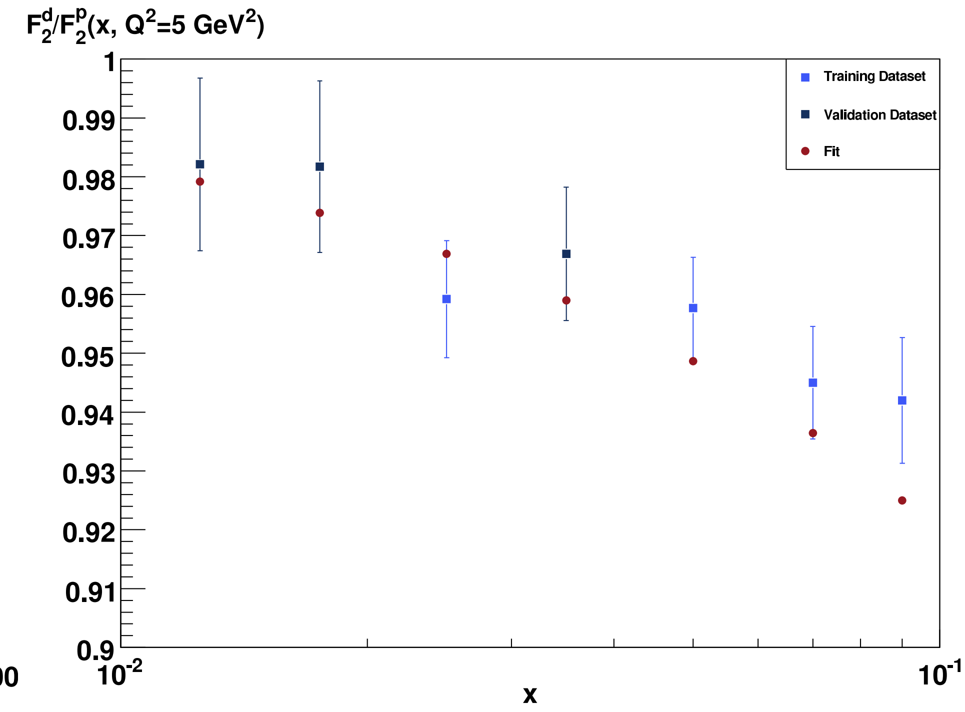
- REPLICAS ARE FITTED TO A DATA SUBSET
- A DIFFERENT SUBSET OF DATA USE FOR EACH REPLICA
- OPTIMAL FIT WHEN FIT TO VALIDATION (CONTROL) DATA STOPS IMPROVING
- THE BEST FIT IS NOT AT THE MINIMUM OF THE χ^2

OVERFITTING

χ^2



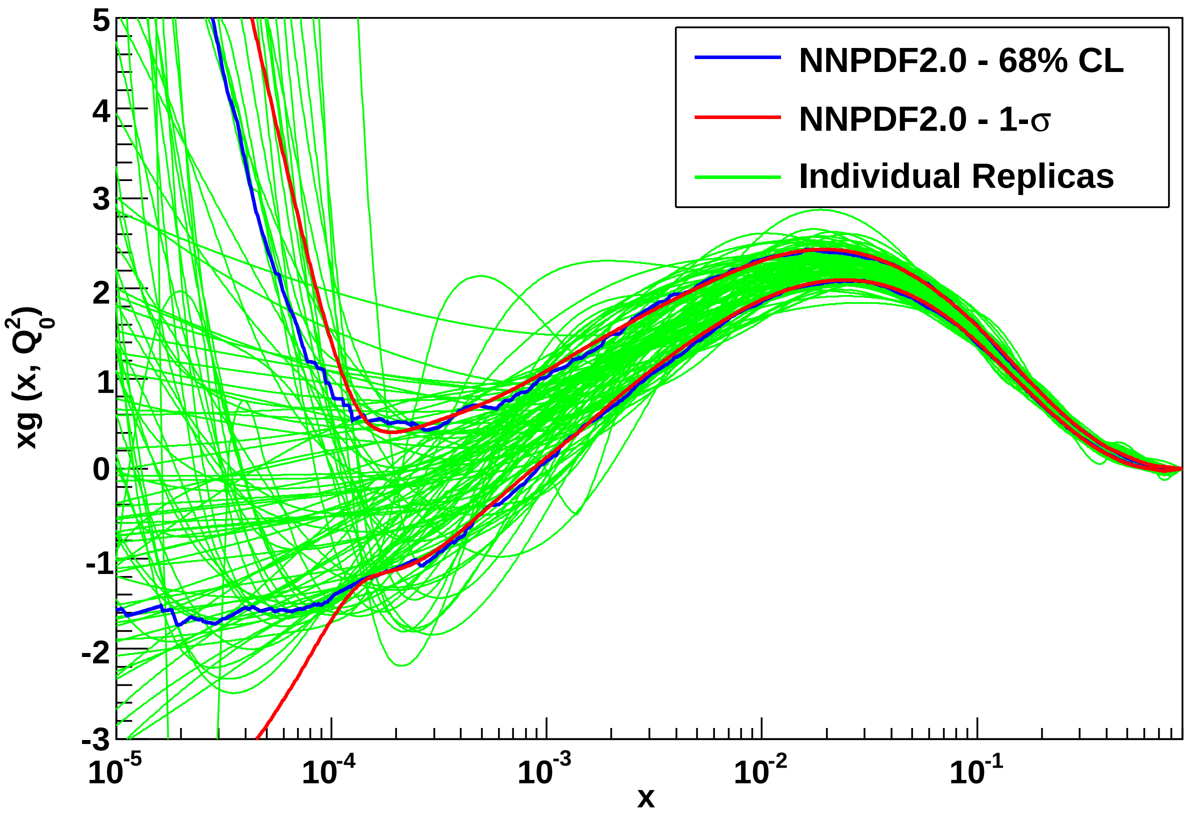
FIT TO DATA



LIKELIHOOD CONTOURS AND UNCERTAINTIES

NNPDF: ONE σ VS. CENTRAL 68% FOR THE MC DISTRIBUTION OF PDFS

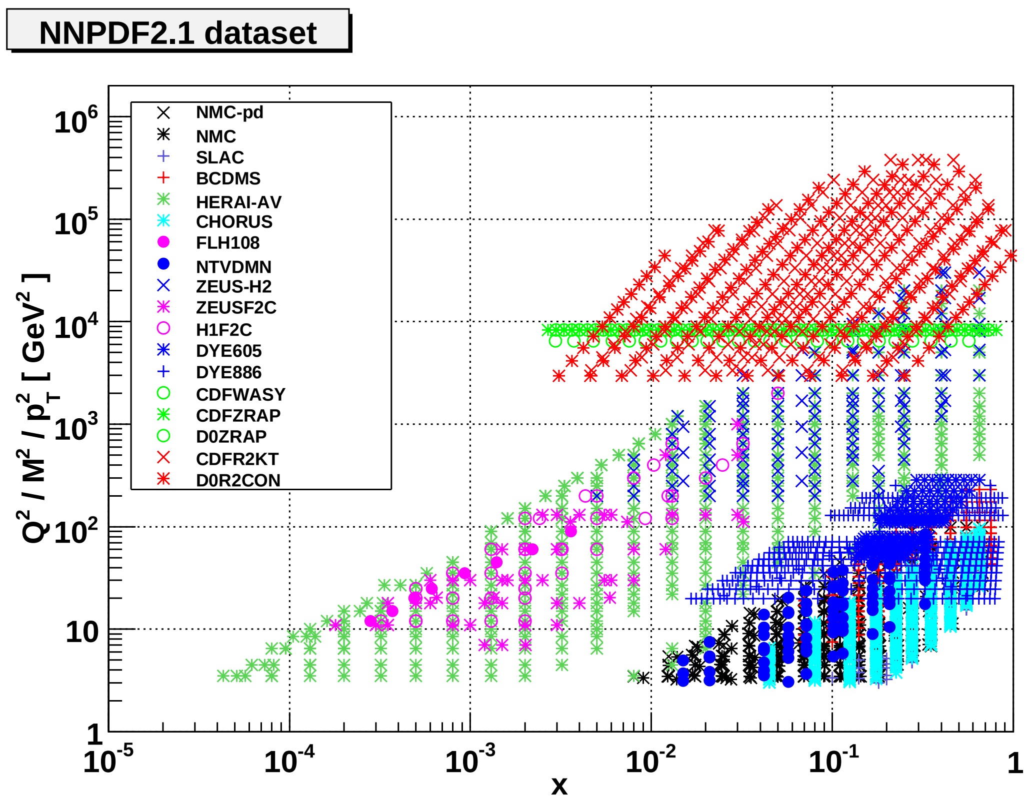
Example: the gluon distribution in the NNPDF2.0 set

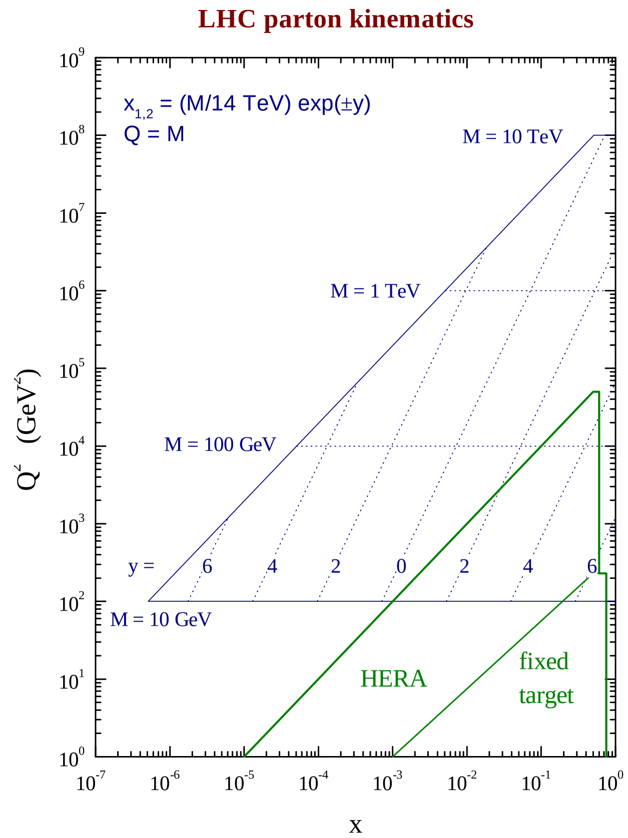


- ENSEMBLE OF REPLICAS $\leftrightarrow$ PROBABILITY DISTRIBUTION OF PDFS
- EXPECTED CENTRAL VALUE $\leftrightarrow$ MEAN; UNCERTAINTY $\leftrightarrow$ STANDARD DEVIATION
- ANY FEATURES OF DISTRIBUTION CAN BE DETERMINED (C.L., CORRELATIONS...)
- DISTRIBUTION NEED NOT BE GAUSSIAN $\rightarrow$ STANDARD DEVIATION $\neq$ 68% C.L. (GLUON $\Leftrightarrow$ STRUCTURE FUNCTION POSITIVITY CONSTRAINTS)

RESULTS

THE NNPDF2.1 PDF DETERMINATION

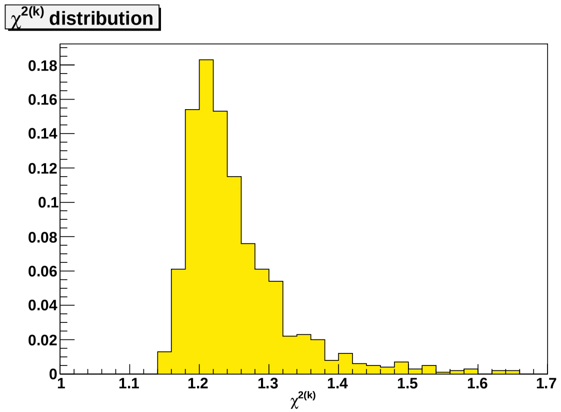




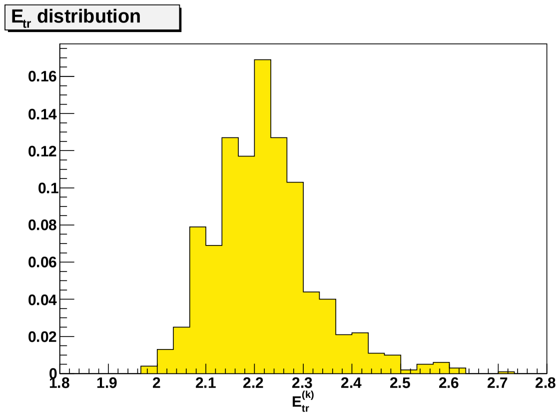
- GLOBAL PDF FIT, INCLUDES
  - DIS: NEUTRAL AND CHARGED CURRENT, CHARGED LEPTON AND NEUTRINO BEAMS, INCLUSIVE AND CHARM-TAGGED
  - DRELL-YAN: FIXED TARGET AND COLLIDER, NEUTRAL γ^* AND Z AND CHARGED CURRENT W PRODUCTION
  - INCLUSIVE JETS
- NLO QCD, HQ MASSES INCLUDED TO $O(\alpha_s)$
- 7 PDFS PARAMETRIZED INDEPENDENTLY, HQ GENERATED DYNAMICALLY
- SETS WITH DIFFERENT VALUES OF α_s, m_c, m_b AVAILABLE

STATISTICAL FEATURES

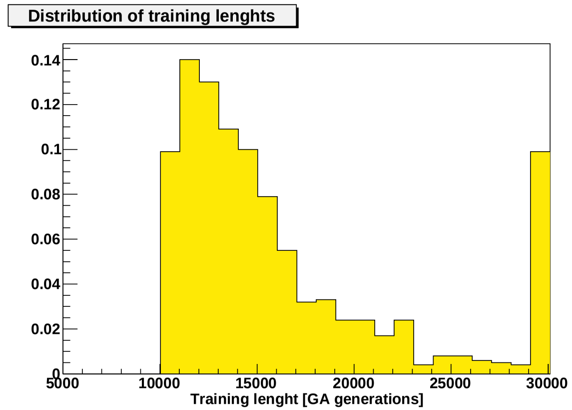
χ^2 TO DATA



χ^2 TO REPLICA



TRAINING LENGTHS



- χ^2 TO REPLICA PEAKED AROUND 2,
  χ^2 TO DATA PEAKED AROUND 1
- χ^2 OF AVERAGE SMALLER THAN AVERAGE OF χ^2
- AVERAGE UNCERTAINTY OF PREDICTION
  SMALLER THAN AVERAGE UNCERTAINTY ON DATA

$\Rightarrow$ FIT "LEARNS" UNDERLYING LAW

χ^2_{tot}	1.16
$\langle E \rangle \pm \sigma_E$	2.24 ± 0.09
$\langle E_{\mathrm{tr}} \rangle \pm \sigma_{E_{\mathrm{tr}}}$	2.22 ± 0.11
$\left\langle E_{\mathrm{val}} \right\rangle \pm \sigma_{E_{\mathrm{val}}}$	2.28 ± 0.12
$\langle \mathrm{TL} \rangle \pm \sigma_{\mathrm{TL}}$	$(1.6 \pm 0.6)\ 10^4$
$\left\langle \chi^{2(k)} \right\rangle \pm \sigma_{\chi^2}$	1.25 ± 0.09
$\left\langle \sigma^{(\mathrm{exp})} \right\rangle_{\mathrm{dat}}$ (%)	11.3%
$\left\langle \sigma^{(\mathrm{net})} \right\rangle_{\mathrm{dat}}$ (%)	4.4%
$\left\langle \rho^{(\mathrm{exp})} \right\rangle_{\mathrm{dat}}$	0.18
$\left\langle \rho^{(\mathrm{net})} \right\rangle_{\mathrm{dat}}$	0.56

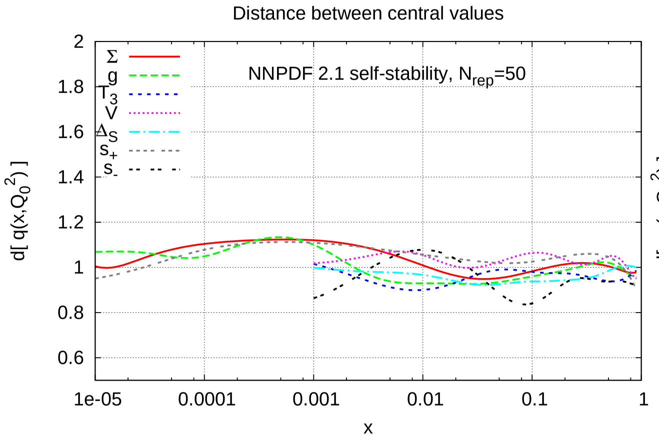
SCALING & STABILITY

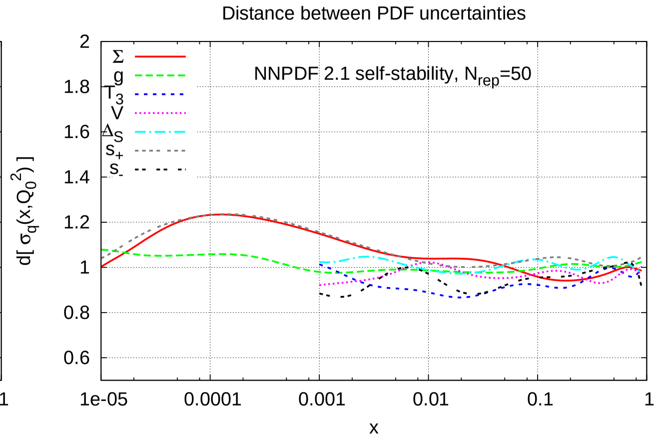
- COMPARE RESULTS BETWEEN DIFFERENT SETS OF REPLICAS $\Rightarrow$ STATISTICALLY EQUIVALENT

DISTANCE

$$d^2\left(\langle q^{(1)}\rangle, \langle q^{(2)}\rangle\right) = \frac{\left(\langle q^{(1)}\rangle_{(1)} - \langle q^{(2)}\rangle_{(2)}\right)^2}{\sigma^2_{(1)}[\langle q^{(1)}\rangle] + \sigma^2_{(2)}[\langle q^{(2)}\rangle]} \quad \text{WITH } \sigma^2_{(i)}[\langle q^{(i)}\rangle] = \frac{1}{N^{(i)}_{\rm rep}}\sigma^2_{(i)}[q^{(i)}]$$

& similarly for uncertainties





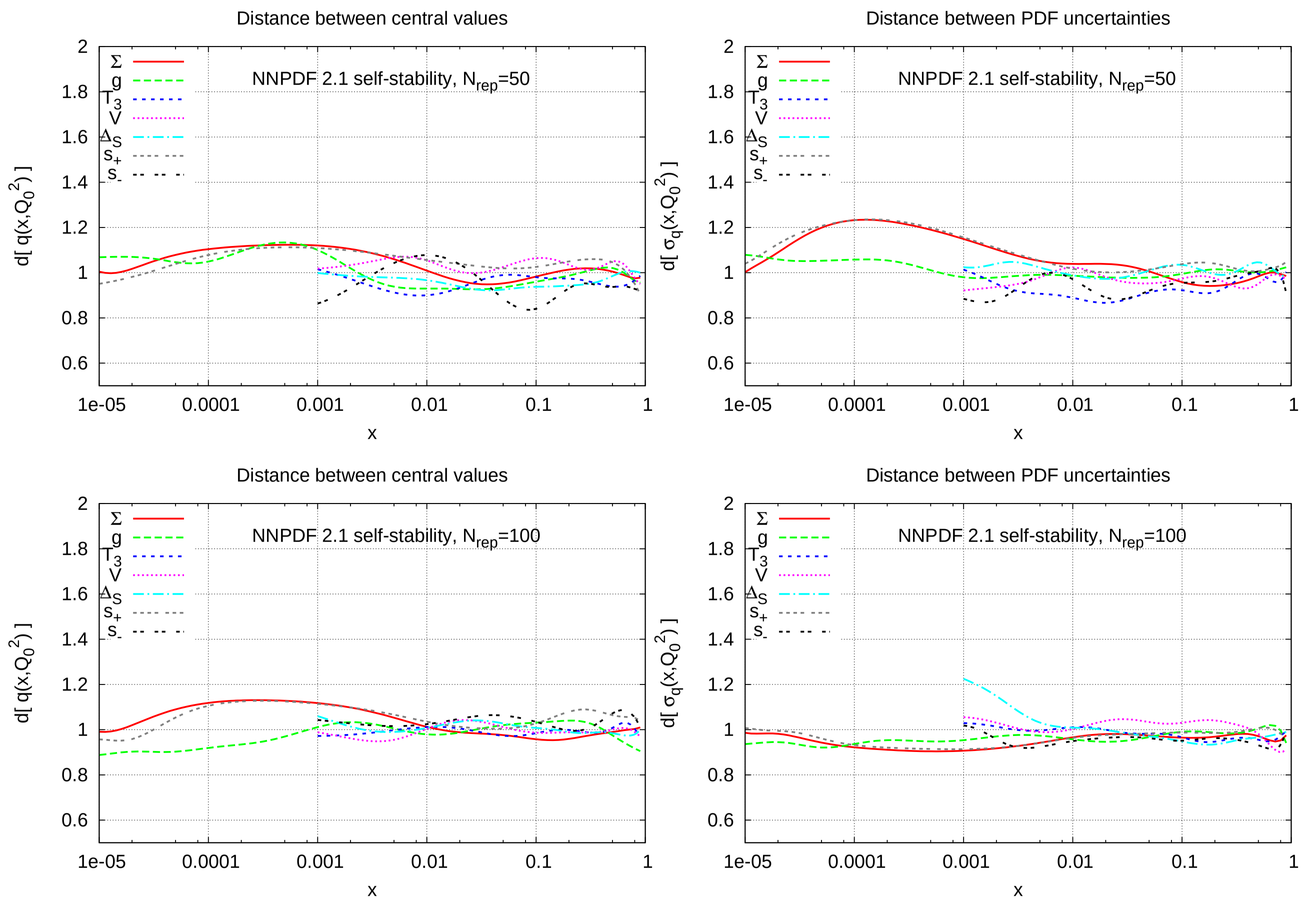
SCALING & STABILITY

- COMPARE RESULTS BETWEEN DIFFERENT SETS OF REPLICAS $\Rightarrow$ STATISTICALLY EQUIVALENT
- REPEAT COMPARISON FOR DIFFERENT $N_{\rm rep}$ $\Rightarrow$ FLUCTUATIONS SCALE WITH $N_{\rm rep}$!

DISTANCE

$$d^2\left(\langle q^{(1)}\rangle,\langle q^{(2)}\rangle\right)=\frac{\left(\langle q^{(1)}\rangle_{(1)}-\langle q^{(2)}\rangle_{(2)}\right)^2}{\sigma^2_{(1)}[\langle q^{(1)}\rangle]+\sigma^2_{(2)}[\langle q^{(2)}\rangle]}\quad \text{WITH } \sigma^2_{(i)}[\langle q^{(i)}\rangle]=\frac{1}{N^{(i)}_{\rm rep}}\sigma^2_{(i)}[q^{(i)}]$$

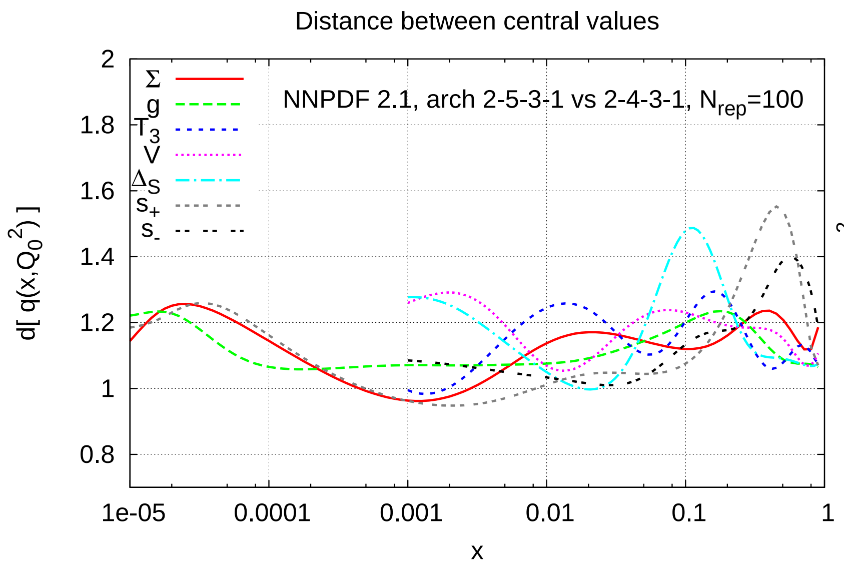
& similarly for uncertainties

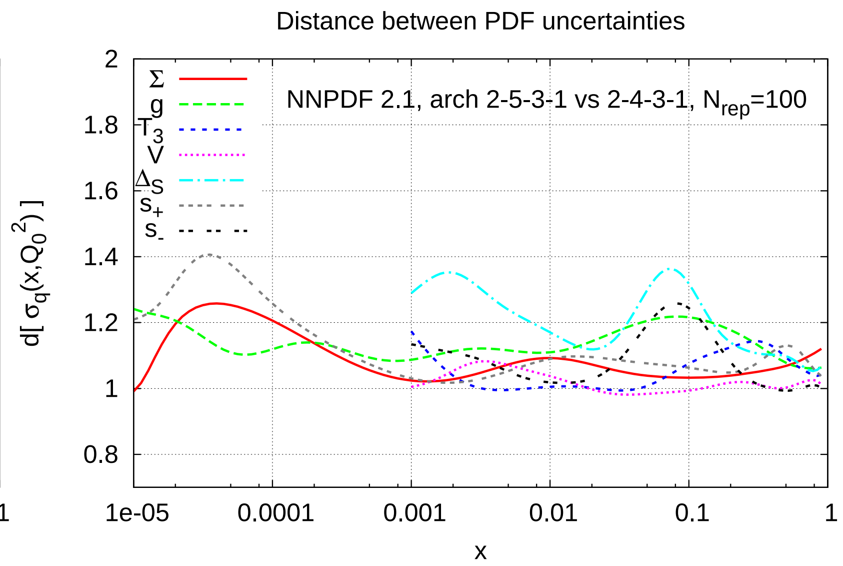


PARAMETRIZATION INDEPENDENCE

COMPARE RESULTS OBTAINED WITH DIFFERENT ARCHITECTURE

OF NEURAL NETWORK: 2-4-3-1 VS 2-5-3-1 (31 PARAMETERS VS 37)





STATISTICALLY EQUIVALENT!

DESPITE USING $6 \times 7 = 42$ LESS PARAMETERS

DO PDF UNCERTAINTIES HAVE A STATISTICAL MEANING?

NEW DATA $\Rightarrow$ BAYES' THEOREM

$$\langle\mathcal{O}\rangle_{\rm new} = \int \mathcal{O}[f]\,\mathcal{P}_{\rm new}(f)\,Df, = \mathcal{N}_\chi \int \mathcal{O}[f]\,\mathcal{P}(\chi^2|f)\mathcal{P}_{\rm old}(f)\,Df,$$

IN A MONTE CARLO APPROACH...

$$\langle\mathcal{O}\rangle_{\rm new} = \frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}} \mathcal{N}_\chi \mathcal{P}(\chi^2|f_k)\mathcal{O}[f_k] = \frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}} w_k\,\mathcal{O}[f_k], \quad w_k = \mathcal{N}(\chi^2_k)^{n/2-1}e^{-\frac{1}{2}\chi^2_k}$$

$\Rightarrow$ EFFECT OF NEW DATA IS ACCOUNTED FOR BY

REWEIGHTING MONTE CARLO AVERAGES

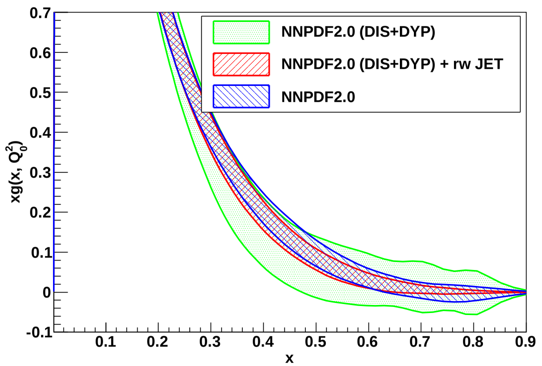
- DETERMINE PDFs INCLUDING SOME DATA BY BAYES' THEOREM (REWEIGHTING)
- DETERMINE PDFs BY ENLARGING THE DATASET TO THE NEW DATA (REFITTING)
- COMPARE RESULTS $\Rightarrow$ STRONG CONSISTENCY CHECK

DO PDF UNCERTAINTIES HAVE A STATISTICAL MEANING?

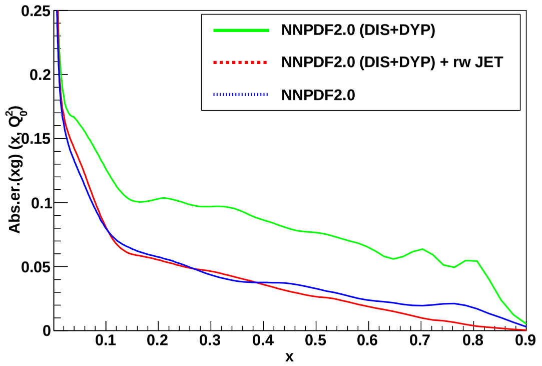
INCLUSION OF JET DATA: REWEIGHTING VS. REFITTING

NNPDF2.0DIS+DY VS. NNPDF2.0FULL

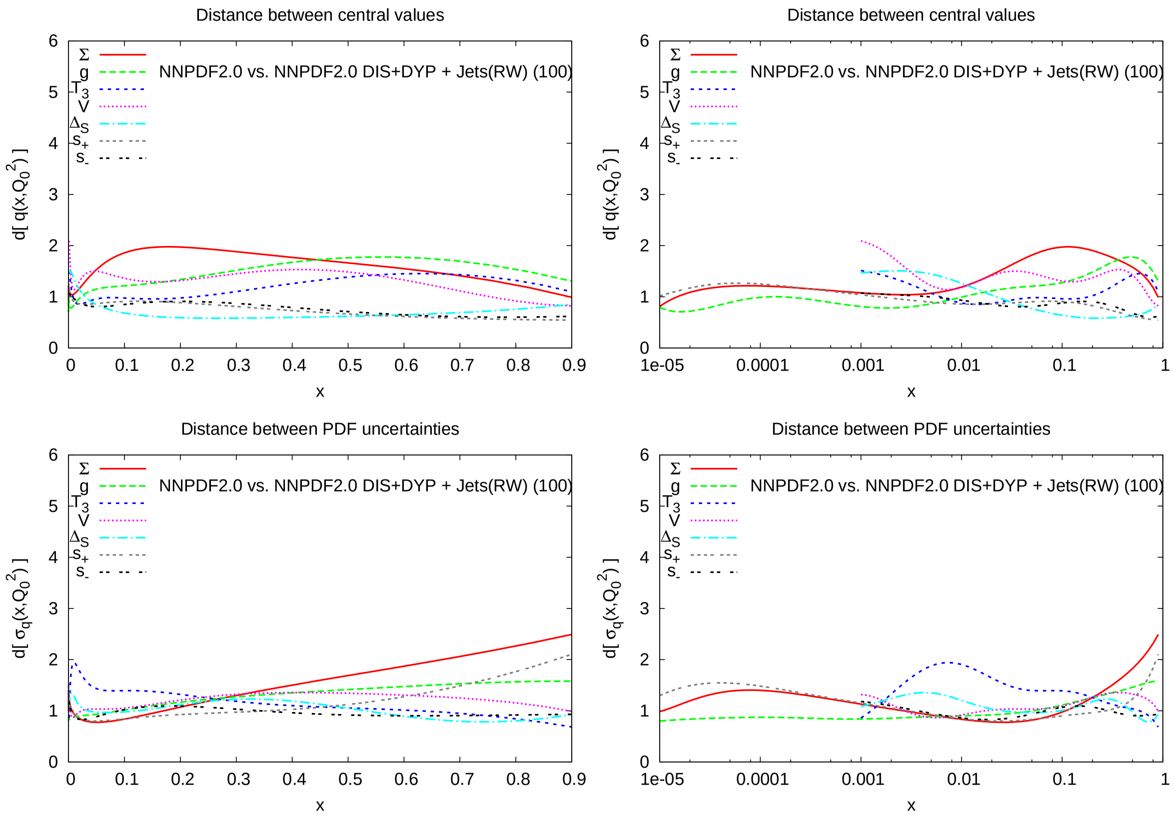
GLUON



GLUON UNCERTAINTY



DISTANCES



YES!

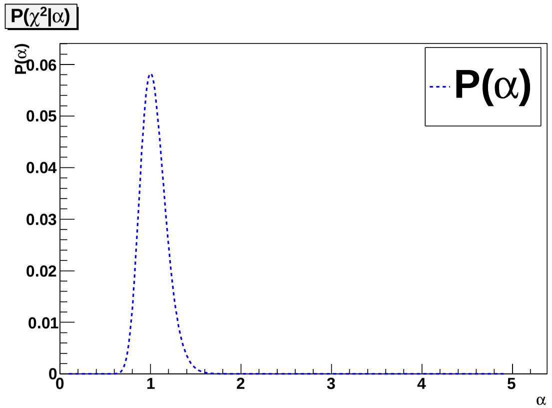
DATA COMPATIBILITY

- INCONSISTENT DATA $\Leftrightarrow$ UNDERESTIMATED UNCERTAINTIES

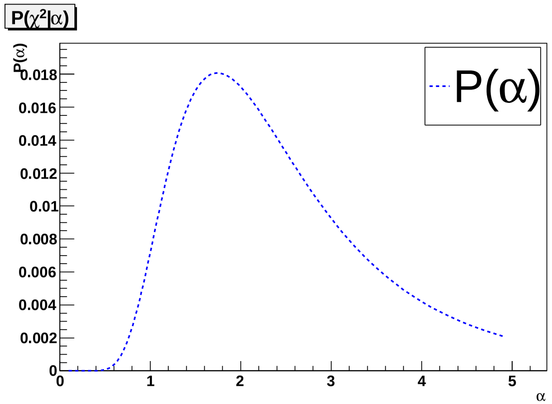
- RESCALE ALL UNCERTAINTIES IN A GIVEN EXPERIMENT BY SOME FACTOR α: $\chi^2_\alpha = \chi^2/\alpha$ (TOLERANCE)

- DETERMINE PROBABILITY DISTRIBUTION OF α VALUES BY BAYES' THEOREM $\Rightarrow$ REWEIGHTING: $\mathcal{P}(\alpha) = \frac{\mathcal{N}}{\alpha} \sum_{k=1}^{N} w_k w_k(\alpha)$.

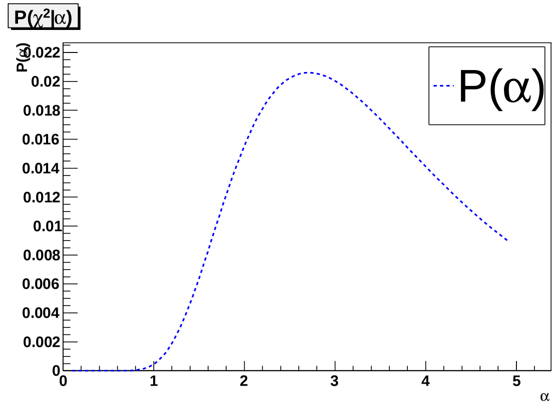
JETS



D0 e INCLUSIVE



D0 e HIGH E_T



- JETS: $\Rightarrow$ CONSISTENT DATA

- $W^\pm$ CHARGE ASYMMETRIES, D0 INCLUSIVE e DATA $\Rightarrow$ UNCERTAINTIES UNDERESTIMATED BY $\sim 30\%$ (PROB. PEAKS AT $\alpha \sim 1.7$)

- $W^\pm$ CHARGE ASYMMETRIES, D0 e DATA WITH $E_T > 35$ GEV $\Rightarrow$ INCONSISTENT DATA

PRECISION PHYSICS: DETERMINATION OF PHYSICAL PARAMETERS

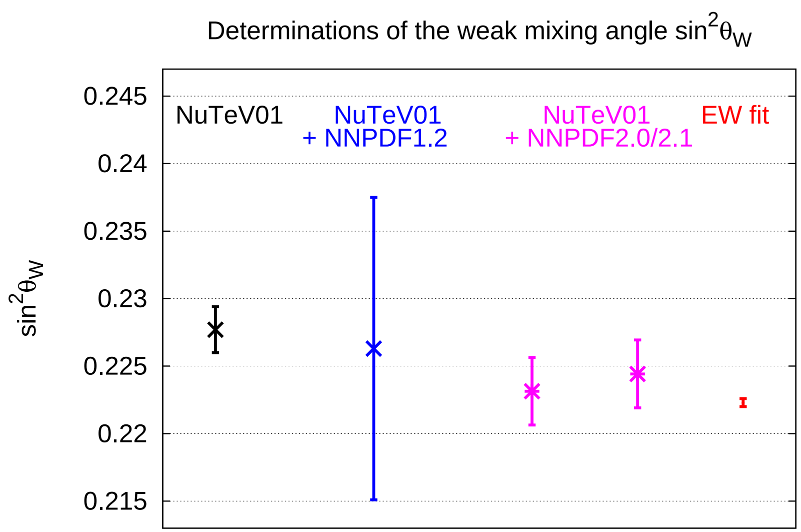
A DISCOVERY THAT WASN'T
THE NUTEV ANOMALY...

$$
\begin{aligned}
R_{\rm PW} &\equiv \frac{\sigma(\nu\mathcal{N}\to\nu X)-\sigma(\bar\nu\mathcal{N}\to\bar\nu X)}{\sigma(\nu\mathcal{N}\to\ell X)-\sigma(\bar\nu\mathcal{N}\to\bar\ell X)}\\
&= \frac{1}{2}-\sin^2\theta_{\rm W}+\left(\frac{(U^- - D^-)+(C^- - S^-)}{\mathcal{Q}^-}\frac{1}{6}\left(3-7\sin^2\theta_{\rm W}\right)\right),
\end{aligned}
$$

- PASCHOS-WOLFENSTEIN RATIO CAN BE MEASURED IN NEUTRINO DIS
- RESULT DEPENDS ON EW MIXING ANGLE, VALENCE ISOSPIN BREAKING (WITH ISOSINGLET TARGET), STRANGENESS VALENCE MOMENTUM ASYMMETRY
- STRANGENESS VALENCE MOMENTUM ASSUMED BY NUTEV TO VANISH $\Rightarrow$ THREE σ DISCREPANCY WITH GLOBAL FIT

...IS GONE

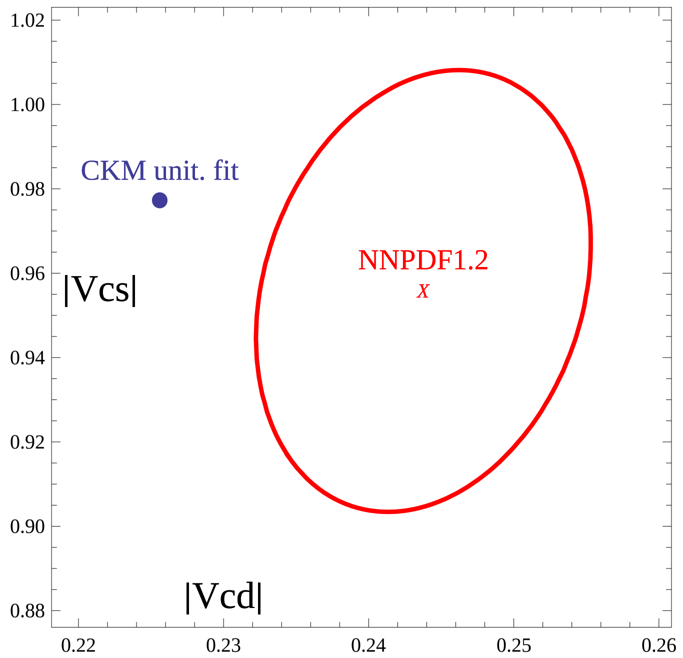
- NNPDF: s, $\bar s$ LIKE ANY OTHER PDF $\to$ 37 PARMS. EACH (CTEQ6.6: $s=\bar s$, TWO PARMS; MSTW08, S4S4, $\bar s$ TWO PARMS EACH)
- IF STRANGENESS UNCERTAINTY KEPT INTO ACCOUNT (DIS ONLY FIT: NNPDF1.2) $\Rightarrow$ EFFECT LOSES STAT. SIGNIFICANCE
- IF HADRONIC DATA INCLUDED (NNPDF2.0/NNPDF2.1 GLOBAL FITS) $\Rightarrow$ STRANGENESS ASYMMETRY DETERMINED QUITE ACCURATELY $\to$ CORRECTED RESULT IN IMPRESSIVE AGREEMENT WITH SM GLOBAL FIT



(NNPDF2.1, 2011)

CKM MATRIX ELEMENTS (NNPDF1.2: 2009)

CAN BE DETERMINED WITH SURPRIZING ACCURACY



Analysis	Determination	$\|V_{cs}\|$
NNPDF1.2	Direct from global PDF analysis	$0.96 \pm 0.07^{\rm tot}$
CDHS	LO from $\nu N \to \mu^+\mu^- X$	≥ 0.59 (90% C.L.)
CCFR	NLO from $\nu N \to \mu^+\mu^- X$	≥ 0.74 (90% C.L.)
PDG08	Average from D decays	1.04 ± 0.06
Hocker	Average from $\nu N \to \mu^+\mu^- X$	1.04 ± 0.16
DELPHI	Direct from $W^+ \to c\bar{s}$ decays	$0.94^{+0.32}_{-0.26} \pm 0.13$
PDG08	CKM unitarity fit	0.97334 ± 0.00023

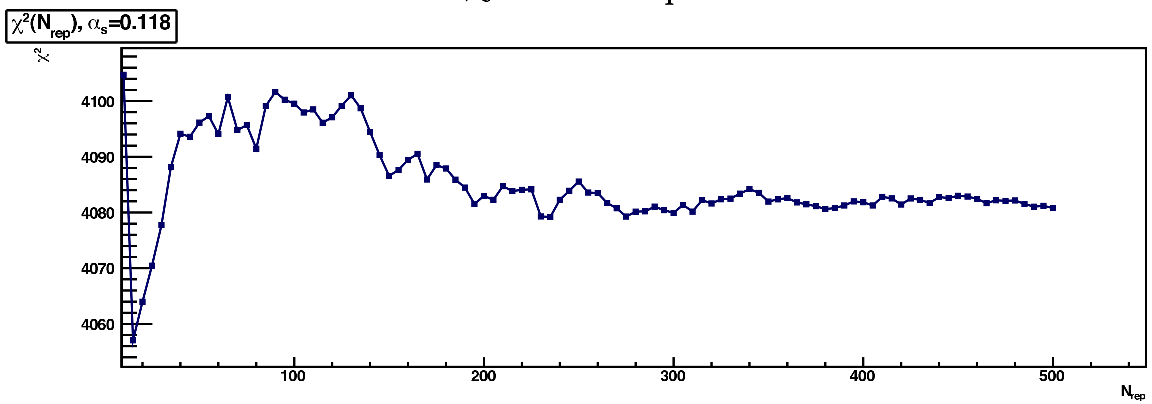
Analysis	Determination	$\|V_{cd}\|$
NNPDF1.2	Direct from global PDF analysis	$0.244 \pm 0.019^{\rm tot}$
CDHS	LO from $\nu N \to \mu^+\mu^- X$	0.24 ± 0.03
CCFR	NLO from $\nu N \to \mu^+\mu^- X$	$0.232^{+0.017}_{-0.019}$
PDG08	Average from $\nu N \to \mu^+\mu^- X$	0.230 ± 0.011
PDG08	Average from $D \to K/\pi l \nu$ decays	0.218 ± 0.023
PDG08	CKM unitarity fit	0.2256 ± 0.0010

- FIRST DETERMINATION (2009) OF CKM MATRIX ELEMENTS FROM DIS
- NNPDF1.2: ONLY DIS DATA
- MORE ACCURATE (V_{cs}) OR COMPETITIVE (V_{cd}) THAN OTHER DIRECT DETERMINATIONS

DETERMINING α_s FROM NNPDF2.1

- **ADVANTAGE**: CAN CHECK STATISTICAL FEATURES OF RESULTS: DEPENDENCE OF UNCERTAINTIES ON SIZE OF MC SAMPLE $\Rightarrow$ ERROR AND ERROR ON THE ERROR
- **DISADVANTAGE** χ^2 IS A RANDOM VARIABLE $\Rightarrow$ FLUCTUATES FOR FINITE SAMPLE SIZE $\Delta\chi^2 \sim \frac{\sqrt{N_{\rm dat}}}{N_{\rm rep}}$ $\Rightarrow$ ADDITIONAL FINITE-SIZE UNCERTAINTY

χ^2 VS $N_{\rm rep}$



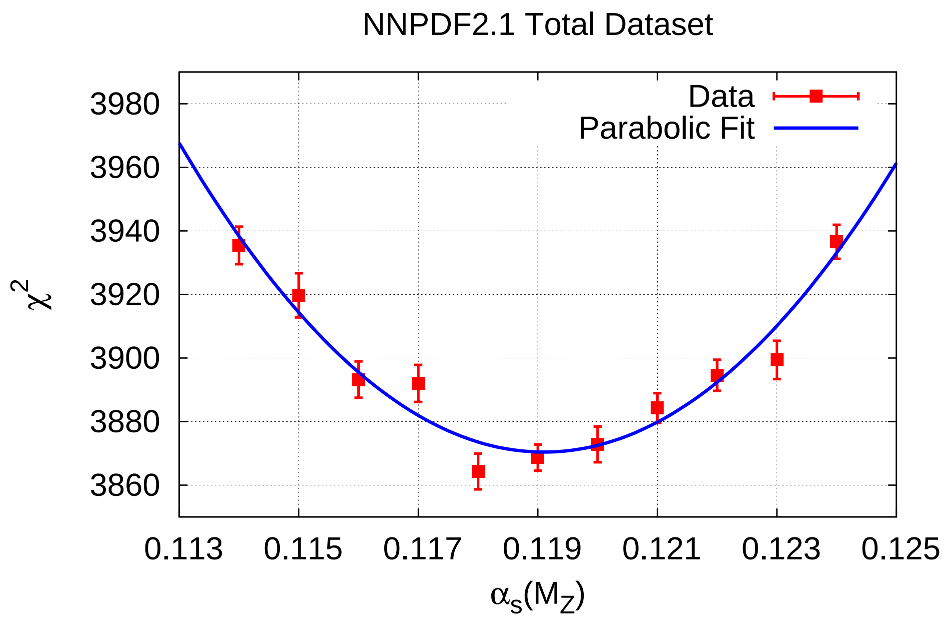
THE PROCEDURE

- PRODUCE $N_{\rm rep}$ REPLICAS FOR A RANGE OF VALUES OF α_s & DETERMINED THE χ^2
- DETERMINE THE STATISTICAL FINITE-SIZE UNCERTAINTY ON EACH χ^2 VALUE
- PERFORM A PARABOLIC FIT TO χ^2 PROFILE, POSSIBLY DISCARD OUTER NON-PARABOLIC POINTS BASED ON FIT QUALITY
- DETERMINE α_s AND STATISTICAL UNCERTAINTY BY $\Delta\chi^2 = 1$ ABOUT THE MINIMUM
- DETERMINE FURTHER PROCEDURAL UNCERTAINTY DUE TO UNCERTAINTY ON BEST-FIT PARMS $\Rightarrow$ REPEAT FOR LARGER $N_{\rm rep}$ UNTIL PROCEDURAL UNCERTAINTY NEGLIGIBLE COMPARED TO STATISTICAL

THE RESULT: GLOBAL FIT

NLO NNPDF2.1 GLOBAL DETERMINATION (ONLY STAT. ERROR KNOWN)
$N_{\mathrm{rep}} = 500$ PER VALUE OF α_s

$$\alpha_s(M_z) = 0.1191 \pm 0.0006(\mathrm{stat.}) \pm 0.0001(\mathrm{proc.})$$



$\chi^2/\mathrm{d.o.f.} = 1.6$ for the parabolic fit

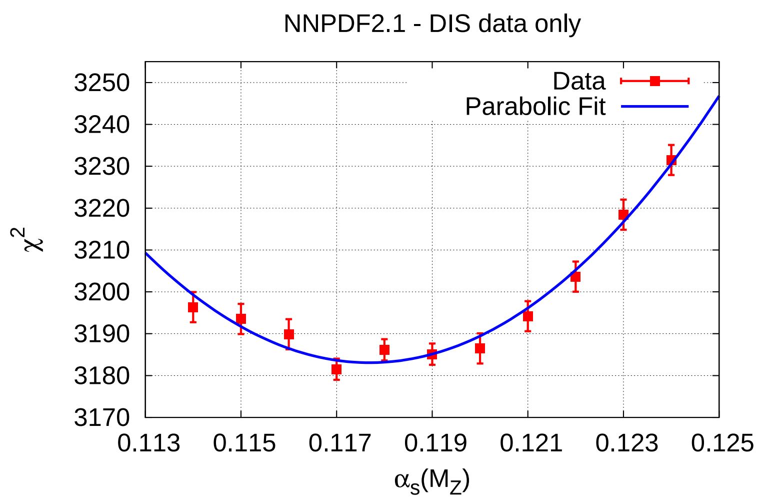
THE RESULT: DIS ONLY

DO DIS DATA PREFER A SMALLER VALUE?

NLO NNPDF2.1 DEEP-INELASTIC DATA

$N_{\rm rep} = 500$ PER VALUE OF α_s

$$\alpha_s(M_z) = 0.1177 \pm 0.0009(\text{stat.}) \pm 0.0002(\text{proc.})$$

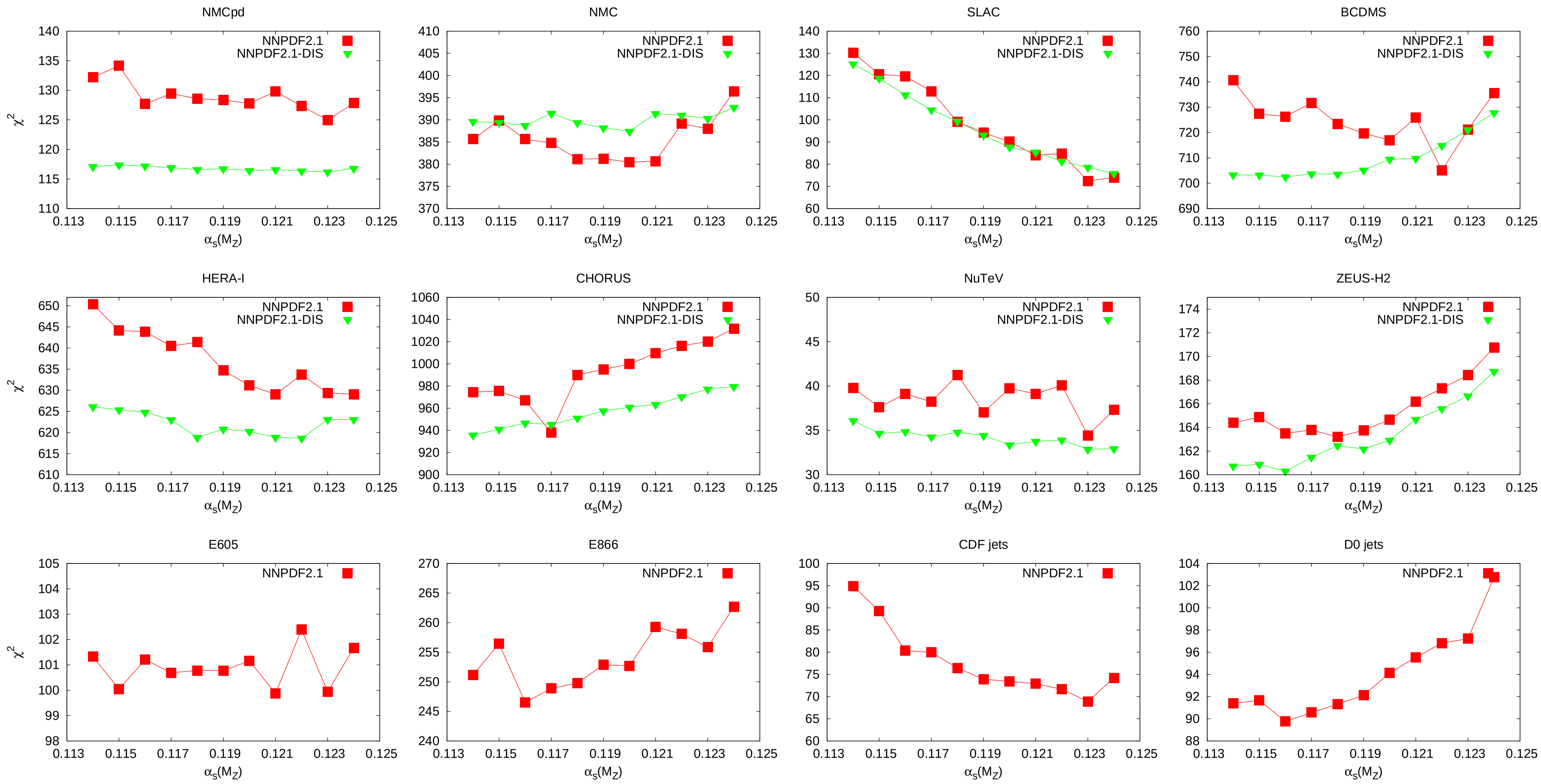


$\chi^2/\text{d.o.f.} = 0.8$ for the parabolic fit

- YES
- BUT NOT MUCH SMALLER & WITH LARGER UNCERTAINTY (COMPATIBLE WITHIN UNCERTAINTIES, AS IT OUGHT TO)

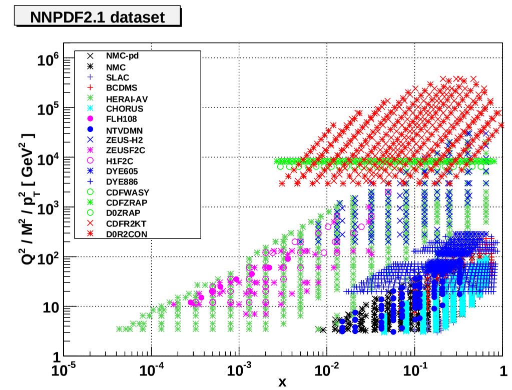
INDIVIDUAL EXPERIMENTS: WHAT'S GOING ON?

χ^2 PROFILE FOR INDIVIDUAL EXPERIMENTS IN GLOBAL & DIS FITS



- BCDMS SIMILAR TO HERA, NMC PRETTY FLAT
- DIS EXPERIMENTS (BCDMS+HERA) IN DIS FIT HAVE A "RUNAWAY DIRECTION" AT SMALL α_s,
  ABSENT IN GLOBAL FIT
- JET EXPERIMENTS EXCLUDE SMALL α_s

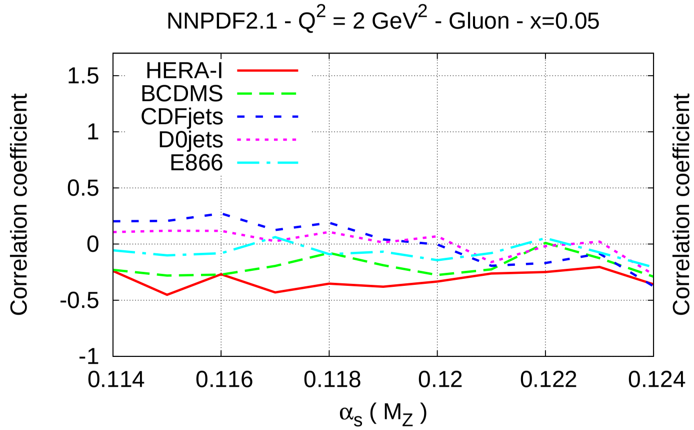
IS THERE A PROBLEM WITH DIS?

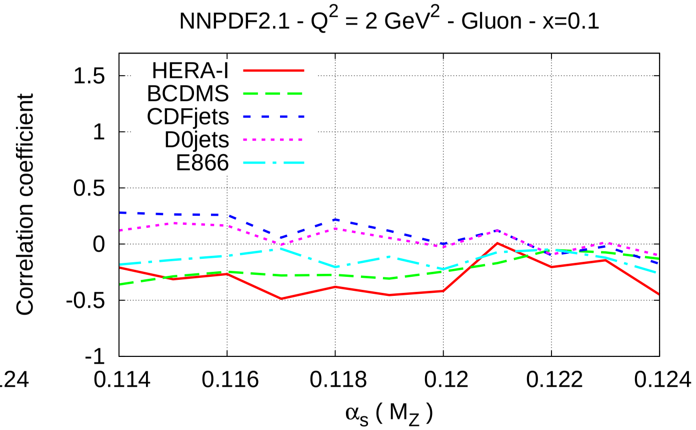


CONJECTURE

- IN DIS, GLUON DETERMINED BY SCALING VIOLATIONS $\Rightarrow$ CAN COMPENSATE SMALLER α_s WITH LARGER GLUON (OR CONVERSELY)
  $\Rightarrow$ RUNAWAY DIRECTIONS POSSIBLE IN DIS FIT
- JET DATA ARE AT MUCH LARGER SCALE, IF α_s TOO SMALL OR LARGE HIGH-SCALE GLUON WILL COME OUT WRONG
  $\Rightarrow$ RUNAWAY DIRECTION QUENCHED IN GLOBAL FIT

THE EVIDENCE



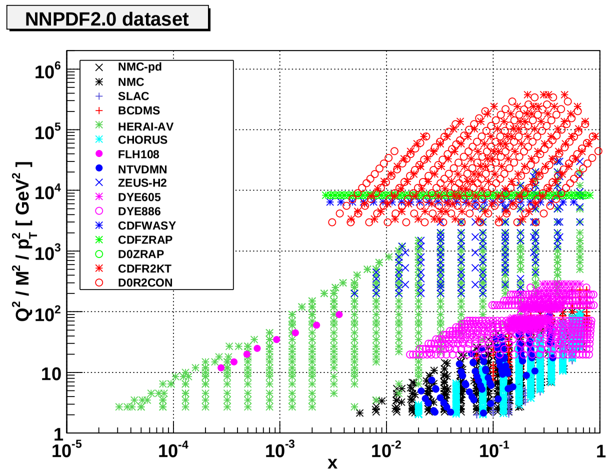


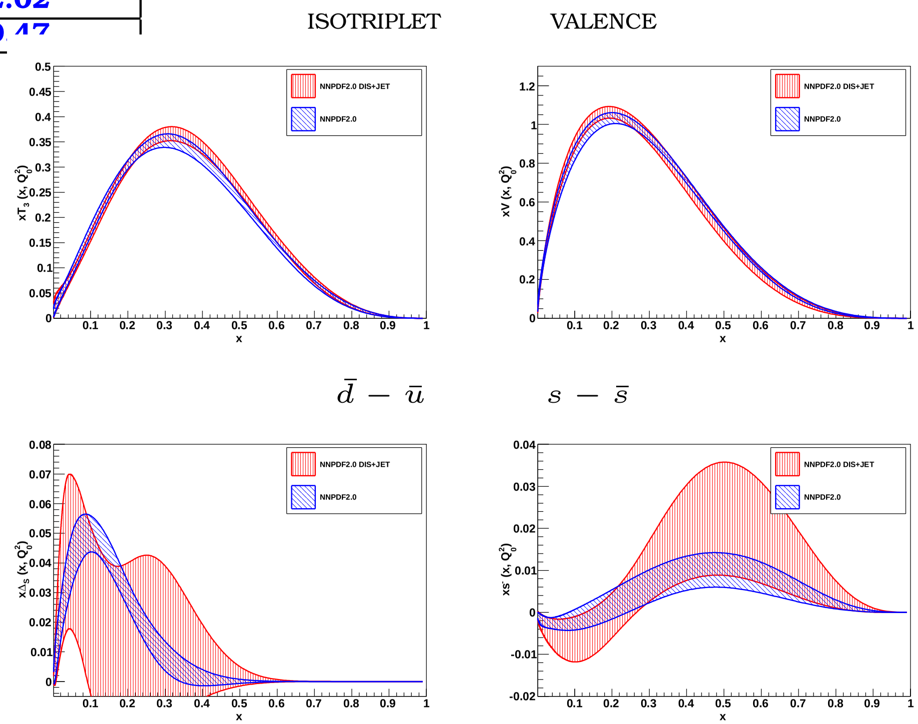
- COMPUTE THE CORRELATION BETWEEN χ^2 & PDFs AT THE GLOBAL BEST FIT
  LARGE (POSITIVE OR NEGATIVE) CORRN. $\Leftrightarrow$ RUNAWAY DIRECTION
- OPPOSITE SIGN CORRELATIONS $\Leftrightarrow$ DATA PULLING IN OPPOSITE DIRECTIONS
- AT LOW $\alpha_s = 0.114$ DIS FIT HAS A RUNAWAY DIRECTION STABILIZED BY JETS
- AT HIGH $\alpha_s = 0.122$ FIT IS GENERALLY STABLE,
  THOUGH STILL SOME PULL IN HERA LOW x

PRECISION PHYSICS FROM PREDICTIONS TO LHC DATA AND BACK

TEVATRON DRELL-YAN+W-PROD. DATA(NNPDF2.0)

	DIS	DIS+JET	NNPDF2.0
$\chi^2_{\rm tot}$	1.20	1.18	1.21
NMC-pd	0.85	0.86	0.99
NMC	1.69	1.66	1.69
SLAC	1.37	1.31	1.34
BCDMS	1.26	1.27	1.27
HERAI	1.13	1.13	1.14
CHORUS	1.13	1.11	1.18
FLH108	1.51	1.49	1.49
NTVDMN	0.71	0.75	0.67
ZEUS-H2	1.50	1.49	1.51
CDFR2KT	0.91	0.79	0.80
D0R2CON	1.00	0.93	0.93
DYE605	7.32	**10.35**	**0.88**
DYE866	2.24	**2.59**	**1.28**
CDFWASY	13.06	**14.13**	**1.85**
CDFZRAP	3.12	**3.31**	**2.02**
D0ZRAP	0.65	**0.68**	**0.47**



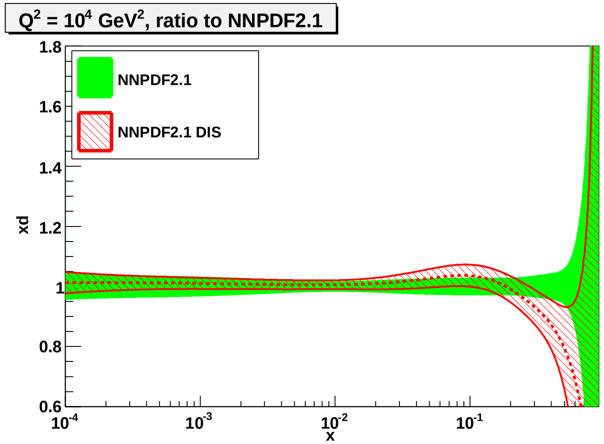


- VERY SUBSTANTIAL IMPROVEMENT IN FIT QUALITY WHEN DATA INCLUDED $\Rightarrow$ SOME PDF COMBINATIONS POORLY DETERMINED WITHOUT THESE DATA
- HUGE IMPROVEMENT IN SEA ASYM $\bar u - \bar d$ & STRANGENESS $s - \bar s$
- SIGNIFICANT IMPROVEMENT IN TOTAL VALENCE $(\sum_i (q_i - \bar q_i))$ & ISOTRIPLET $(u + \bar u - (d + \bar d))$

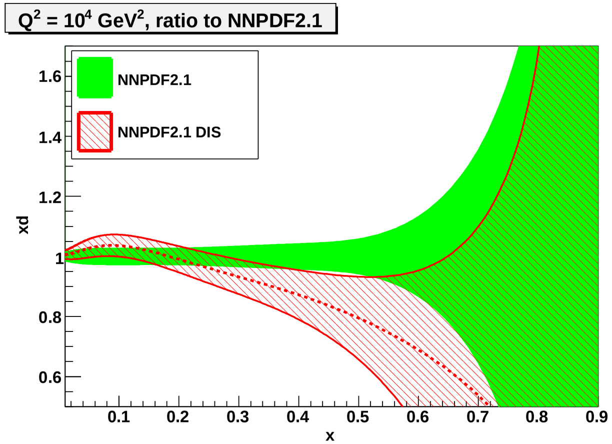
TEVATRON DRELL-YAN+W-PROD. DATA

QUARKS: DOWN

SMALL x



LARGE x

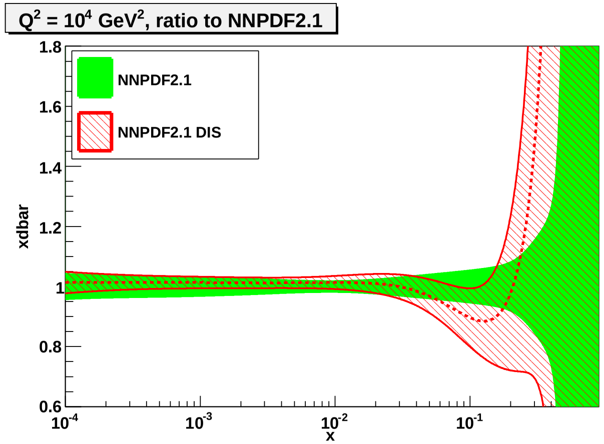


- WELL DETERMINED BY DIS AT SMALL X x
- SOME IMPROVEMENTS IN ACCURACY AT LARGE x
- TENSION BETWEEN DIS AND DY AT LARGE x

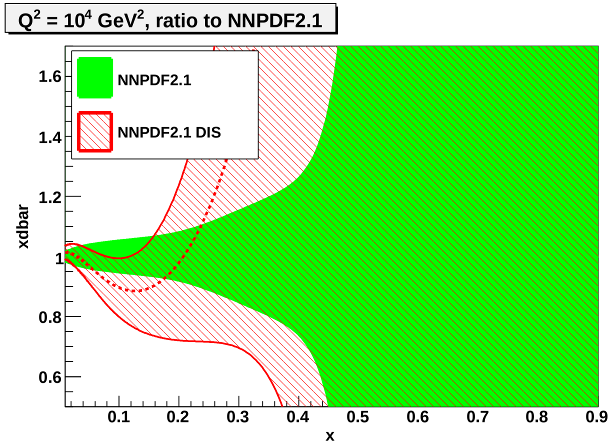
ANTIQUARKS: ANTIDOWN

- AT SMALL x ALL FLAVORS ARE ALIKE $\longrightarrow$ WELL DETERMINED BY DIS DATA
- SOME TENSION BETWEEN DIS AND DY AT MEDIUM-SMALL x
- DRAMATIC IMPROVEMENT IN ACCURACY AT LARGE x

SMALL x



LARGE x



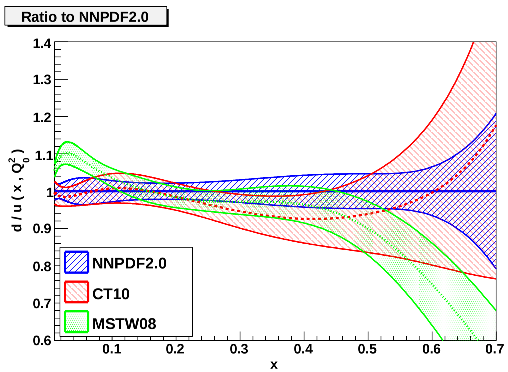
TENSION?

- SHAPES OF THE d QUARK FAVORED BY DIS AND DY DATA DO NOT AGREE
- IS THERE A PROBLEM WITH THE u/d RATIO?

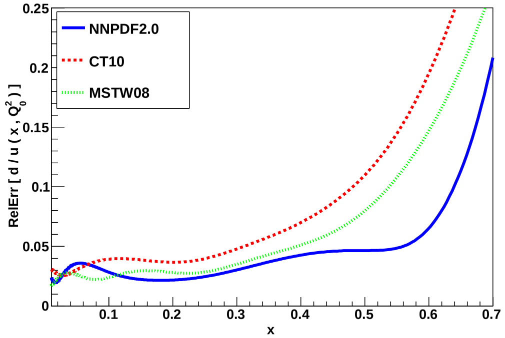
THE UP/DOWN RATIO

COMPARISON BETWEEN PDF SETS

u/d AT $Q^2 = 2$ GEV^2



RELATIVE UNCERTAINTY

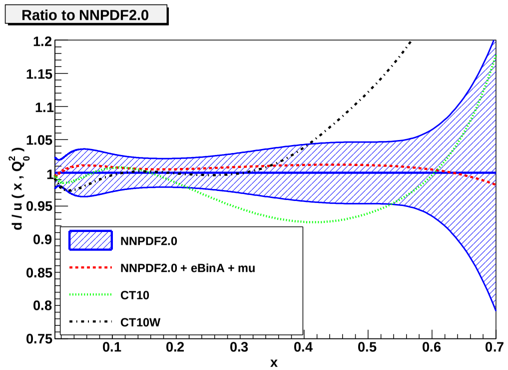


- u/d RATIO IN DIS DOMINATED BY BCDMS AND NMC DATA KNOWN TO HAVE CONSISTENCY PROBLEMS
- PERHAPS NUCLEAR CORRECTIONS & HIGHER TWISTS PLAY A ROLE IN NMC & BCDMS
- DIFFERENT PDF SETS BASED ON THE SAME DATA (BUT WITH DIFFERENT FIT QUALITY) IN PARTIAL DISAGREEMENT

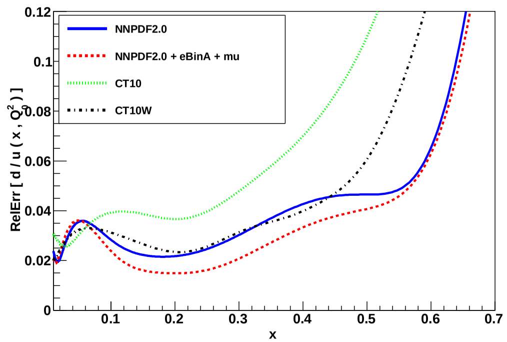
THE RUN II W ASYMMETRY DATA

- PRECISE D0 RUN II LEPTON-LEVEL DATA INCLUDED VIA REWEIGHTING
- NNPDF: SIZABLE IMPROVEMENT IN FIT QUALITY
- CTEQ: NO GLOBALLY CONSISTENT FIT FOUND: CT10W WITH W DATA, CT10 WITHOUT

u/d AT $Q^2 = 2$ GEV^2



RELATIVE UNCERTAINTY



THE W ASYMMETRY AT THE TEVATRON

- PRECISE W ASYMMETRY DATA DETERMINE THE u/d RATIO ACCURATELY
- EVENTUALLY, IT WOULD BE BETTER TO DO WITHOUT NMC/BCDMS DATA (NO LOW Q^2 DATA, NO NUCLEAR TARGETS)

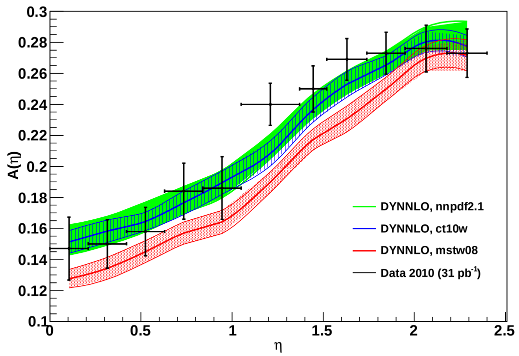
THE W LEPTON ASYMMETRY AT THE LHC

THE $p\bar{p} \to W^{\pm} \to l^{\pm}\nu$ ASYMMETRY: $A_\ell(y_\ell) = \dfrac{d\sigma(W^+)/dy_\ell - d\sigma(W^-)/dy_\ell}{d\sigma(W^+)/dy_\ell + d\sigma(W^-)/dy_\ell}$

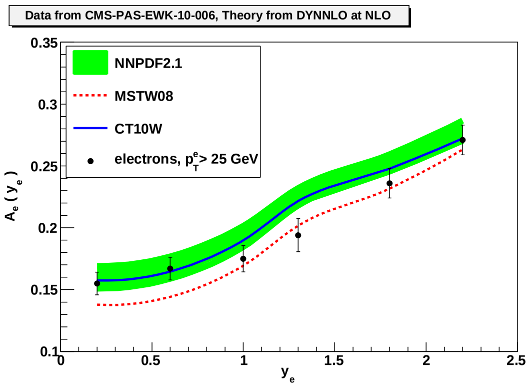
- ACCURATE ATLAS (arXiv:1103.2929) AND CMS (arXiv:1103.3470) DATA PUBLISHED EARLIER THIS MONTH
- NNPDF1.2 PREDICTION AS GOOD AS CTEQ (BETTER FOR ATLAS), BOTH RATHER BETTER THAN MSTW

$\chi^2/N_{\rm dat}$ (EL)	ATLAS	CMS $p_T^l \geq 25$ GEV	CMS $p_T^l \geq 30$ GEV
NNPDF2.1	0.68	1.8	1.6
MSTW08	3.24	1.8	2.3
CT10W	0.81	1.2	1.3

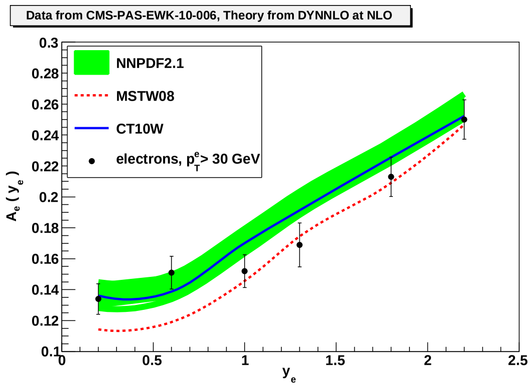
ATLAS



CMS $p_t > 25$ GEV



CMS $p_t > 30$ GEV



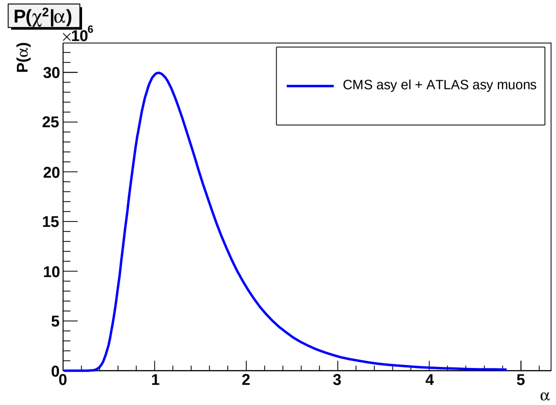
LHC W asymmetry $\Rightarrow$ PDFs

reweighted observables

- LHC data can be included in the PDF determination using the reweighting method

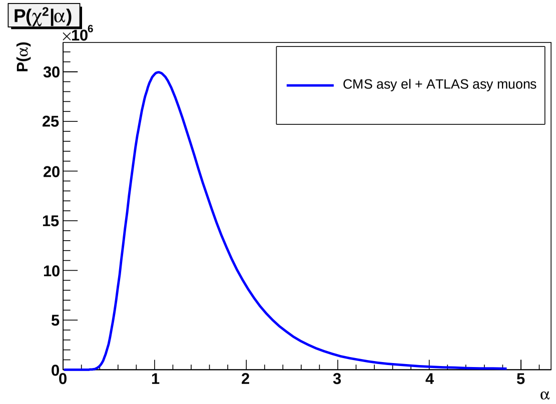
LHC W ASYMMETRY $\Rightarrow$ PDFS
REWEIGHTED OBSERVABLES

- LHC DATA CAN BE INCLUDED IN THE PDF DETERMINATION USING THE REWEIGHTING METHOD
- CAN CHECK FOR CONSISTENCY

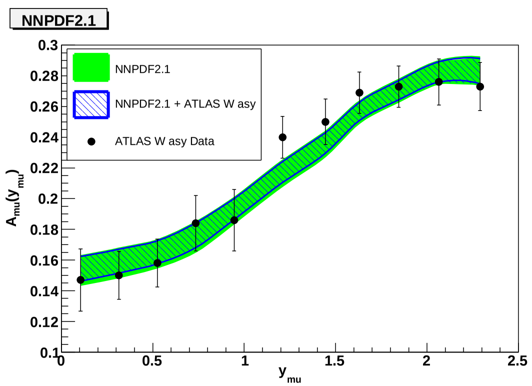


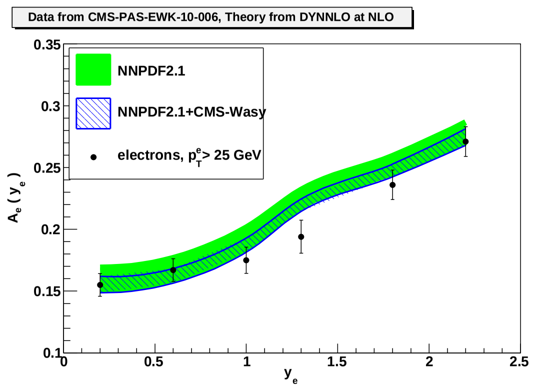
LHC W ASYMMETRY $\Rightarrow$ PDFS
REWEIGHTED OBSERVABLES

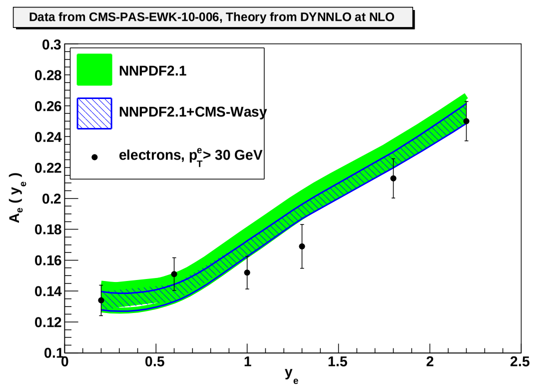
- LHC DATA CAN BE INCLUDED IN THE PDF DETERMINATION USING THE REWEIGHTING METHOD
- CAN CHECK FOR CONSISTENCY
- RECOMPUTE OBSERVABLES AFTER INCLUSION & VERIFIY FIT QUALITY



$\chi^2/N_{\rm dat}$ (EL)	ATLAS	CMS $p_T^l \geq 25$ GEV	CMS $p_T^l \geq 30$ GEV	ALL
NNPDF2.1	0.68	1.8	1.6	0.95
NNPDF2.1+EACH DATASET	0.63	1.03	1.16	-
NNPDF2.1+ALL DATA	-	-	-	0.92





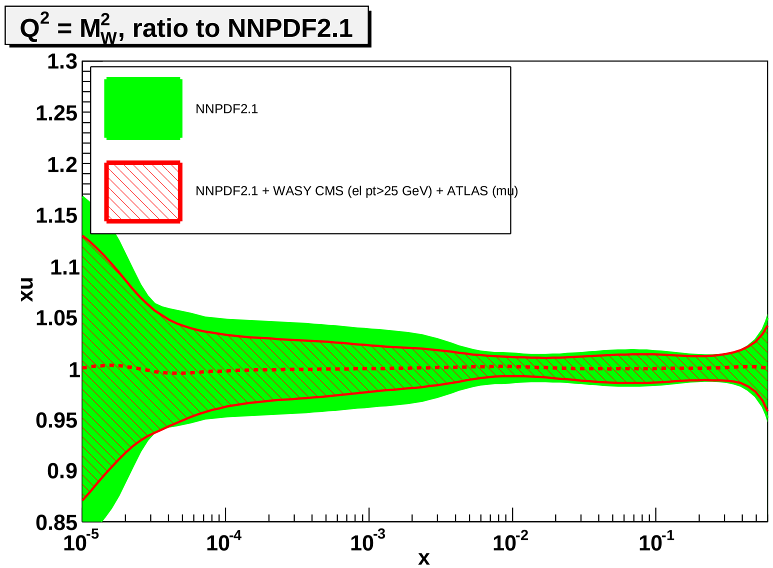


LHC W ASYMMETRY ⇒ PDFs
REWEIGHTED PDFs

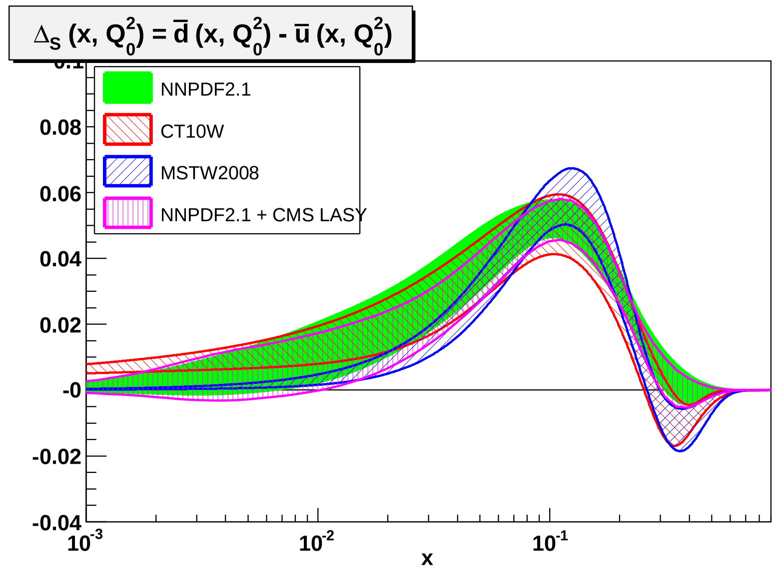
WHAT IS THE EFFECT OF THE LHC DATA ON PDFs?

- ALL LIGHT QUARK UNCERTAINTIES REDUCED
- GOOD CONSISTENCY ⇒
  - CENTRAL VALUES DON'T CHANGE MUCH
  - UNCERTAINTIES SHRINK
- DIFFERENCE W.T. TO OTHER PDFs: MOSTLY $\bar{u} - \bar{d}$
  - NNPDF CENTRAL VALUE UNCHANGED WITH SMALLER UNCERTAINTY
  - MSTW (CTEQ) SHAPE SOMEWHAT DIFFERENT ⇒ MSTW/CTEQ PREDICTION LESS GOOD

THE UP DISTRIBUTION



$\bar{u} - \bar{d}$

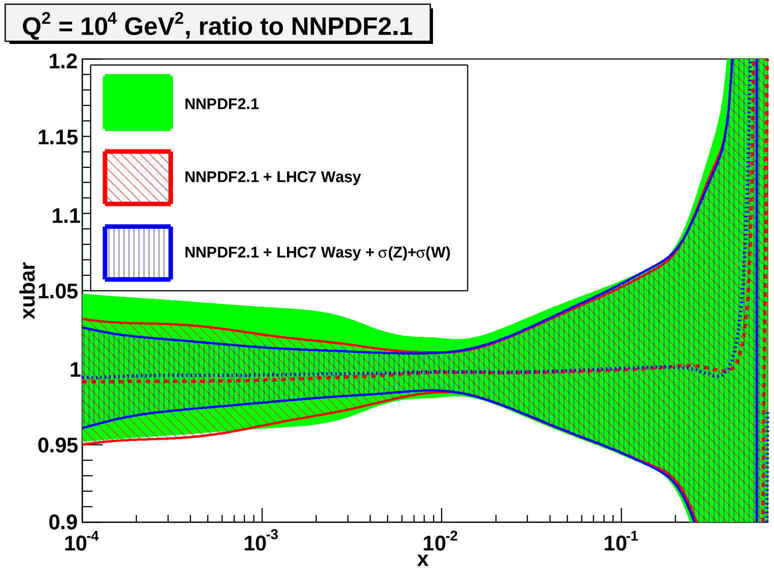


FUTURE LHC DATA

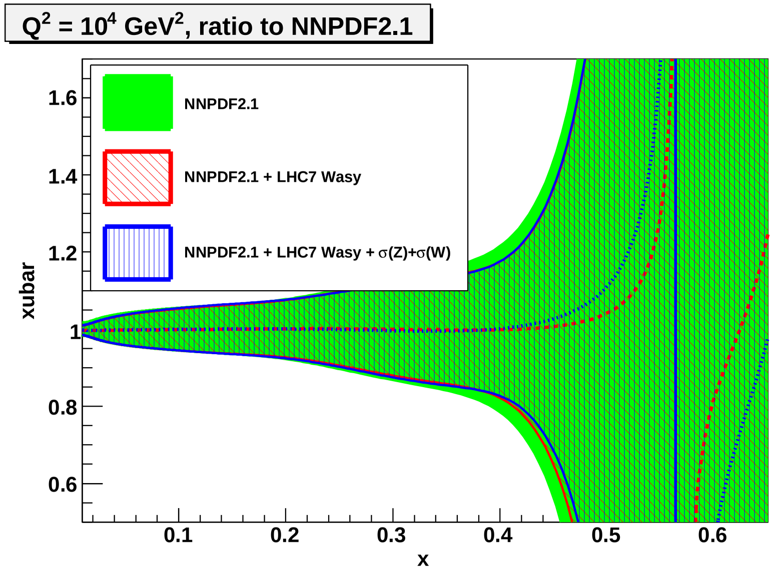
TEST THE IMPACT OF LHC 7TeV W AND Z DATA USING REWEIGHTING IN THE FOLLOWING "TOY" SCENARIOS:

- W ASYMMETRY MEASURED WITH $\sim 5\%$ ACCURACY AT ATLAS (KINEMATICS COURTESY OF A. GLAZOV) $\Rightarrow$ SIGNIFICANT IMPROVEMENT FOR ALL FLAVORS AT SMALL x
- W ASYMMETRY + TOTAL W AND Z CROSS-SECTIONS MEASURED TO 2% ACCURACY $\Rightarrow$
  - FURTHER IMPROVEMENT AT SMALL x
  - NOTE EFFECT SEEN ON ALL FLAVORS AND ANTIFLAVORS

ANTIUP SMALL x



ANTIUP LARGE x



PROGRESS:
TOWARDS NNLO PDFs

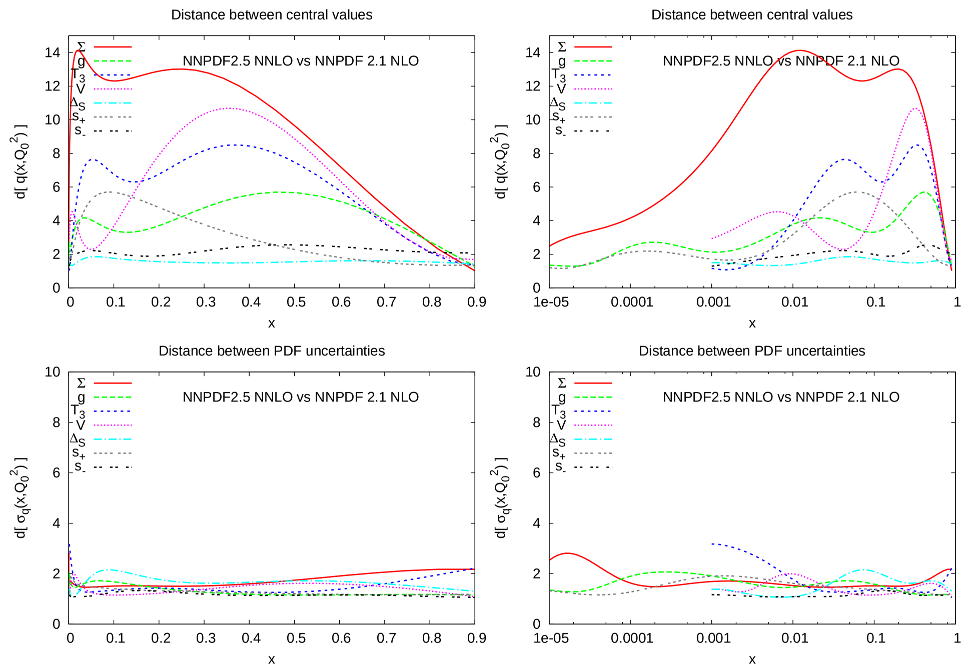
NNPDF @ NNLO

- FULL INCLUSION OF NNLO COMPUTATIONALLY VERY INTENSIVE
  - ABKM IS TRULY NNLO, BUT DOES NOT INCLUDE ANY HADRON COLLIDER DATA
  - MSTW, CTEQ: DRELL-YAN BEYOND LO INCLUDED WITH (LOCAL) K–FACTORS
- NNPDF $\Rightarrow$ FASTKERNEL METHOD FOR INCLUSION OF HO TERMS, IMPLEMENTED FOR DIS AT NNLO AND FOR DY, JETS AT NLO
  - PRELIMINARY RESULTS WITH DIS @ NNLO, DY, JETS AT NLO
  - GLOBAL (DIS+DY) NNLO FIT IN PREPARATION

IMPACT ON PDFS

- CENTRAL VALUES MOVE BY $1.5 - 2\ \sigma$
- UNCERTAINTIES ALMOST UNCHANGED

DISTANCES



NNPDF @ NNLO

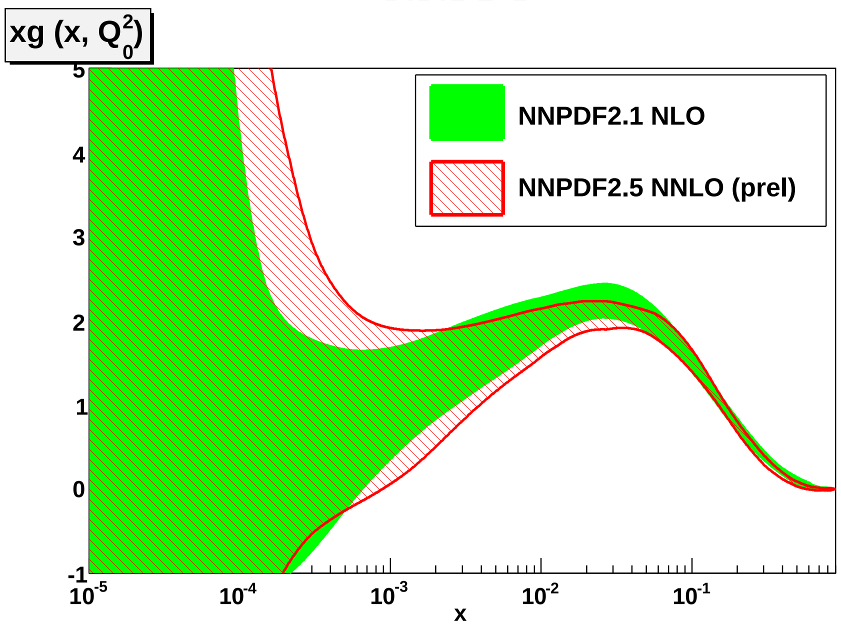
- FULL INCLUSION OF NNLO COMPUTATIONALLY VERY INTENSIVE
  - ABKM IS TRULY NNLO, BUT DOES NOT INCLUDE ANY HADRON COLLIDER DATA
  - MSTW, CTEQ: DRELL-YAN BEYOND LO INCLUDED WITH (LOCAL) K–FACTORS
- NNPDF $\Rightarrow$ FASTKERNEL METHOD FOR INCLUSION OF HO TERMS, IMPLEMENTED FOR DIS AT NNLO AND FOR DY, JETS AT NLO
  - PRELIMINARY RESULTS WITH DIS @ NNLO, DY, JETS AT NLO
  - GLOBAL (DIS+DY) NNLO FIT IN PREPARATION

IMPACT ON PDFS

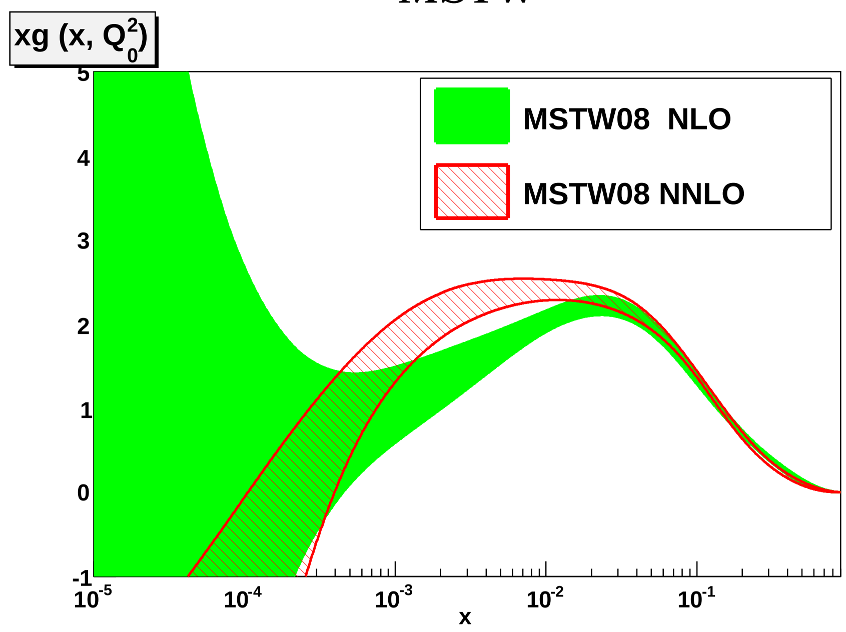
- STABLE PDFS INCLUDING GLUON!
- UNLIKE OTHER NNLO SETS (MSTW)

NLO VS NNLO

NNPDF



MSTW



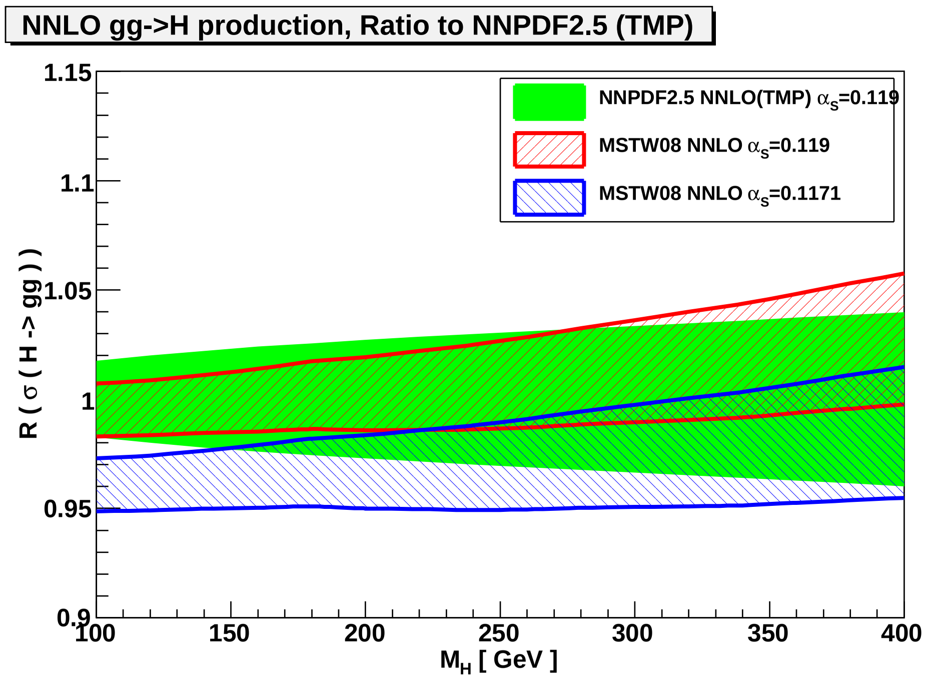
NNPDF @ NNLO

- FULL INCLUSION OF NNLO COMPUTATIONALLY VERY INTENSIVE
  - ABKM IS TRULY NNLO, BUT DOES NOT INCLUDE ANY HADRON COLLIDER DATA
  - MSTW, CTEQ: DRELL-YAN BEYOND LO INCLUDED WITH (LOCAL) K–FACTORS
- NNPDF $\Rightarrow$ FASTKERNEL METHOD FOR INCLUSION OF HO TERMS, IMPLEMENTED FOR DIS AT NNLO AND FOR DY, JETS AT NLO
  - PRELIMINARY RESULTS WITH DIS @ NNLO, DY, JETS AT NLO
  - GLOBAL (DIS+DY) NNLO FIT IN PREPARATION

IMPACT ON PHYSICAL OBSERVABLES: HIGGS

- GOOD AGREEMENT BETWEEN MSTW AND NNPDF
- VALUE OF α_s VERY IMPORTANT!

THE HIGGS CROSS SECTION



THE NNPDF TIMELINE

- 2002 FIT OF NUCLEON STRUCTURE FUNCTION TO TWO DATASETS
- 2005 FIT OF NUCLEON STRUCTURE FUNCTION TO MANY DATASETS
- 2007 FIT OF ONE SINGLE (ISOTRIPLET) PARTON DISTRIBUTION TO TWO DATASETS
- 2008 FIRST FULL PARTON SET: NNPDF1.0 (WORLD DIS DATA)
- 2009 PARTON SET WITH ACCURATE STRANGENESS: NNPDF1.2 (DIS+ DIMUON DATA)
- 2010 FIRST GLOBAL PARTON SET: NNPDF2.0
- JANUARY 2011: NNPDF2.1, A GLOBAL NLO SET WITH FONLL/ACOT GENERAL MASS SCHEME
- EARLY 2011: PRECISION STUDIES AT NLO: α_s, LHC DATA
- SPRING 2011: NNLO GLOBAL SET
- 2011-2012: STUDIES OF TH. UNCERTAINTIES; RESUMMED PDF SETS; POLARIZED PDF SETS

Conclusion

- PDFs can be determined, assessed and used in a quantitatively accurate way
- LHC itself (together with HERA) will provide the necessary experimental input
- we are ready for discovery physics at the LHC

EXTRAS

NMC data: a problem?

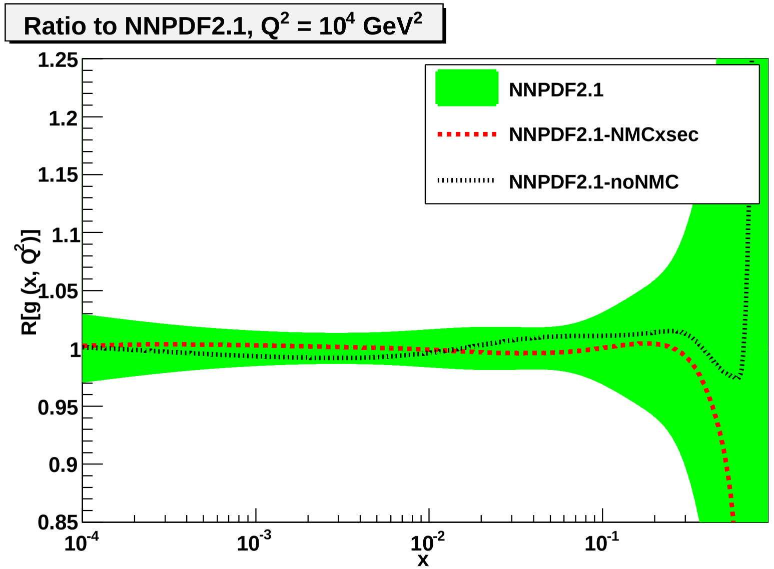
- NMC proton and deuteron data for DIS F_2 structure function
- F_2structure function data used in MSTW, CTEQ, NNPDF fits, would the use of cross section data make a difference? (Alekhin, Blümlein, Moch, 2011)
  $\widetilde{\sigma}(x, y, Q^2) = F_2(x, Q^2)\left(2 - 2y + y^2/[1 + R(x, Q^2)]\right)$

NMC DATA: A PROBLEM?

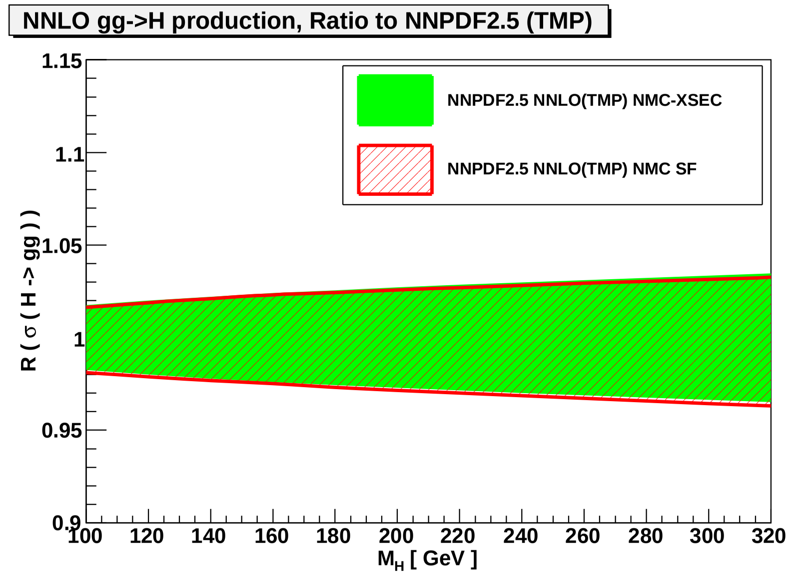
- NMC PROTON AND DEUTERON DATA FOR DIS F_2 STRUCTURE FUNCTION
- F_2STRUCTURE FUNCTION DATA USED IN MSTW, CTEQ, NNPDF FITS, WOULD THE USE OF CROSS SECTION DATA MAKE A DIFFERENCE? (Alekhin, Blümlein, Moch, 2011)
  $\widetilde{\sigma}(x,y,Q^2) = F_2(x,Q^2)\left(2-2y+y^2/[1+R(x,Q^2)]\right)$

NNPDF FIT REPEATED WITH NMC XSECT OR NO NMC

EFFECT ON GLUON



EFFECT ON HIGGS XSECT



NO SIGNIFICANT CHANGE OBSERVED

Monte Carlo data generation

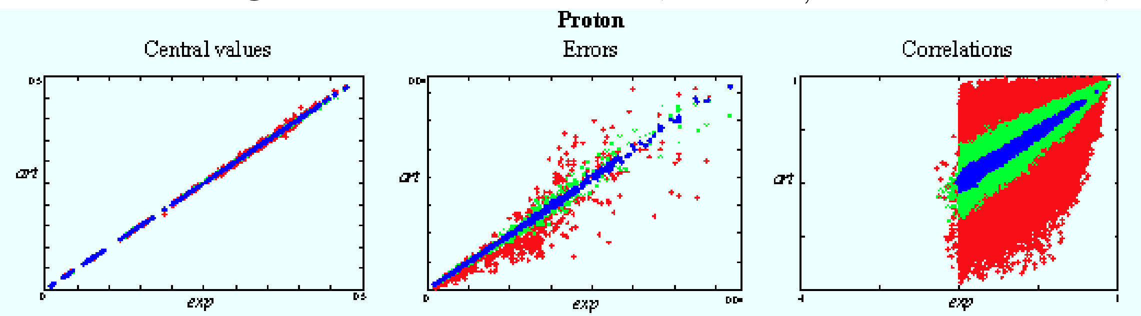
- BCDMS+ NMC proton & deuteron F_2 data (full correlated systematics available), taken at 4 beam energies
- on top of stat. errors, 4 systematics + 1 normalization (NMC) or 6 systematics + 1 absolute & 2 relative normalizations (BCDMS), with various forms of correlation (full, or for each target, or for each beam energy)

Generate data according to a multigaussian distribution

$$F_i^{(art)\,(k)} =$$

$$(1 + r_5^{(k)}\,\sigma_N)\sqrt{1 + r_{i,6}^{(k)}\,\sigma_{N_t}}\sqrt{1 + r_{i,7}^{(k)}\,\sigma_{N_b}}\left[F_i^{(exp)} + \frac{r_{i,1}^{(k)}\,f_b + r_{i,2}^{(k)}\,f_{i,s} + r_{i,3}^{(k)}\,f_{i,r}}{100}F_i^{(exp)} + r_{i,s}^{(k)}\,\sigma_s^i\right]$$

r univariate gaussian random nos., one $r_{i,s}$ for each data, but single $r_{i,j}$ for all correlated data

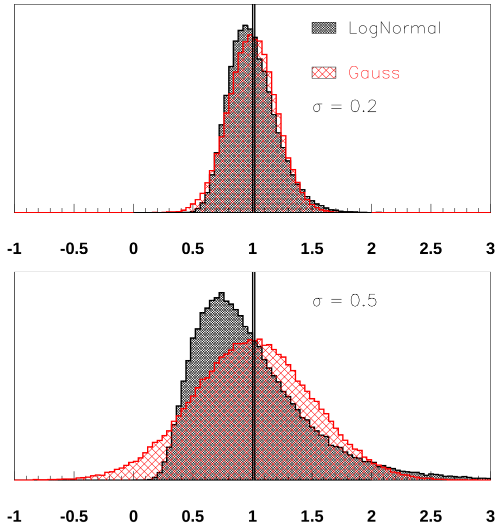


Scatter plot art. vs. exp. for 10 (red) 100 (green) and 1000 (blue) replicas

Need 1000 replicas to reproduce correlations to percent accuracy

PARAMETRIZATION UNCERTAINTIES?
NONGAUSSIAN BEHAVIOUR?

LOGNORMAL VS. GAUSSIAN

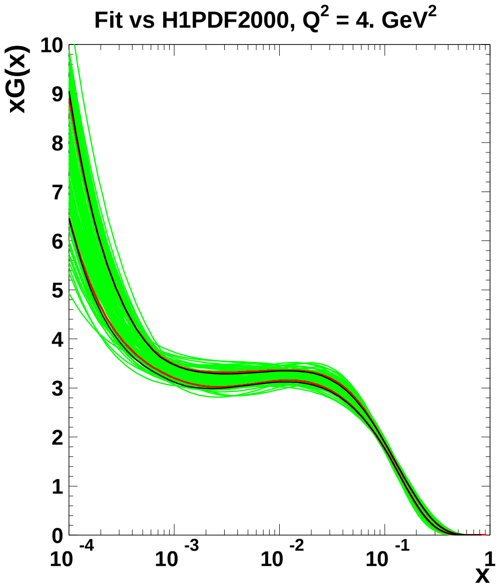


THE HERALHC BENCHMARK

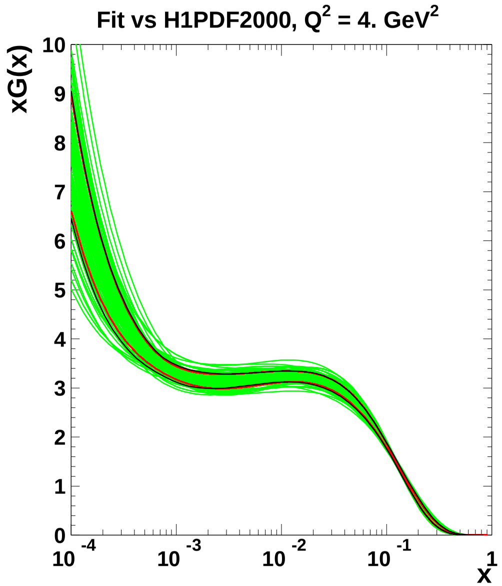
(Feltesse, Glazov, Radescu + NNPDF 2008)

- TRY EXPERIMENTAL SYSTEMATICS GIVEN BY EITHER GAUSSIAN OR LOGNORMAL DISTRIBUTION
- REPEAT (BENCHMARK) HERAPDF, WITH MONTECARLO LOGNORMAL OR GAUSSIAN, IN EITHER CASE DETERMINE UNCERTAINTY EITHER WITH HESSIAN OR MONTECARLO

LOGNORMAL: HESS. VS MC



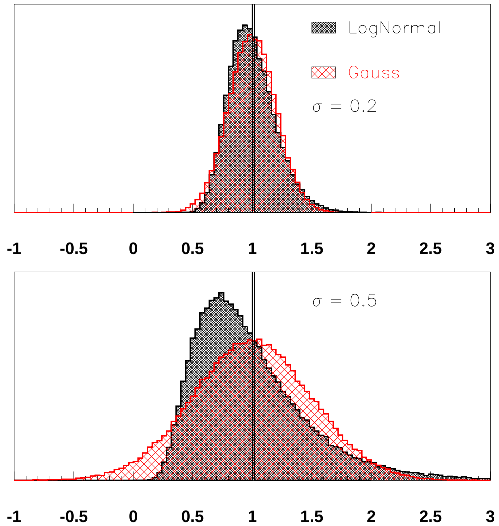
GAUSSIAN: HESS. VS MC



- NO DIFFERENCE BETWEEN LOGNORMAL, GAUSSIAN, MC, HESSIAN

PARAMETRIZATION UNCERTAINTIES?
NONGAUSSIAN BEHAVIOUR?

LOGNORMAL VS. GAUSSIAN

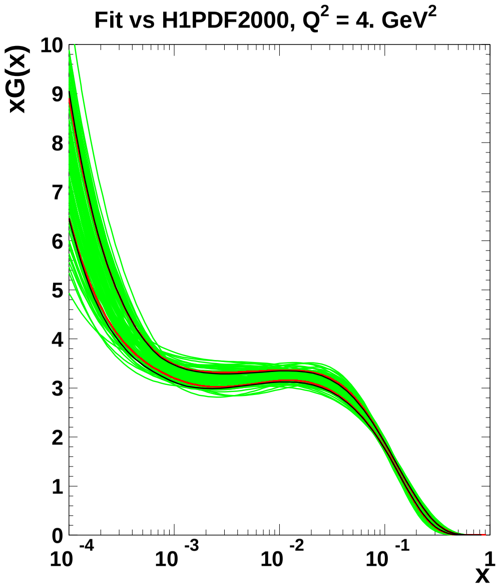


THE HERALHC BENCHMARK

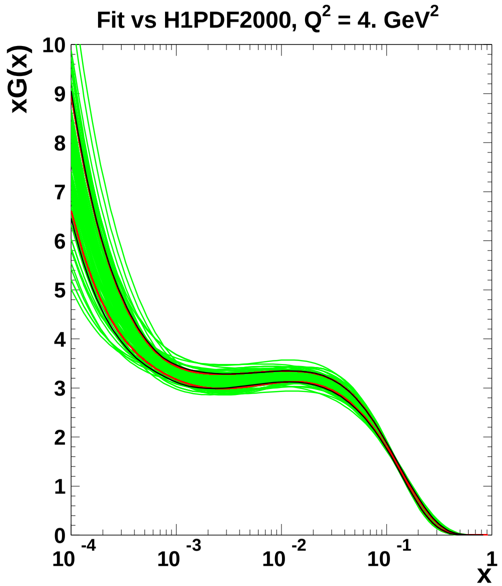
(Feltesse, Glazov, Radescu + NNPDF 2008)

- TRY EXPERIMENTAL SYSTEMATICS GIVEN BY EITHER GAUSSIAN OR LOGNORMAL DISTRIBUTION
- REPEAT (BENCHMARK) HERAPDF, WITH MONTECARLO LOGNORMAL OR GAUSSIAN, IN EITHER CASE DETERMINE UNCERTAINTY EITHER WITH HESSIAN OR MONTECARLO
- COMPARE TO NNPDF FIT TO SAME DATA

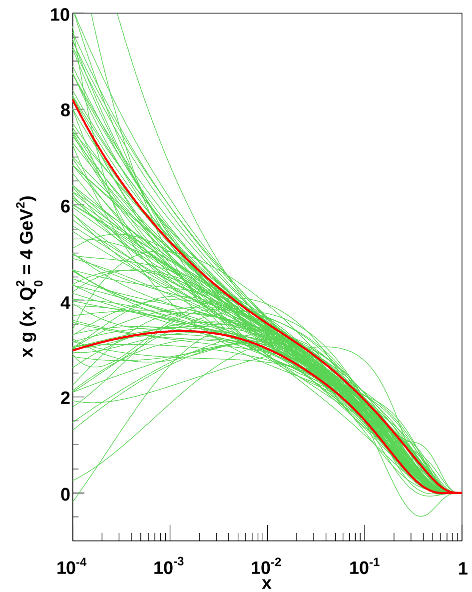
LOGNORMAL: HESS. VS MC



GAUSSIAN: HESS. VS MC



NNPDF

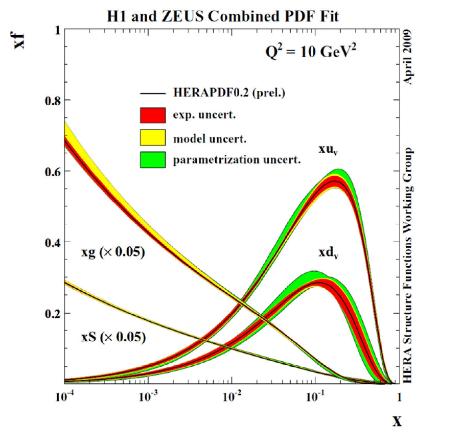


- NO DIFFERENCE BETWEEN LOGNORMAL, GAUSSIAN, MC, HESSIAN
- SIZABLE DIFFERENCE WR TO FLEXIBLE NNPDF PARAMETRIZATION

PARAMETRIZATION UNCERTAINTIES?

EXPLORING THE SPACE OF PARAMETERS: HESSIAN APPROACH

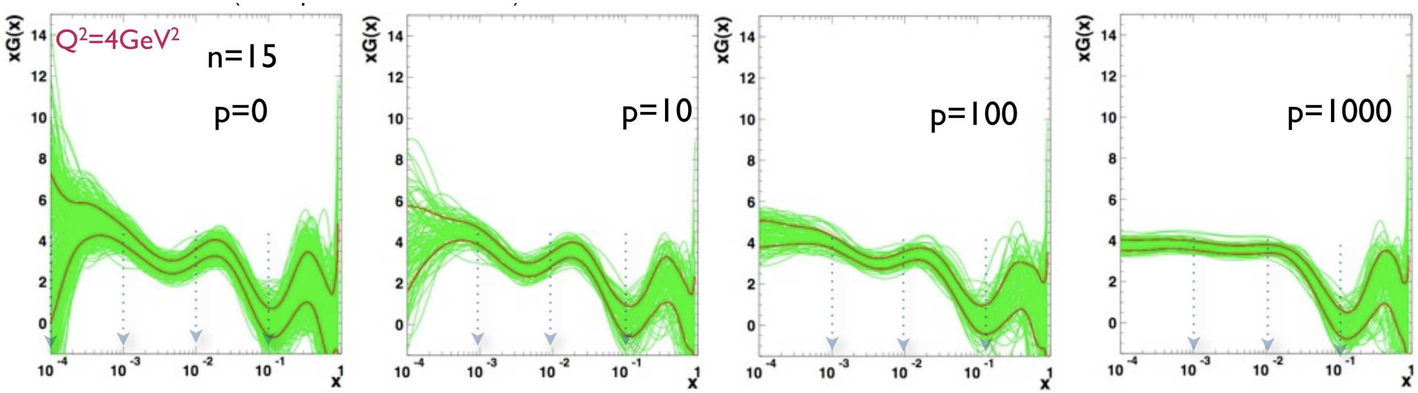
- IN HESSIAN APPROACH CAN VARY THE FUNCTIONAL FORM, ASSUMPTIONS, STARTING SCALE
- DONE IN THE HERAPDF1.0 FIT: VARIATION OF STRANGENESS FRACTION, LARGE x BEHAVIOUR, HIGHER ORDER POLYNOMIAL TERMS
- NO TOLERANCE ($\Delta\chi^2 = 1$), UNCERTAINTY DOUBLED



- Red: experimental uncertainties
- Yellow: model uncertainties
- Green: pdf parametrization uncertainties

ORTHOGONAL POLYNOMIALS

- OLD IDEA (PARISI, SOURLAS, 1978; ZOMER 1996): EXPAND PDFS OVER BASIS OF ORTHOGONAL POLYNOMIALS
- GLAZOV, RADESCU, 2009: COUPLED TO MONTE CARLO METHOD
- LENGTH PENALTY TO STABILIZE THE FIT



(Glazov, Radescu, 2009)

NORMALIZATION UNCERTAINTIES

- NORMALIZATION UNCERTAINTIES IN COVARIANCE MATRIX
  $(\mathrm{cov}_{t_0})_{IJ} = \sigma_{I,n}\sigma_{J,n}F_I F_J$
  $\Rightarrow$ MAXIMUM-LIKELIHOOD RESULT BIASED (d'Agostini, 1994)

- "PENALTY TRICK": RESCALE BY λ & ADD $\frac{(\lambda-1)^2}{\sigma_n}^2$ TO χ^2
  SOMETIMES ALSO HIGHER ORDER POWERS
  $\Rightarrow$ ALSO BIASED, THOUGH BIAS DOES NOT GROW WITH N_{dat}

- BIAS DUE TO χ^2 NOT QUADRATIC IN MEASURED QUANTITY
  (NONGAUSSIAN AND IMPROPERLY NORMALIZED LIKELIHOOD)

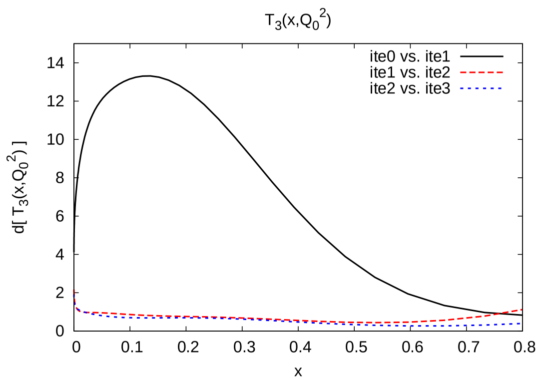
THE t_0 METHOD (R.D. Ball et al., 2010)

- NORMALIZATION UNCERTAINTIES IN COVARIANCE MATRIX, BUT COMPUTED AS FUNCTION OF RESULT OF PREVIOUS FIT $F_I^{(0)}$: $(\mathrm{cov}_{t_0})_{IJ} = \sigma_{I,n}\sigma_{J,n}F_I^{(0)}F_J^{(0)}$

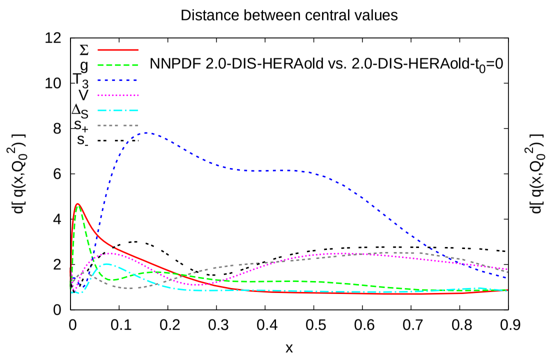
- ITERATE UNTIL CONVERGENCE

DISTANCE BETWEEN ITERATIONS

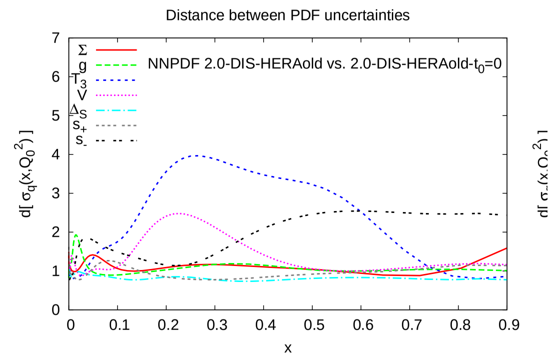
(TRIPLET)



t_0: IMPACT ON CENTRAL VALS



t_0: IMPACT ON UNCERTAINTIES



PREPROCESSING

- PDFs PARAMETRIZED AS $f(x,Q_0^2) = A_f\,(1-x)^{m_f}\,x^{-n_f}\mathrm{N}N_f(x)$, WITH $\mathrm{N}N_f(x)$ A NEURAL NETWORK
- PREFACTOR NECESSARY IF ONE WANTS $f(x)$ UNBOUNDED IN $0,1$ BECAUSE OUTPUT OF NEURAL NET BOUNDED
- WITH ARBITRARY LONG TRAINING, ANY ADMISSIBLE BEHAVIOUR CAN BE LEARNT, BUT TRAINING VERY LONG IF NN HAS TO OFFSET SHARP RISE OR DROP
- EXPONENTS m_f, n_f CHOSEN A RANDOMN WITH FLAT DISTRIBUTION IN WIDE RANGE
- RANGE DETERMINED BY WIDENING UNTIL FIT QUALITY DETERIORATES
- TEST FOR UNIFORM FIT QUALITY BY DETERMINING CORRELATION BETWEEN χ^2 AND PREPROCESSING EXPONENTS IN GIVEN RANGE

PREPROCESSING EXPONENTS: RANGE AND CORRELATION

PDF	$[m_{\rm min}, m_{\rm max}]$	$[n_{\rm min}, n_{\rm max}]$	$r\left[\chi^2, m\right]$	$r\left[\chi^2, n\right]$
$\Sigma(x,Q_0^2)$	$[2.55, 3.45]$	$[1.05, 1.35]$	-0.018	0.131
$g(x,Q_0^2)$	$[1.05, 1.35]$	$[1.05, 1.35]$	-0.002	0.050
$T_3(x,Q_0^2)$	$[2.55, 3.45]$	$[0, 0.5]$	-0.023	-0.130
$V_T(x,Q_0^2)$	$[2.55, 3.45]$	$[0, 0.5]$	0.003	-0.068
$\Delta_S(x,Q_0^2)$	$[12, 14]$	$[-0.95, -0.65]$	0.000	-0.069
$s^+(x,Q_0^2)$	$[2.55, 3.45]$	$[1.05, 1.35]$	0.021	-0.055
$s^-(x,Q_0^2)$	$[2.55, 3.45]$	$[0, 0.5]$	-0.027	-0.015

THE “HESSIAN MONTE CARLO”

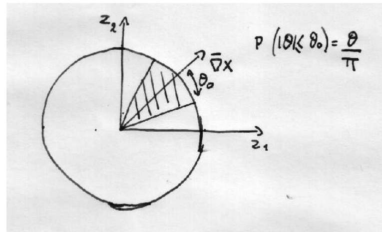
Q: IF ONE PICKS REPLICAS AT RANDOM ON THE ONE-SIGMA CONTOUR WHAT IS THE CHANCE OF “FILLING” THE ENVELOPE?

A: DETERMINE THE PROBABILITY FOR AT LEAST ONE REPLICA TO BE WITHIN ANGLE θ OF DIRECTION $\vec{\nabla} X$ OF MAX

TWO PARAMETERS: ONE REPLICA WITH $\theta < \theta_0 \Rightarrow P(2, 1 : \theta_0) = \frac{\theta_0}{\pi}$

PROBABILITY OF MAX(ENVELOPE)=$\sigma_X \cos\theta_0$

$\Rightarrow$ ALL n REPLICAS HAVE $\theta > \theta_0 \Rightarrow P(2, n; \theta_0) = \left(1 - \frac{\theta_0}{\pi}\right)^n$

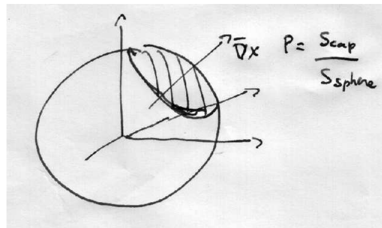


d PARAMETERS: ONE REPLICA WITH $\theta < \theta_0$

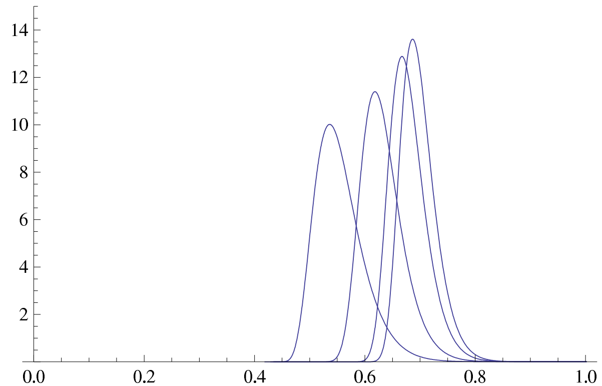
$$\Rightarrow P(d, 1 : \theta_0) = \frac{\Gamma\left(\frac{d}{2}\right)}{(d-1)\sqrt{\pi}\,\Gamma\left(\frac{d-1}{2}\right)} \theta_0^{d-1}(1 + O(\theta_0)) \approx \frac{\theta_0^{d-1}}{\sqrt{2\pi d}}$$

PROBABILITY OF MAX(ENVELOPE)=$\sigma_X \cos\theta_0$

$$\Rightarrow P(d, n; \theta_0) = \left(1 - \frac{\theta_0^{d-1}}{\sqrt{2\pi d}}\right)^n$$



PROBABILITY FOR THE WIDTH OF THE ENVELOPE TO BE SMALLER BY A FACTOR R THAN THE STANDARD DEVIATION σ_X PLOTTED VS R FOR $d = 23$ PARAMETERS AND $n = 10, 100, 500, 1000$ REPLICAS



MONTE CARLO ERROR ESTIMATES

PARAMETER SPACE: NOT ADVISABLE

OBSERVABLE X DEPENDS ON PARAMETERS $\vec{z}$

VARIANCE: $\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$

AVERAGES: $\langle X \rangle = \int d^d z X(\vec{z}) P(\vec{z})$, WITH

$P(\vec{z}) \Rightarrow$ PROBABILITY DISTN. OF PARAMETER VALUES

& INTEGRAL PERFORMED BY MONTE CARLO SAMPLING

HOW MANY REPLICAS DOES ONE NEED? THREE BINS PER PARM $\Rightarrow 3^d$ BINS

FOR 23 PARMS., NEED $> \ 10^{11}$ REPLICAS

DATA SPACE

DIAGONALIZATION: CHOOSE PARM z_1 ALONG $\vec{\nabla} X$

ALL OTHER PARMS $\Rightarrow$ FLAT DIRECTIONS

AVERAGES: $\langle X \rangle = \int dz_1 X(\vec{z}) P(z_1)$

HOW MANY REPLICAS DOES ONE NEED? ONE–DIMENSIONAL AVERAGE OF n REPLICAS CONVERGES TO TRUE AVERAGE WITH STANDARD DEV. $\frac{\sigma}{\sqrt{n}}$

10 REPLICAS ENOUGH FOR $\frac{\sigma}{3}$ ACCURACY

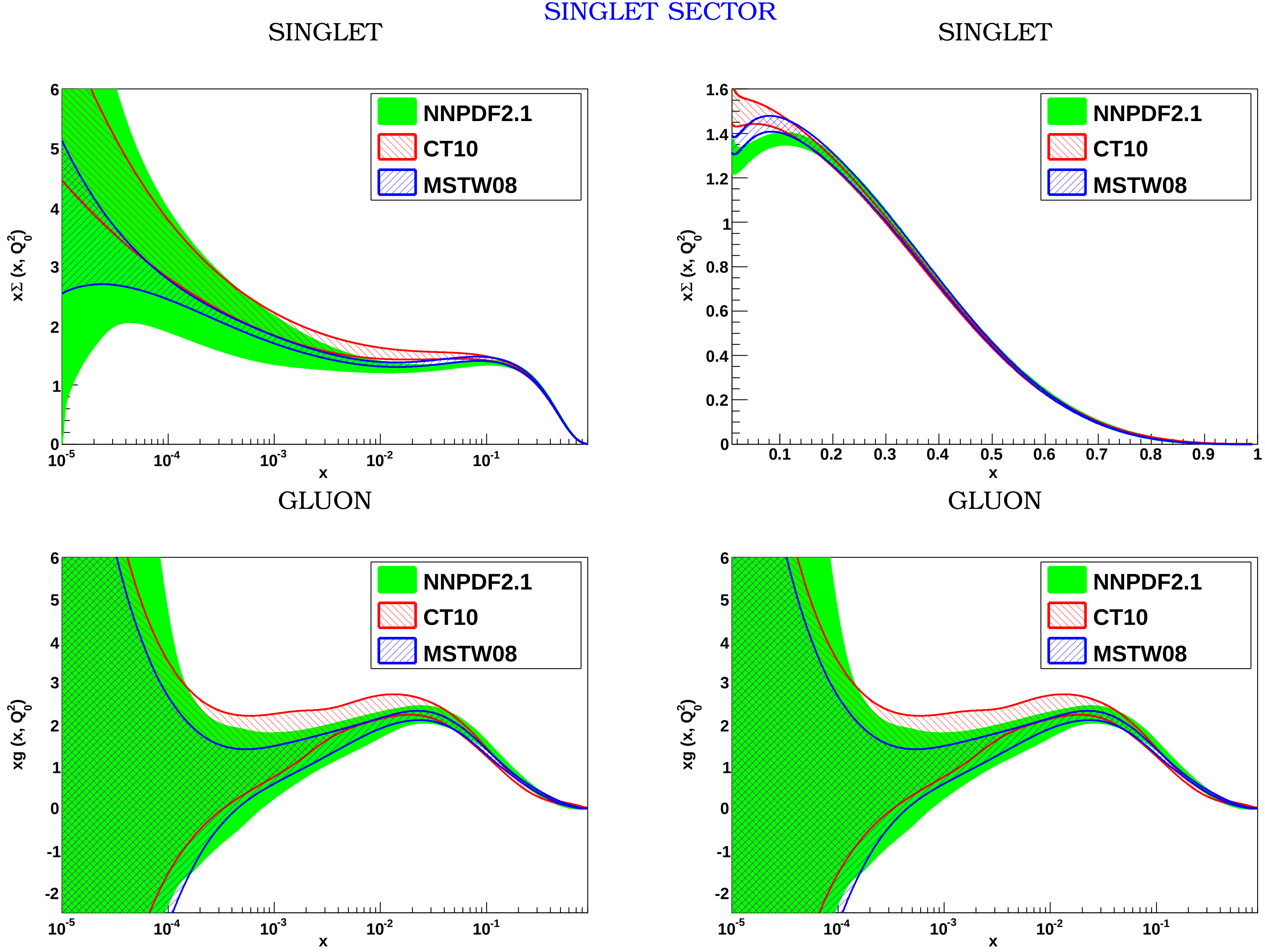
Q: HOW IS IT DONE IN PRACTICE?

A: CHOOSE REPLICAS OF THE DATA, DISTRIBUTED AS THE DATA

NNPDF2.1 PDFs

COMPARED TO OTHER GLOBAL PDF SETS (MSTW08, CT10)

SINGLET SECTOR

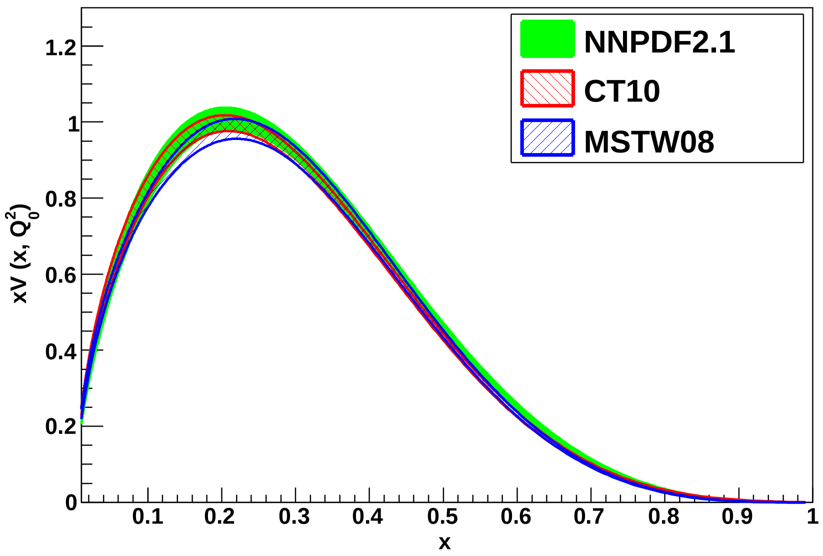


NNPDF2.1 PDFs

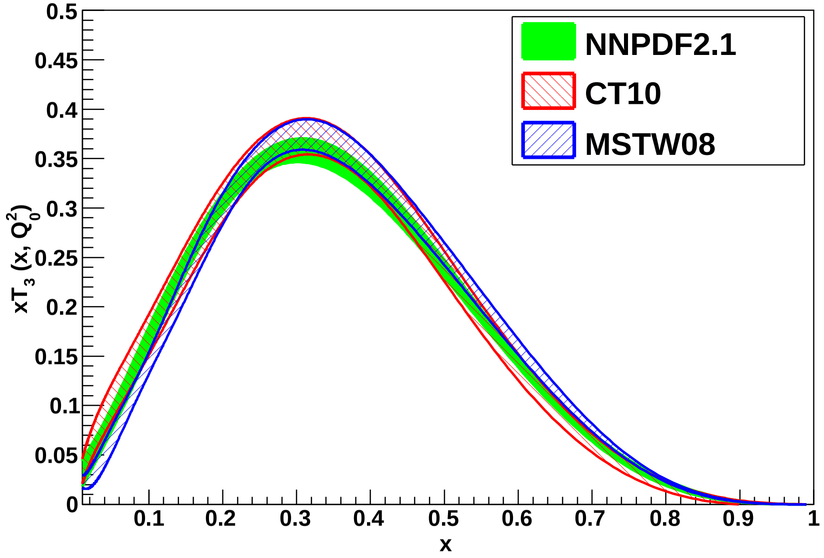
COMPARED TO OTHER GLOBAL PDF SETS (MSTW08, CT10)

NONSINGLET SECTOR

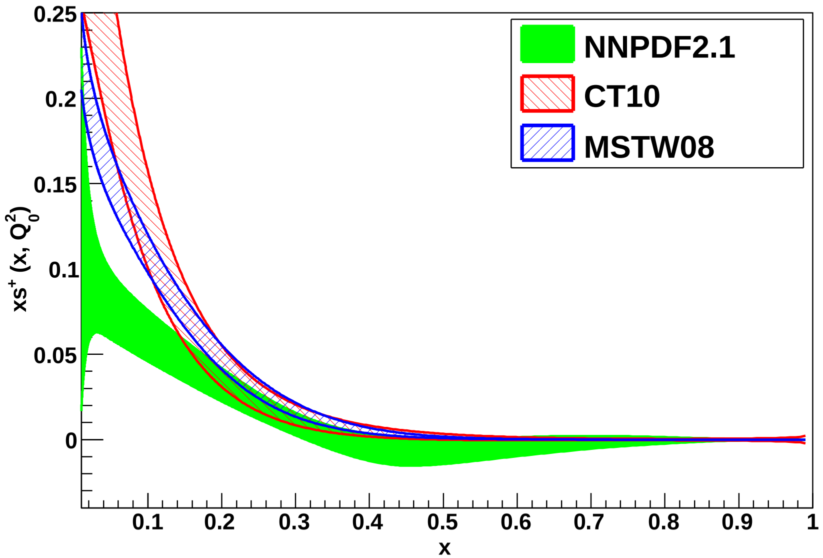
TOTAL VALENCE



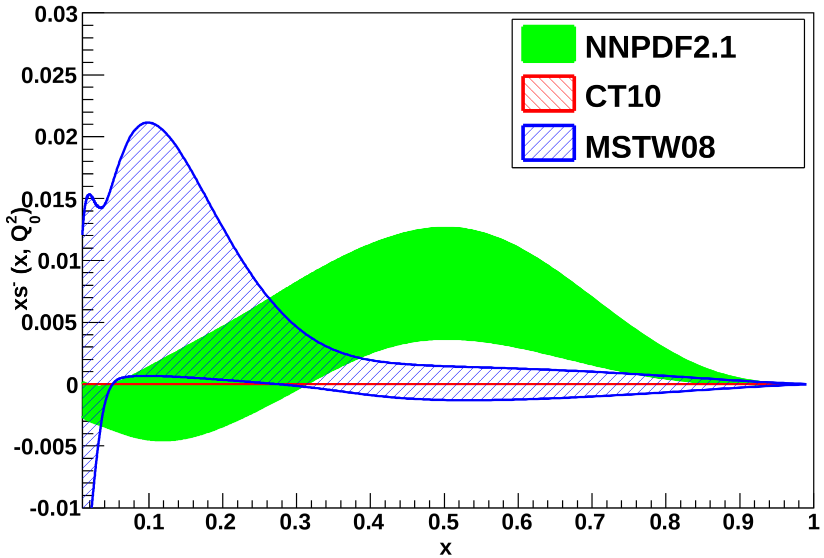
ISOSPIN TRIPLET



$s + \bar{s}$



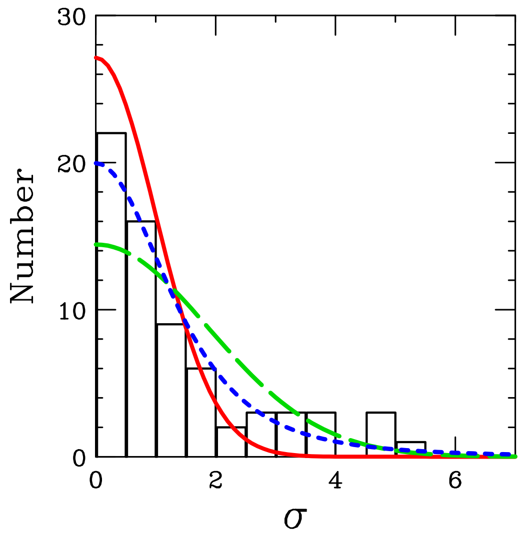
$s - \bar{s}$



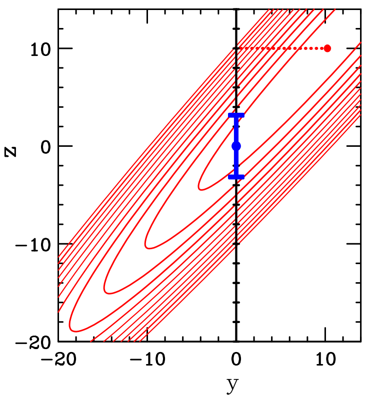
WHERE IS THE UNCERTAINTY COMING FROM?
WHY DOES ONE NEED LARGE TOLERANCES?

DATA INCOMPATIBILITY (Pumplin, 2009)

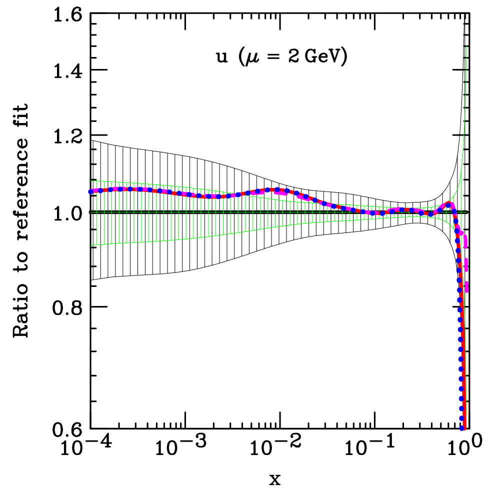
- CAN "REDIAGONALIZE": DIAGONALIZE SIMULTANEOUSLY χ^2 FOR TOTAL AND i–TH EXPT $\Rightarrow$ COMPATIBILITY OF EACH EXPT WITH GLOBAL FIT
- STUDY DISTRIBUTION OF DISCREPANCIES
- APPROX. GAUSSIAN WITH UNCERTAINTIES RESCALED BY $2 \Rightarrow \Delta\chi^2 \sim 10$ FOR 90%C.L.



FUNCTIONAL BIAS (Pumplin, 2010)



- IF PARM. NOT GENERAL ENOUGH, GLOBAL MIN. IS NOT TRUE MIN.
- ONE-σ VARIATION ABOUT FAKE MIN CORRESP. TO LARGE χ^2 VARIATION
- USE OF CHEBYSHEV POLYNOMIALS SUGGESTS "MOST GENERAL" PARM. COMPATIBLE WITH $\Delta\chi^2 = 100$ RANGE OF CT10 PARM.



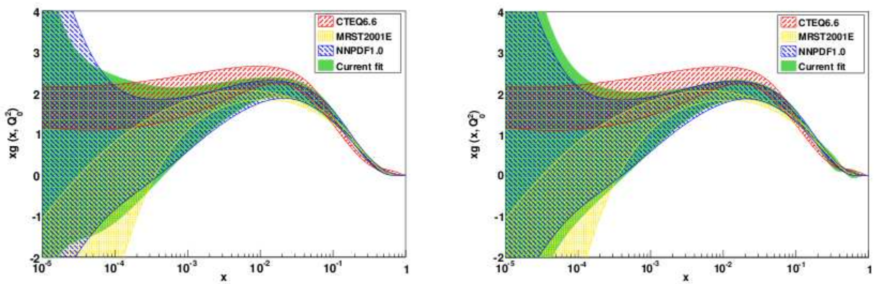
WHERE IS THE UNCERTAINTY COMING FROM?

FIT TO REPLICAS VS RANDOM SUBSET OF CENTRAL VAL.S

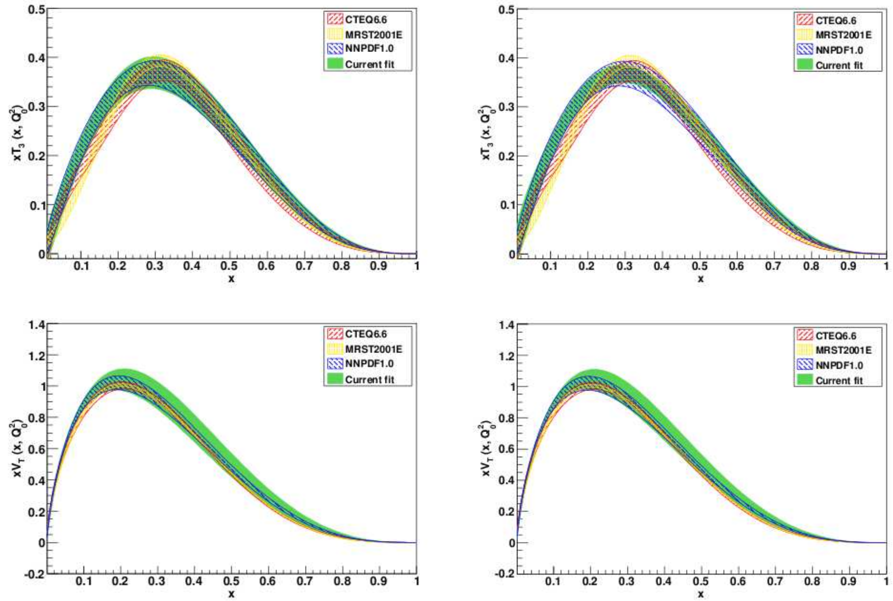
	REPLICAS	CENTRAL V.
χ^2	1.32	1.32
$\langle\chi^2\rangle_{\rm rep}$	2.79 ± 0.24	1.65 ± 0.20
$\langle\sigma^{\rm dat}\rangle$	0.039	0.035

GLUE

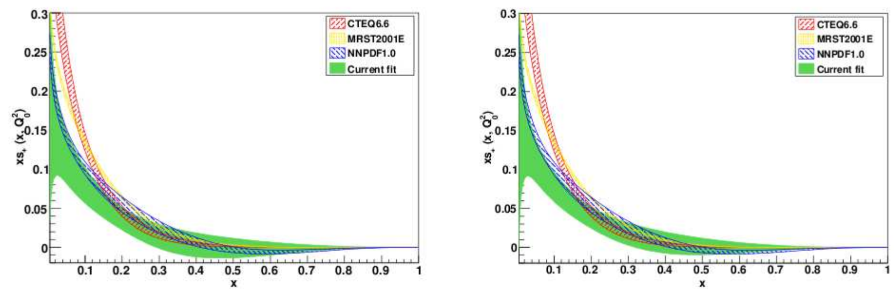
replicas c. vals.



LIGHT QUARKS



STRANGE



- QUALITY OF FIT &PDFS UNCHANGED
- REDUCTION OF $\langle\chi^2\rangle_{\rm rep}$ BY FACTOR $\sim 2 \Rightarrow$ FLUCTUATIONS ABOUT TRUE VALUE HALVED
- UNCERTAINTY ON DATA ONLY REDUCED BY 1.1 $\Rightarrow$ EXPT. UNCERTAINTIES UNDERESTIMATED OR UNDERLYING INCOMPRESSIBLE UNCERTAINTY

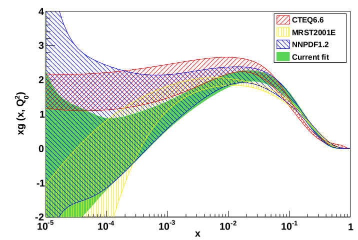
WHERE IS THE UNCERTAINTY COMING FROM?

CENTRAL VALUES: VARYING PARTITION VS FIXED PARTITION

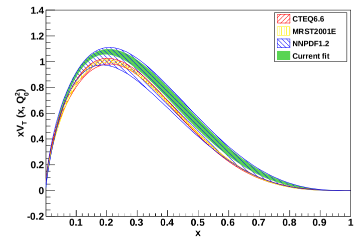
	REPLICAS	CENTRAL VALUE	FIXED PARTITION
χ^2	1.32	1.32	$\sim$1.3
$\langle\chi^2\rangle_{\rm rep}$	2.79 ± 0.24	1.65 ± 0.20	$\sim 1.6 \pm 0.2$
$\langle\sigma^{\rm dat}\rangle$	0.039	0.035	$\sim$0.03

fixed partition results obtained averaging over 5 different choices of partition (100 replicas each); more partitions needed for accurate results

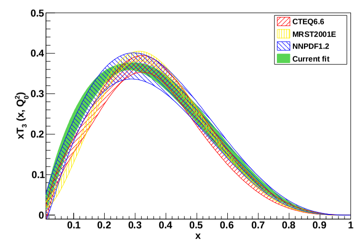
GLUE



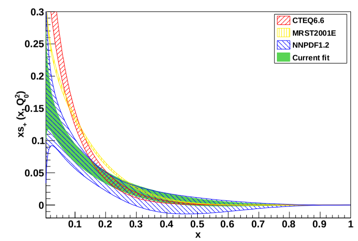
VALENCE



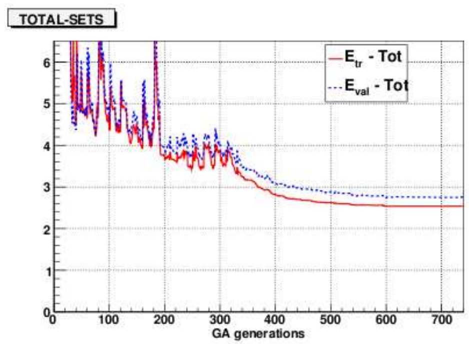
TRIPLET

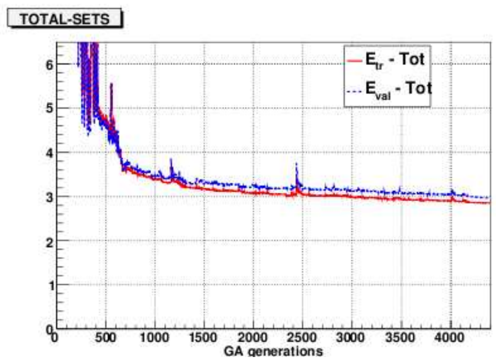


STRANGE



- QUALITY OF FIT UNCHANGED
- $\langle\chi^2\rangle_{\rm rep}$ UNCHANGED $\Rightarrow$ CENTRAL FIT UNCHANGED
- UNCERTAINTY ON PREDICTION (I.E. ON PDFS) REDUCED

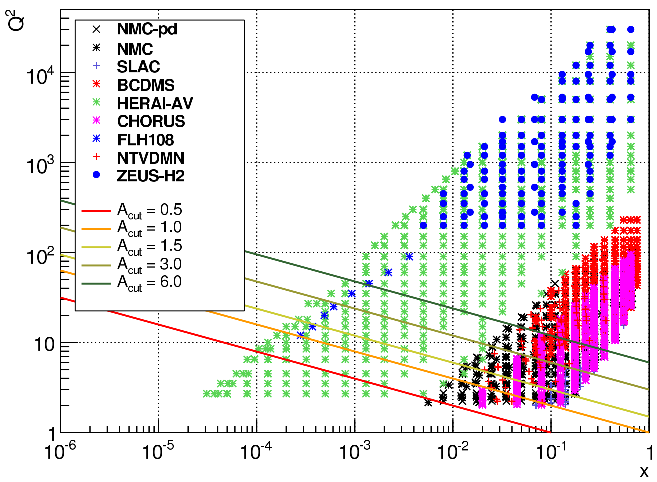




FUNCTIONAL UNCERTAINTY

- MORE THAN HALF OF UNCERTAINTY DUE TO "FUNCTIONAL FORM": $\langle\sigma^{\rm dat}\rangle =\sim 0.03$ SMALLER FOR HERA DATA
- REMAINING UNCERTAINTY ROUGHLY SCALES WITH DATA UNCERTAINTY: $\langle\sigma^{\rm dat}\rangle =\sim 0.005$ CENT.; $\langle\sigma^{\rm dat}\rangle =\sim 0.009$ REP.

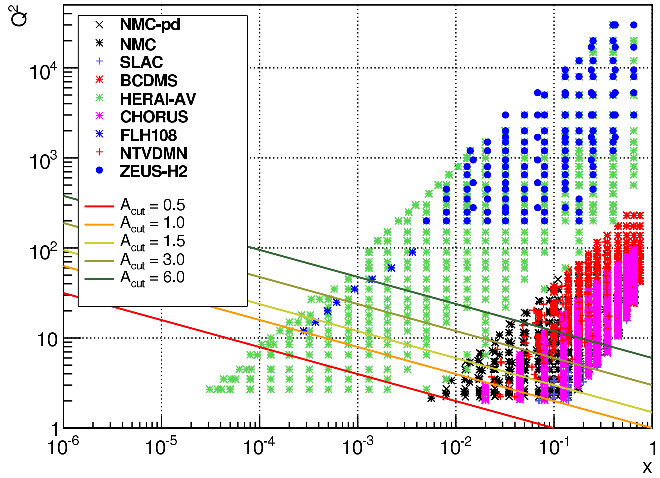
BEYOND DGLAP?
DISCOVERING A NEW QCD EFFECT IN HERA DATA



IDEA: (Gélis, 2008, $\Rightarrow$ Caola, s.f. ,Rojo 2010)

- CUT OUT DATA IN THE "DANGEROUS" (SMALL x) REGION
- DETERMINE PDFs IN THE "SAFE" (LARGE x AND Q^2) REGION
- EVOLVE BACKWARDS AND COMPARE TO DATA

BEYOND DGLAP?
DISCOVERING A NEW QCD EFFECT IN HERA DATA

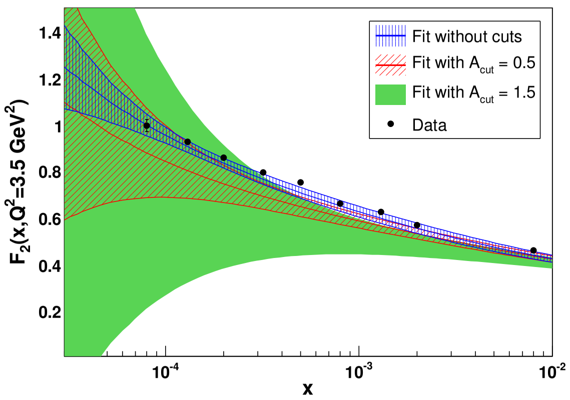


IDEA: (Gélis, 2008, $\Rightarrow$ Caola, s.f. ,Rojo 2010)

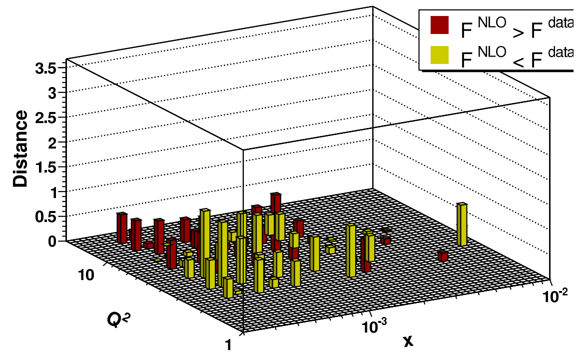
- CUT OUT DATA IN THE "DANGEROUS" (SMALL x) REGION
- DETERMINE PDFs IN THE "SAFE" (LARGE x AND Q^2) REGION
- EVOLVE BACKWARDS AND COMPARE TO DATA

OLD HERA DATA

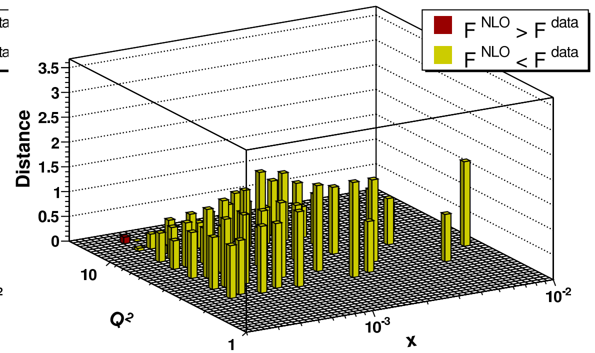
BACKWARD EV. VS DATA



DAT/TH DIST: NO CUT



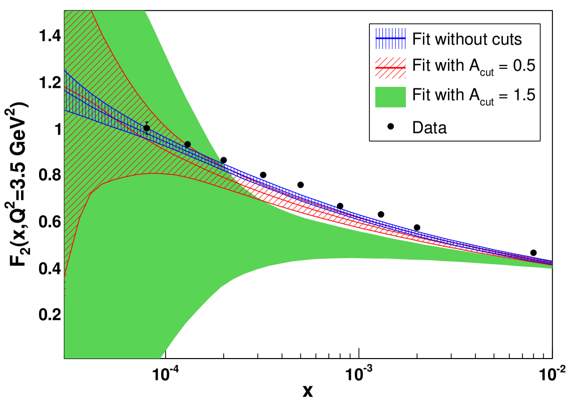
DAT/TH DIST: CUT



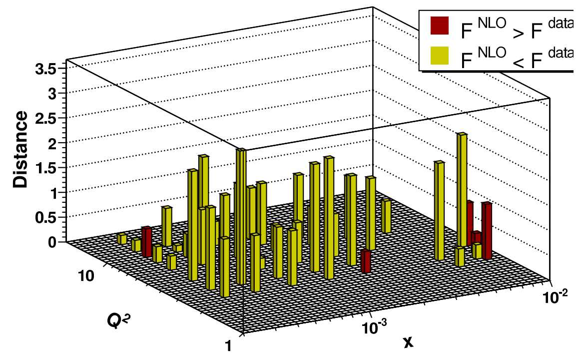
- BACKWARD EVOLVED FIT LIES SYSTEMATICALLY BELOW DATA
- DATA AT LOW x AND Q^2 SHOW LESS EVOLUTION THAN PREDICTED BY NLO DGLAP
- IF LOW x AND Q^2 DATA INCLUDED, THE FIT COMPENSATES READJUSTING PDFs

NEW COMBINED HERA DATA

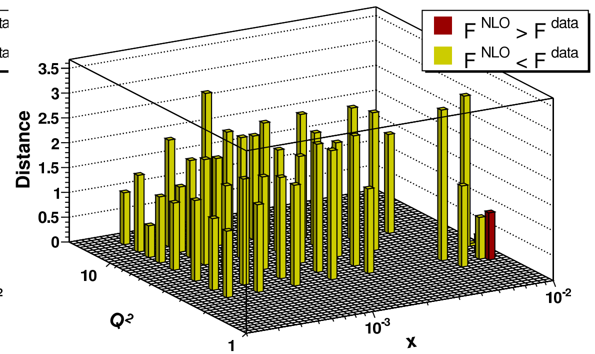
BACKWARD EV. VS DATA



DAT/TH DIST: NO CUT



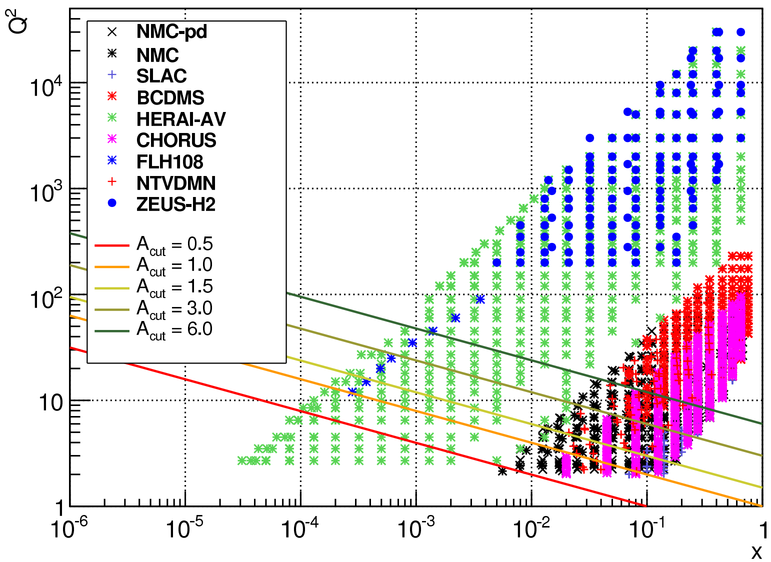
DAT/TH DIST: CUT

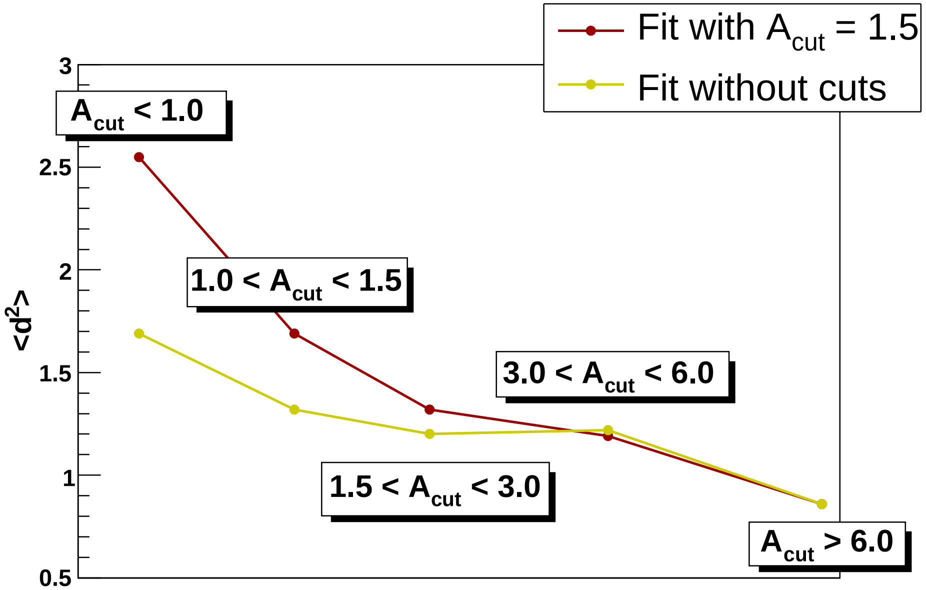


- DATA AT LOW x AND Q^2 SHOW LESS EVOLUTION THAN PREDICTED BY NLO DGLAP
- BACKWARD EVOLVED FIT LIES SYSTEMATICALLY BELOW DATA
- WITH MORE PRECISE DATA, THE FIT NO LONGER MANAGES TO COMPENSATE BY READJUSTING THE PDFs: EVEN FULL FIT LIES BELOW DATA

DETERIORATION IN FIT QUALITY:

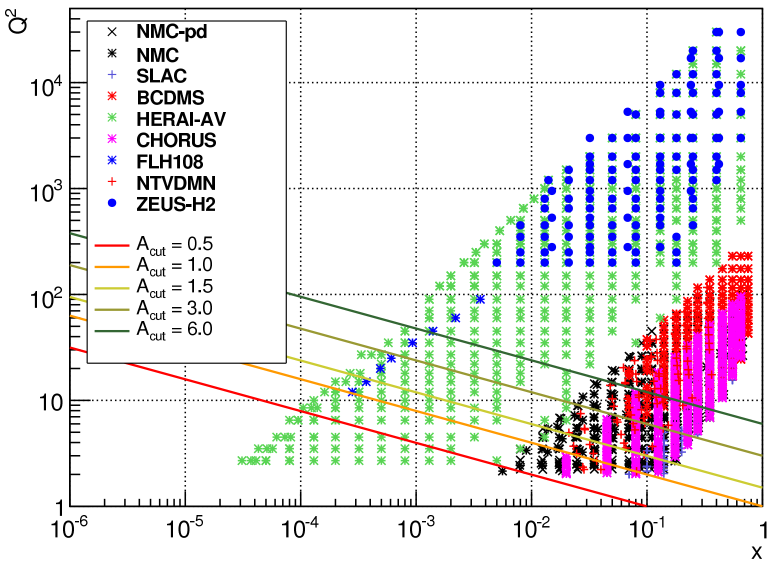
$\langle d^2 \rangle$ VS x SLICES



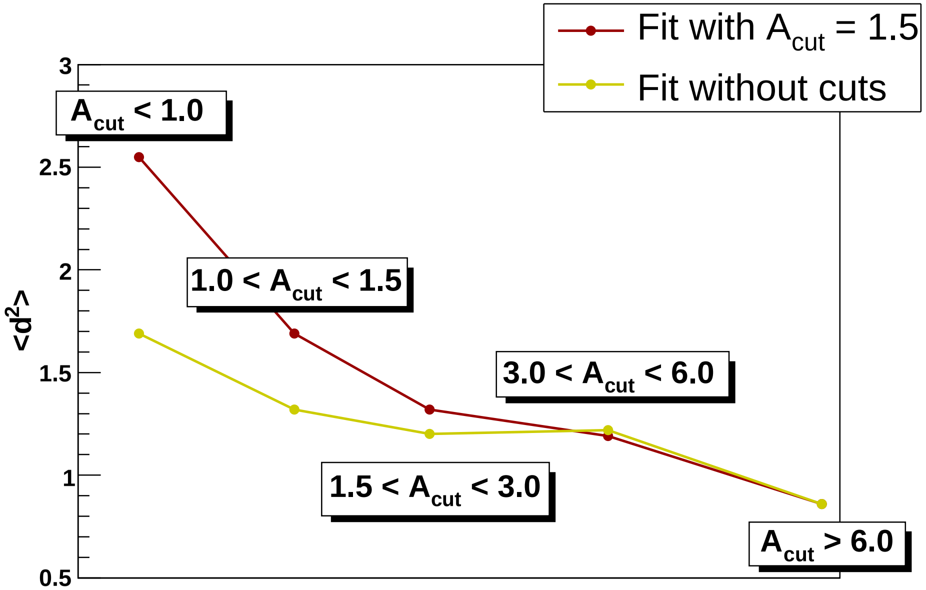


- QUALITY OF UNCUT FIT DETERIORATES IN LOW x REGIONS
- QUALITY OF CUT FIT INCREASINGLY POOR AS x DECREASES
- DISTANCE RISES DESPITE HUGE INCREASE IN UNCERTAINTY

DETERIORATION IN FIT QUALITY:

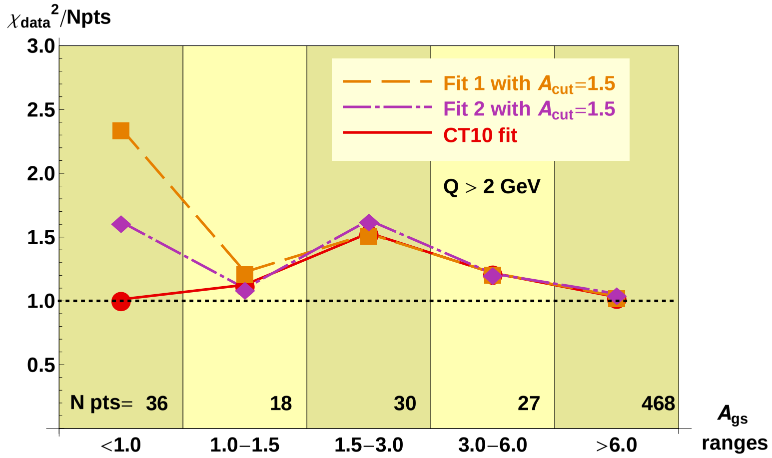


$\langle d^2 \rangle$ VS x SLICES

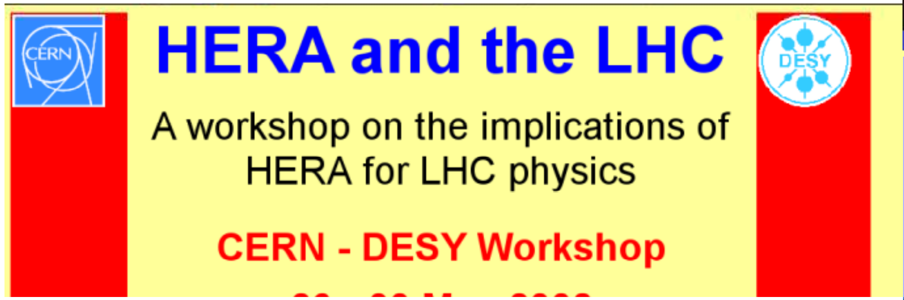


- QUALITY OF UNCUT FIT DETERIORATES IN LOW x REGIONS
- QUALITY OF CUT FIT INCREASINGLY POOR AS x DECREASES
- DISTANCE RISES DESPITE HUGE INCREASE IN UNCERTAINTY
- IN HESSIAN FIT (CTEQ) RESULTS DEPEND ON PARAMETRIZATION $\Rightarrow$ EVIDENCE INCONCLUSIVE

χ^2 VS x SLICES FOR CT10



AN ONGOING EFFORT



- HERALHC (2004-2008): A WORKSHOP TO TRANSFER KNOWN-HOW FROM HERA TO THE LHC COMMUNITY
  - two sessions (2004-2005; 2006-2007), four plenary meetings, five mid-term and working group meetings
  - two CERN/DESY reports (yellow books): **HEP-PH/0601012-HEP-PH/0601013**; **ARXIV:0903.3861**.
- PDF4LHC (2008 - ONGOING): A PERMANENT WORKING GROUP TO PROVIDE GUIDANCE ON PDF TO LHC EXPERIMENTS AND PHENOMENOLOGY
  - quarterly meetings, 12 since inception in February 2008
  - website `http://www.hep.ucl.ac.uk/pdf4lhc/` and wiki `https://wiki.terascale.de/index.php?title=PDF4LHC_WIKI` resources available
  - fruitful interactions with the LHC Higgs Cross Section WG