PROTON STRUCTURE FOR LHC DISCOVERIES

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

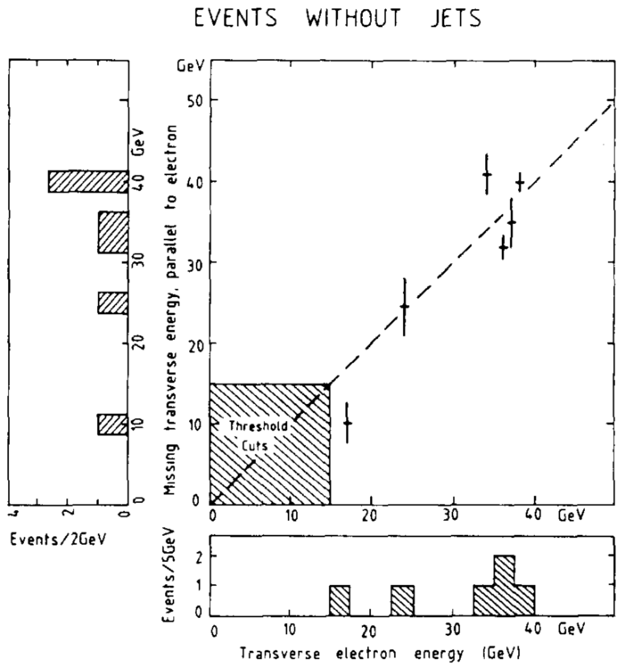
INFN Istituto Nazionale di Fisica Nucleare

HIGGS-MAXWELL MEETING

EDINBURGH, FEBRUARY 8, 2012

DISCOVERY AT THE S$p\bar{p}$S AND AT THE LHC

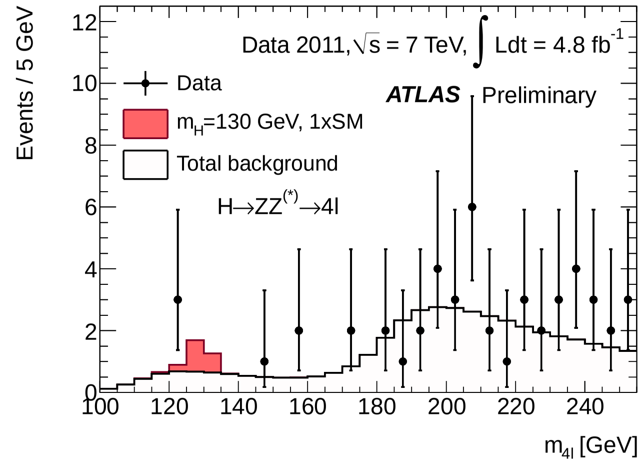
DISCOVERY OF THE W (1984)



p_T distribution of the e and ν

from W decay at UA1 (1984)

DISCOVERY OF THE HIGGS (2011?)



transverse energy distribution

in the four lepton channel at ATLAS (2011)

- HADRON COLLIDER CIRCA 1985 $\Rightarrow$ “CLEAN” DISCOVERY SIGNAL
- HADRON COLLIDER CIRCA 2010 $\Rightarrow$ PRECISION PHYSICS $\leftrightarrow$ NEW PHYSICS

PDFs FROM THE S$p\bar{p}$S TO THE LHC

PDFs IN 1984

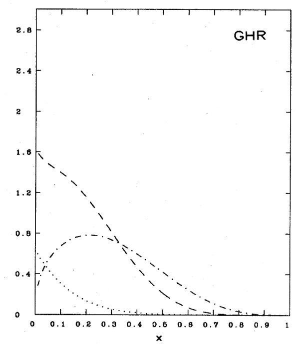


FIG. 25. Parton distributions of Glück, Hoffmann, and Reya (1982), at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and q_v (dotted line).

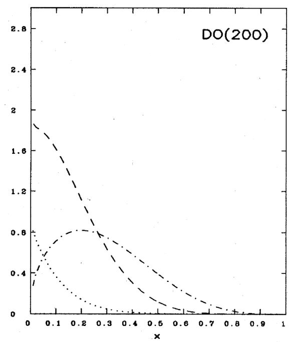


FIG. 27. "Soft-gluon" ($\Lambda=200$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

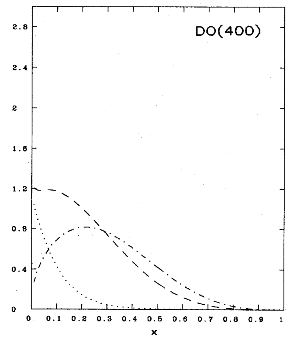


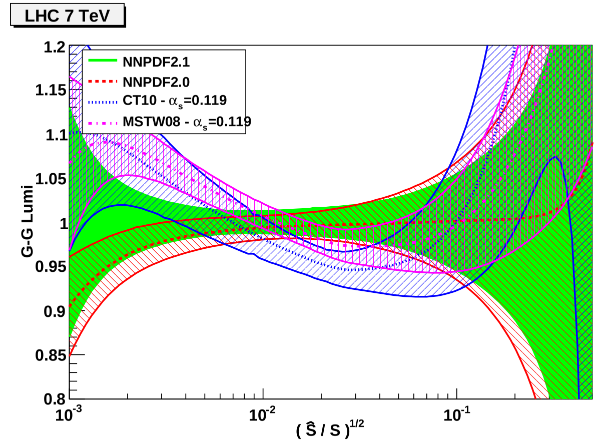
FIG. 26. "Hard-gluon" ($\Lambda=400$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

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

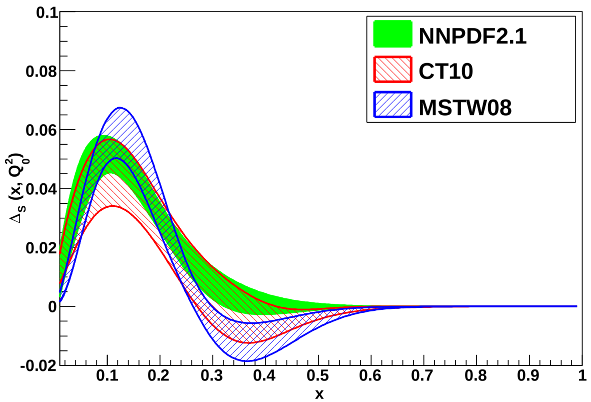
GHR VS DUKE-OWENS

PDFs IN 2011

GLUON-GLUON LUMI



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



NNPDF VS MSTW VS CTEQ, 2011

- 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

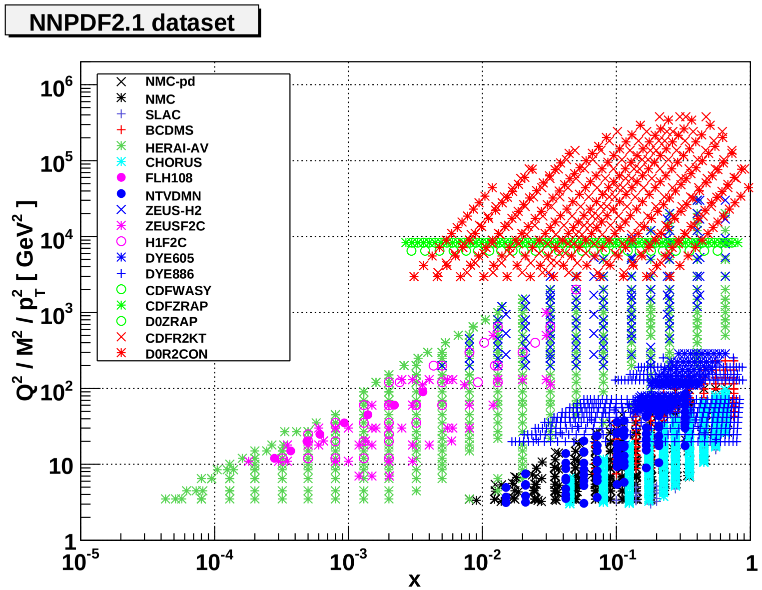
- PROTON STRUCTURE TODAY
  - PDFS, LUMINOSITIES, STANDARD CANDLES
  - RECENT PROGRESS AT NLO AND NNLO
- ISSUES
  - THE MEANING OF PDF UNCERTAINTIES
  - DATA CONSISTENCY
- THE IMPACT OF HADRON COLLIDER DATA
  - JETS AND THE GLUON
  - FLAVOR SEPARATION: DRELL-YAN, W AND Z
- PRECISION PHYSICS AT THE LHC
  - COLLIDER ONLY PDFS
  - THE ROLE OF THE LHC DATA
- A ROAD MAP FOR THE FUTURE

THE STATE OF THE ART

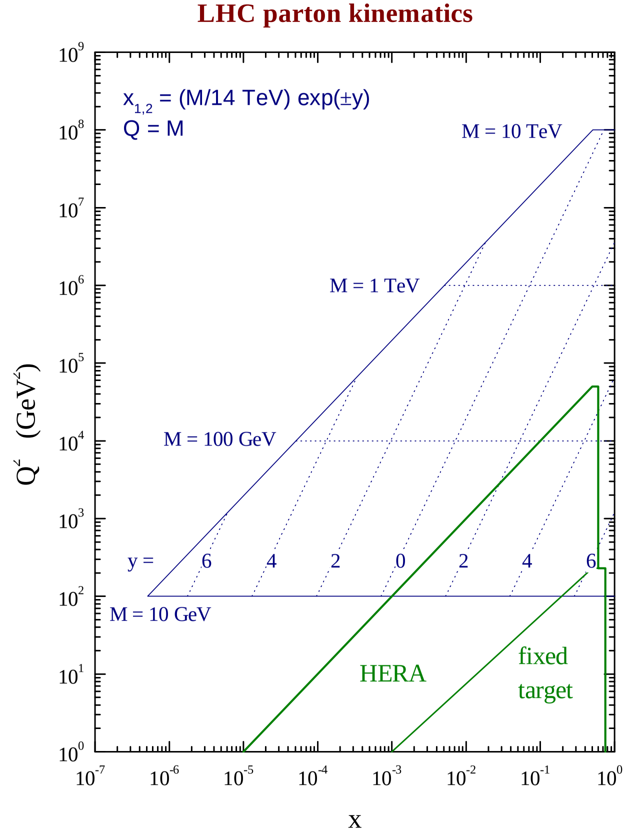
CURRENT PDF SETS

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

data in global fits



LHC kinematics



- CT10: GLOBAL, LO+NLO, SEVERAL α_s VALUES, VFN
- MSTW08: GLOBAL, LO+NLO+NNLO, SEVERAL α_s VALUES, VFN
- NNPDF2.1-2.2: GLOBAL (INCL. LHC), LO+NLO+NNLO, SEVERAL α_s VALS., VFN
- ALEKHIN ABKM: DIS+SOME DY, NLO+NNLO, SINGLE α_s VALUE, FFN
- HERAPDF1.0 (1.5?): HERA DIS, NLO+NNLO, SEVERAL α_s VALS, VFN
- SETS BASED ON MODEL ASSUMPTIONS (GRV/GJR, STATISTICAL PDFS,...)

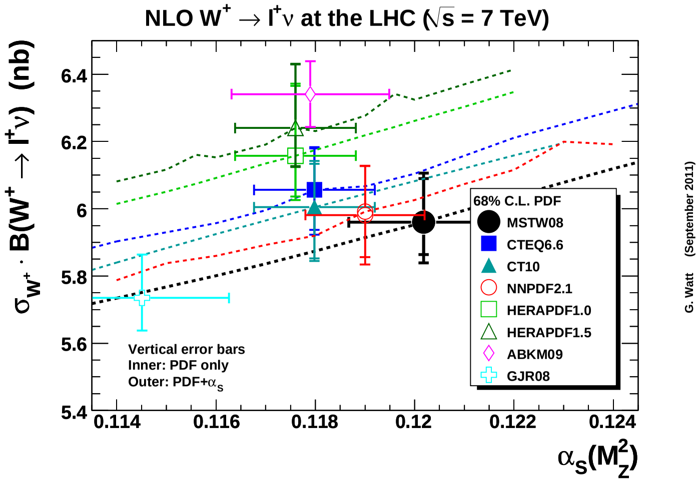
FEATURES, TRADEOFFS AND CHOICES

- DATASET: CTEQ, MSTW, NNPDF FIXED TARGET + COLLIDER, eP AND $\bar{P}P$ DATA (NNPDF: LHC); ABKM, GJR GLOBAL DIS+ FIXET-TARGET DY; HERAPDF: HERA DIS 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, MSTW)
- 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: LO+NLO; MSTW: LO+NLO+NNLO; NNPDF LO+NLO+NNLO; ABKM, HERAPDF, GJR: NLO+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 ($N_f = 3$)

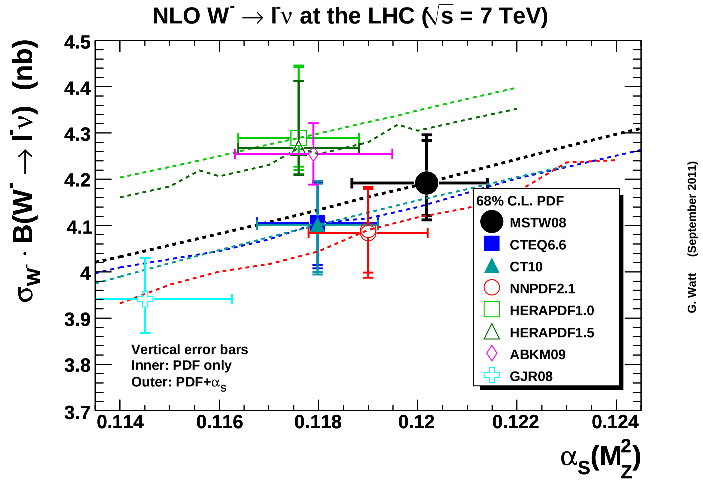
WHERE DO WE STAND?

LHC STANDARD CANDLES (NLO)

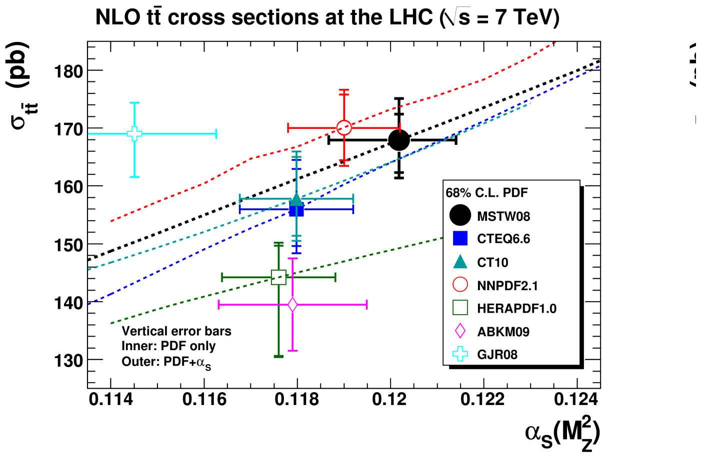
W^+



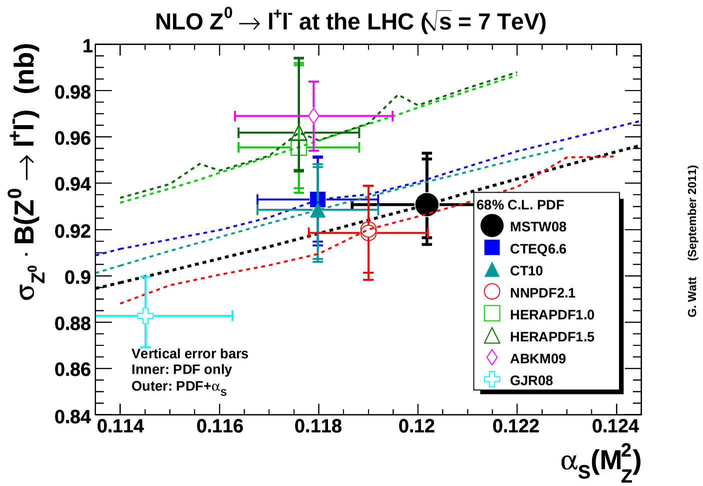
W^-



TOP



Z



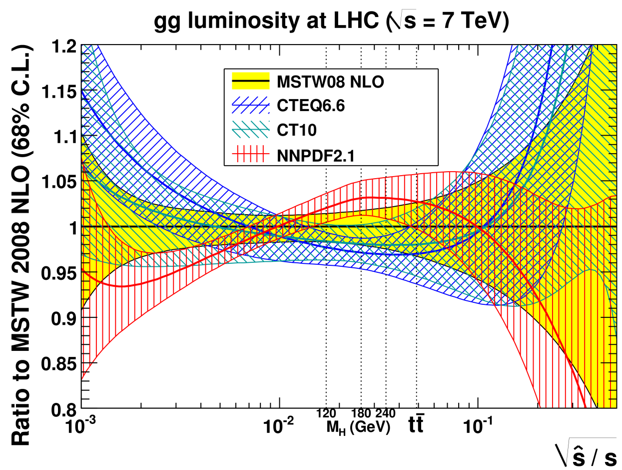
(G. Watt, 2011)

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

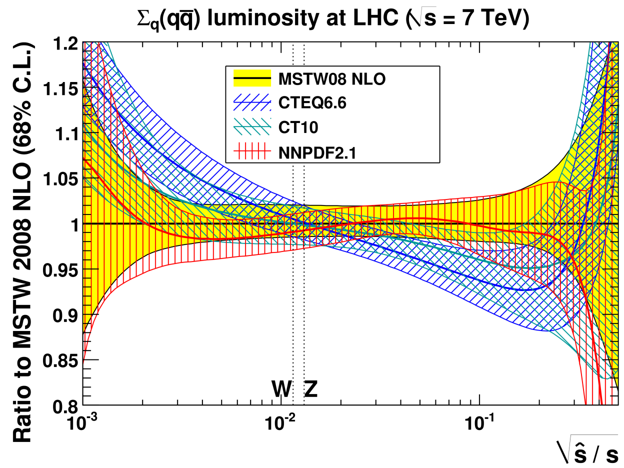
WHERE DO WE STAND?

PARTON LUMINOSITIES (NLO)

GLUON-GLUON

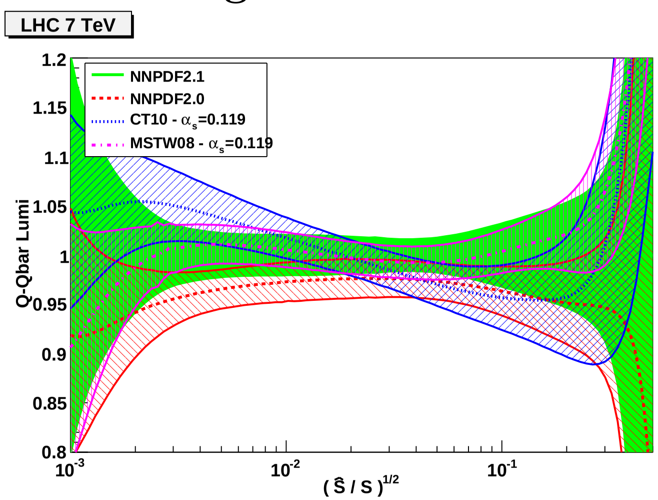


QUARK-QUARK

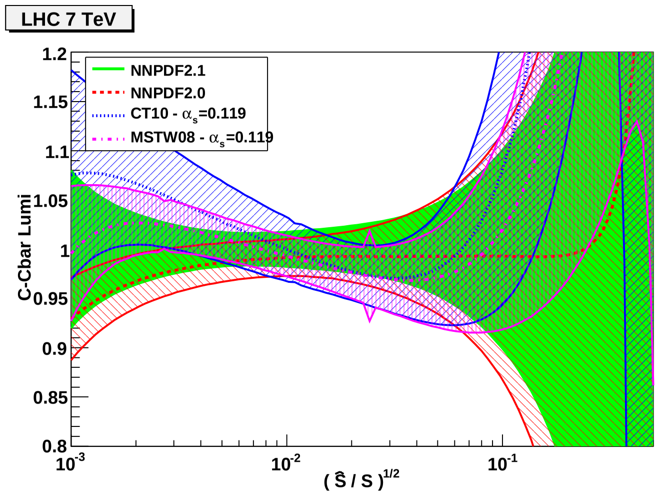


(G. Watt, 2011)

QUARK-GLUON



CHARM-CHARM



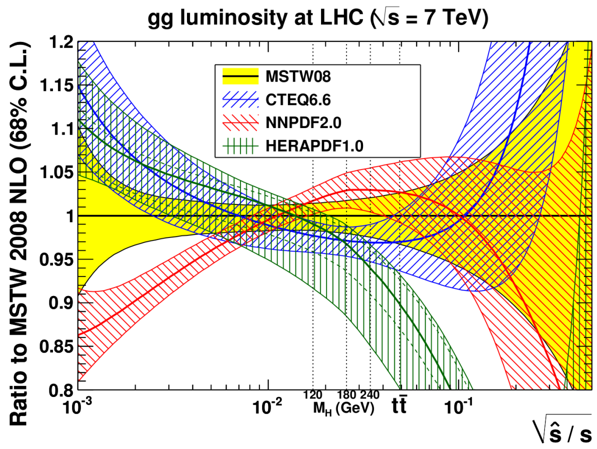
(NNPDF, 2011)

CONTINUOUS PROGRESS

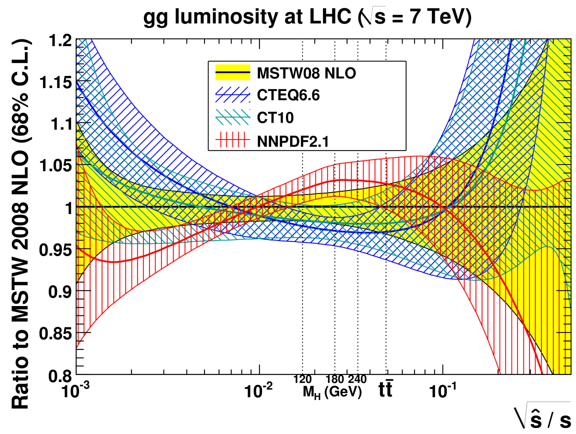
PARTON LUMINOSITIES (NLO)

GLUON-GLUON

2010

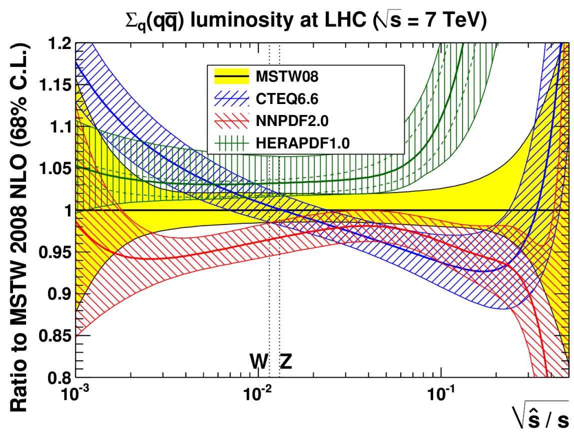


2011

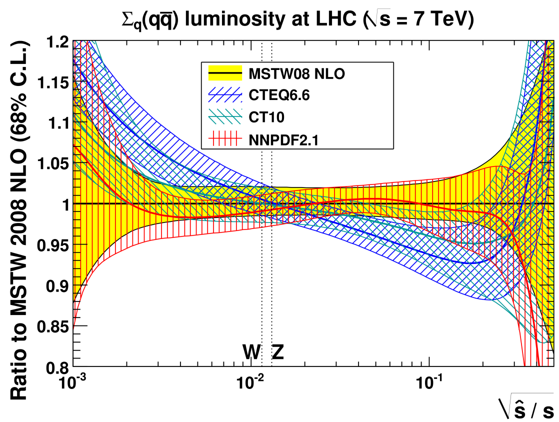


QUARK-QUARK

2010



2011



(G. Watt, 2011)

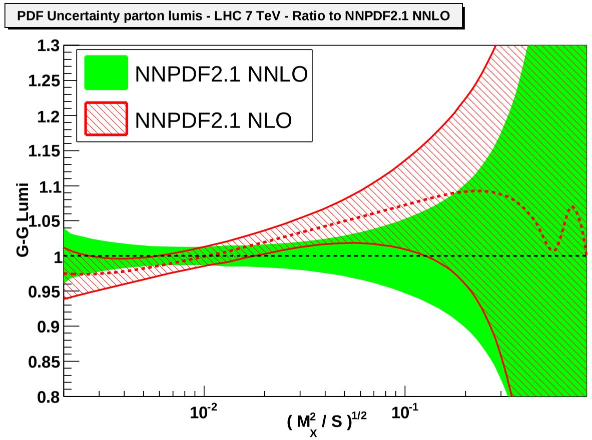
- NNPDF2.0 → NNPDF2.1: INCLUSION OF HW MASSES
- CTEQ6.6 → CT10: IMPROVED STATISTICS, PARAMETRIZATION, DATA
- CONSIDERABLY IMPROVED AGREEMENT!

CONTINUOUS PROGRESS

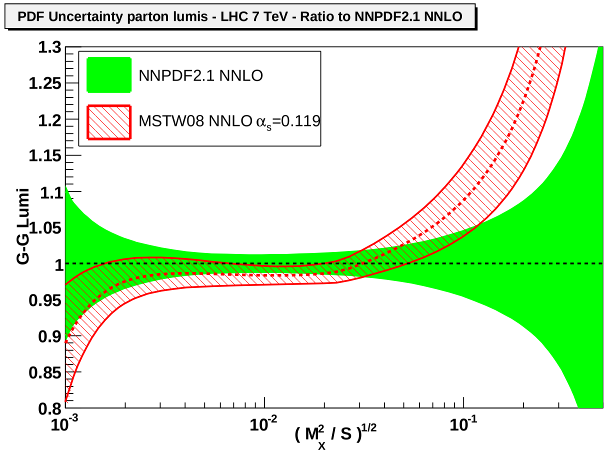
PARTON LUMINOSITIES (NNLO)

GLUON-GLUON

NNPDF NLO vs. NNLO

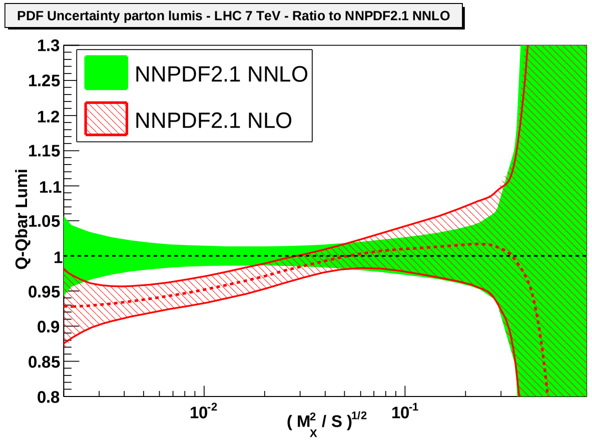


NNLO NNPDF2.1 vs. MSTW08

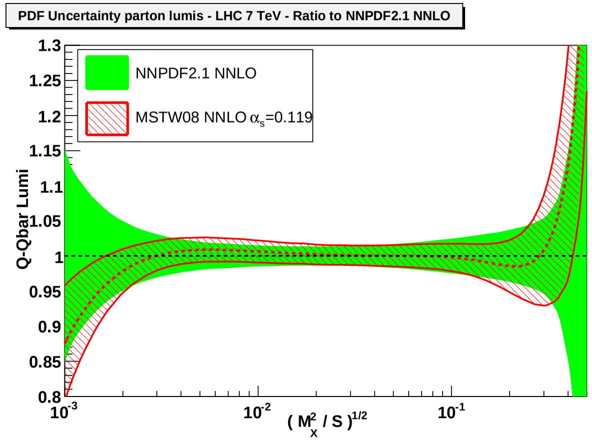


QUARK-QUARK

NNPDF NLO vs. NNLO



NNLO NNPDF2.1 vs. MSTW08



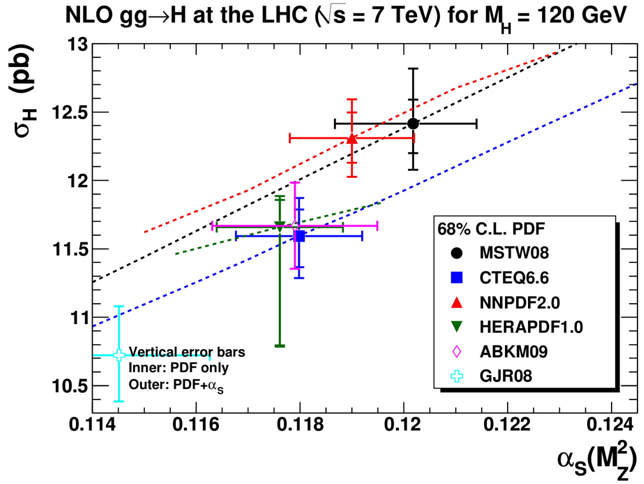
(NNPDF, 2011)

- NNPDF2.1 NLO VS NNLO $\rightarrow$ STABILITY
- NNPDF2.1 VS MSTW08 $\rightarrow$ GOOD AGREEMENT IN EW $M_x \sim 100$ GEV REGION, BUT MSTW GLUON UNSTABLE AT SMALL x

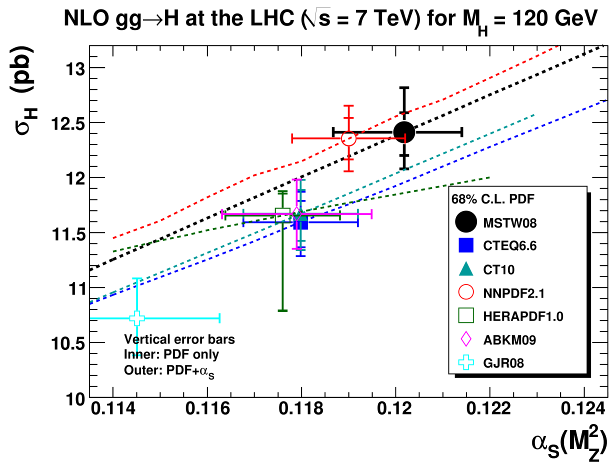
CONTINUOUS PROGRESS

THE HIGGS CROSS SECTION

$m_H = 120$ GeV: 2010

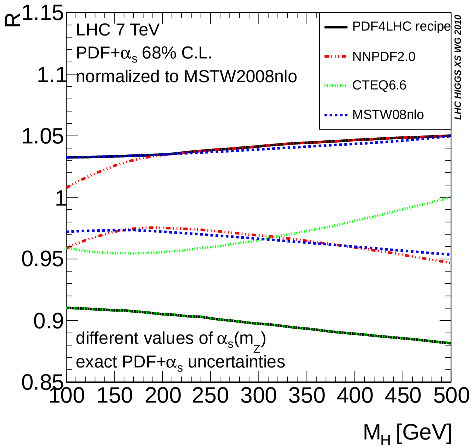


$m_H = 120$ GeV: 2011

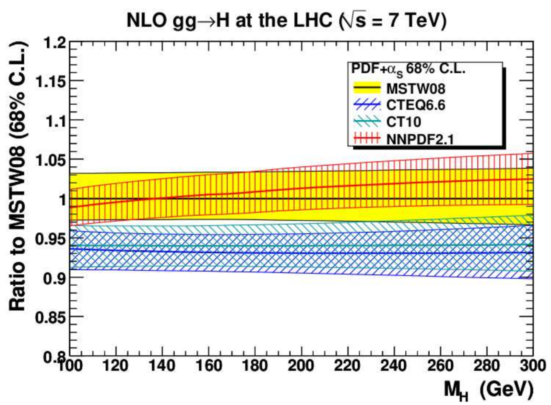


(G. Watt)

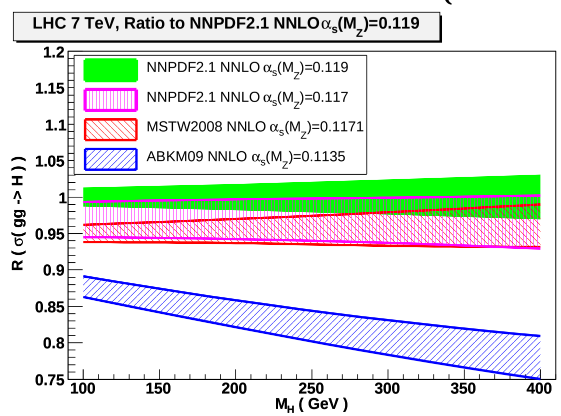
HIGGS WG NLO 2010



NLO 2011 ENVELOPE (WATT)



NNLO 2011 ENVELOPE (NNPDF)



- NLO: 2010 $\to$ 2011 IMPROVED AGREEMENT OF GLOBAL FITS
- NNLO: MSTW-NNPDF AGREEMENT ALMOST PERFECT

ISSUES

PDF UNCERTAINTIES:
DO THEY HAVE A STATISTICAL MEANING?

SOMETIMES STATED THAT "PDF UNCERTAINTIES ARE THEORETICAL UNCERTAINTIES" (THUS DEVOID OF STATISTICAL MEANING) IS IT TRUE?

PDF UNCERTAINTIES: DO THEY HAVE A STATISTICAL MEANING?

SOMETIMES STATED THAT "PDF UNCERTAINTIES ARE THEORETICAL UNCERTAINTIES" (THUS DEVOID OF STATISTICAL MEANING) IS IT TRUE?

A WAY OF TESTING:
NEW DATA ⇒ 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...

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

⇒ EFFECT OF NEW DATA IS ACCOUNTED FOR BY
REWEIGHTING MONTE CARLO AVERAGES

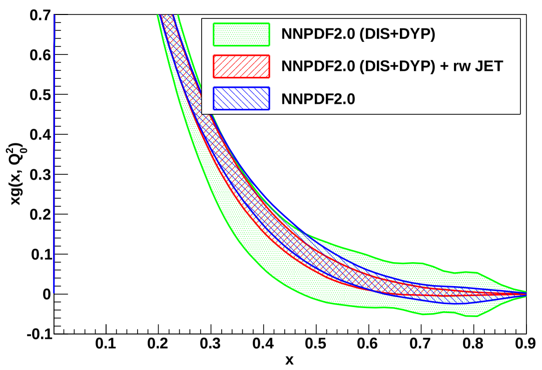
(CAN ALSO BE IMPLEMENTED IN HESSIAN FRAMEWORK (de Lorenzi, McNulty, 2010-2011))

- DETERMINE PDFS INCLUDING SOME DATA BY BAYES' THEOREM (REWEIGHTING)
- DETERMINE PDFS BY ENLARGING THE DATASET TO THE NEW DATA (REFITTING)
- COMPARE RESULTS ⇒ IF THEY AGREE, PDFS DO HAVE A STATISTICAL INTERPRETATIONS ALRIGHT!

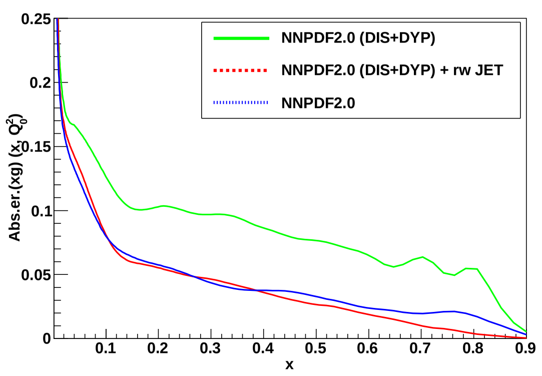
RUNNING THE TEST

INCLUSION OF JET DATA: REWEIGHTING VS. REFITTING
NNPDF2.0DIS+DY VS. NNPDF2.0FULL

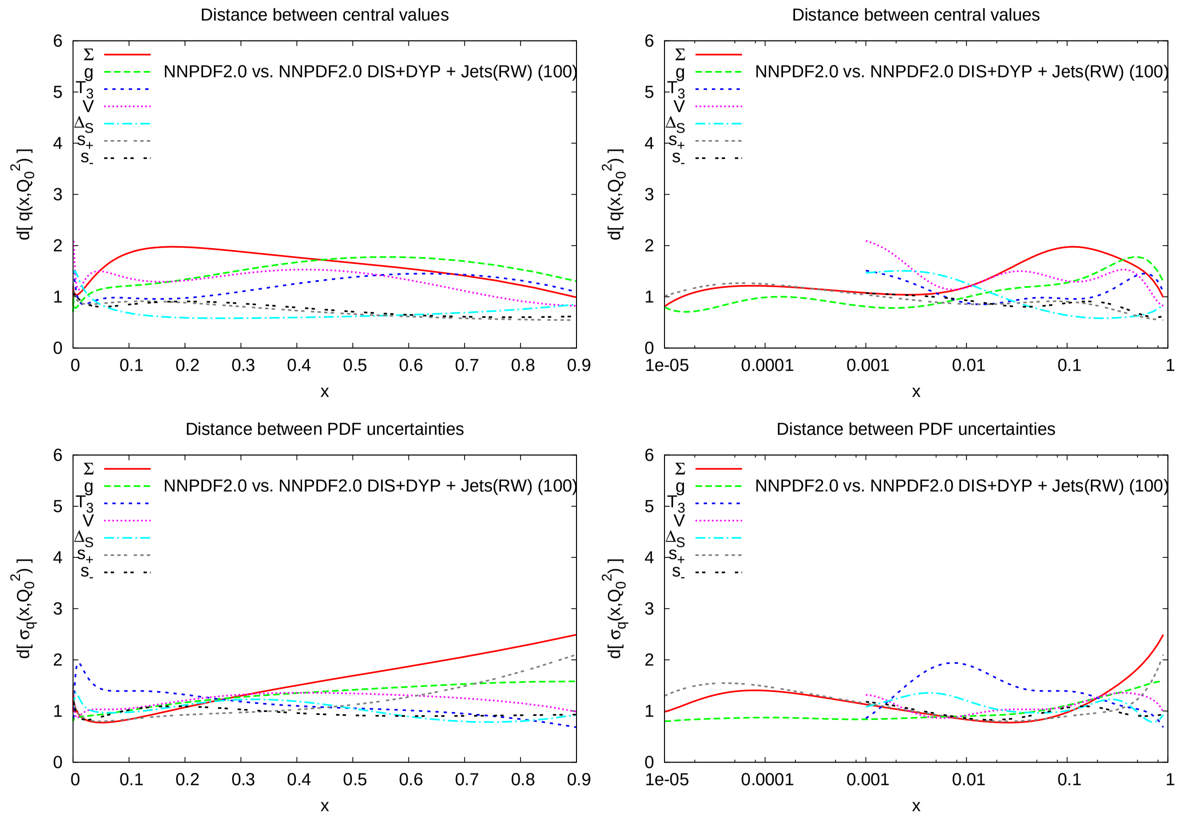
GLUON



GLUON UNCERTAINTY



DISTANCES

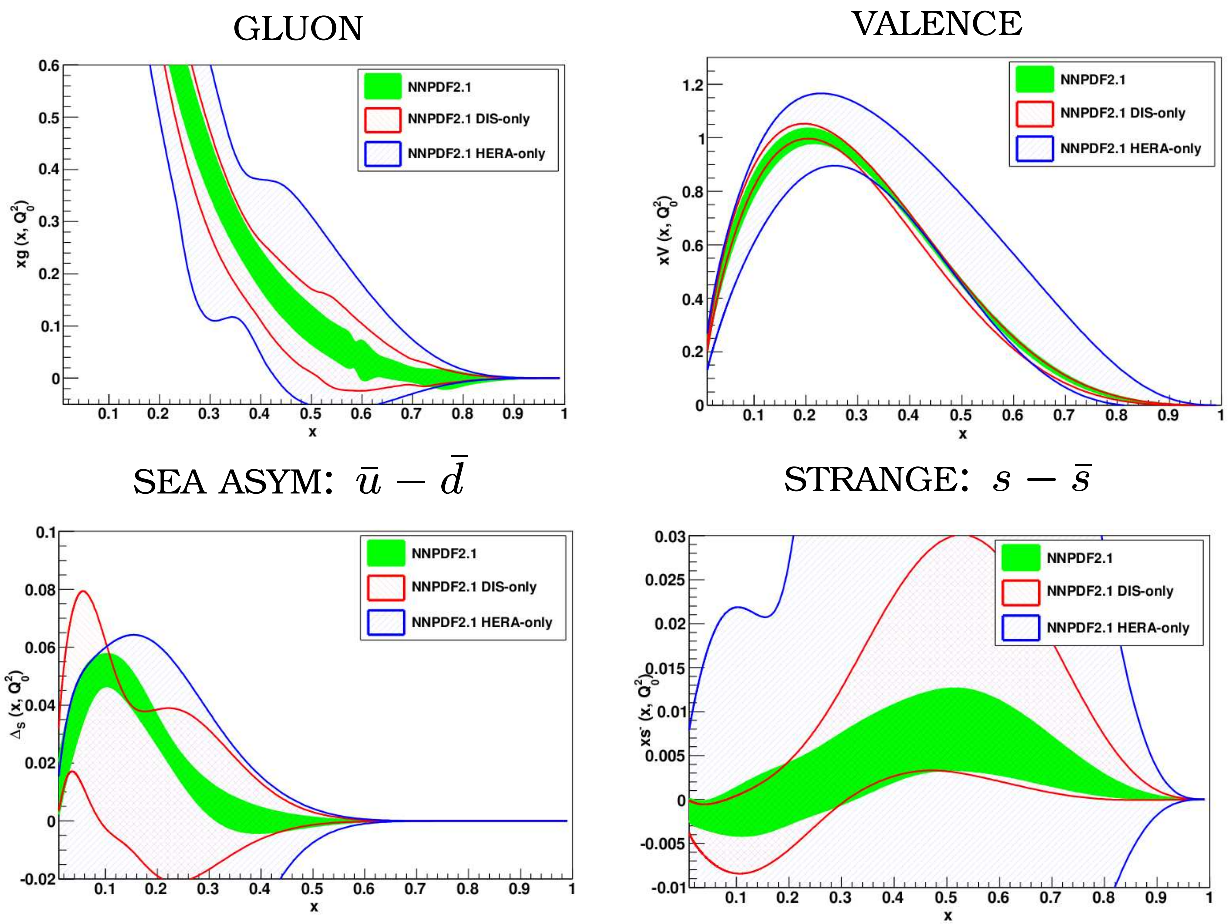


$d \sim 1 \Rightarrow$ STATISTICAL EQUIVALENCE
($d = n \Leftrightarrow n\ \sigma$ DISCREPANCY)

STATISTICAL INTERPRETATION VALIDATED

DATA CONSISTENCY: WHY WE NEED IT

- DIS ONLY FIT $\Rightarrow$ SIGNIFICANT LOSS IN ACCURACY
- LARGE x GLUON: NEED JETS!
- LIGHT AND STRANGE FLAVOUR SEPARATION: NEED DRELL-YAN AND W PRODUCTION

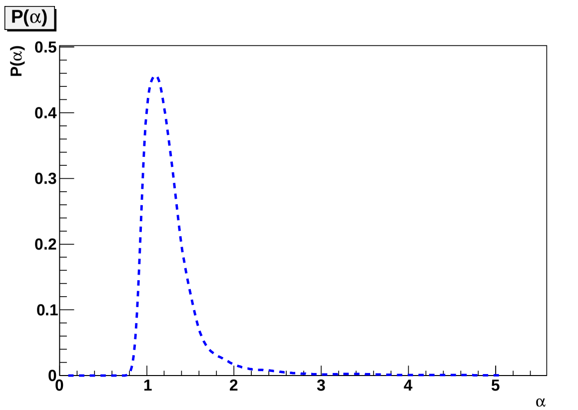


A HERA ONLY FIT IS A TERRIBLE IDEA!

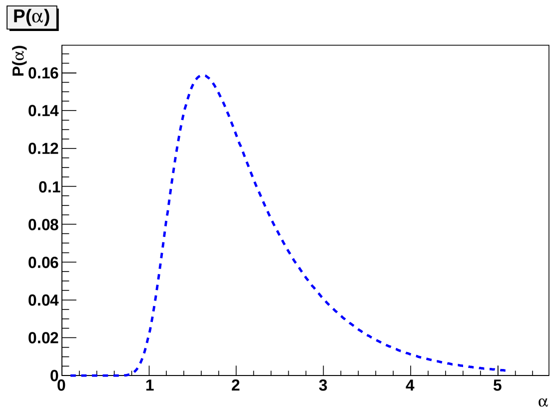
DATA CONSISTENCY: HOW TO TEST FOR IT

- INCONSISTENT DATA $\Leftrightarrow$ UNDERESTIMATED UNCERTAINTIES
- RESCALE ALL UNCERTAINTIES IN A GIVEN EXPERIMENT BY SOME FACTOR α: $\chi^2_\alpha = \chi^2/\alpha^2$ (TOLERANCE)
- DETERMINE PROBABILITY DISTRIBUTION OF α VALUES BY BAYES' THEOREM $\Rightarrow$ REWEIGHTING: $\mathcal{P}(\alpha) = \frac{\mathcal{N}}{\alpha} \sum_{k=1}^{N} w_k(\alpha)$.

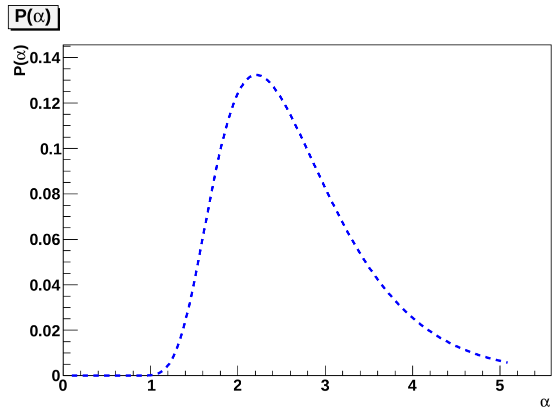
JETS



D0 e INCLUSIVE



D0 e HIGH E_T

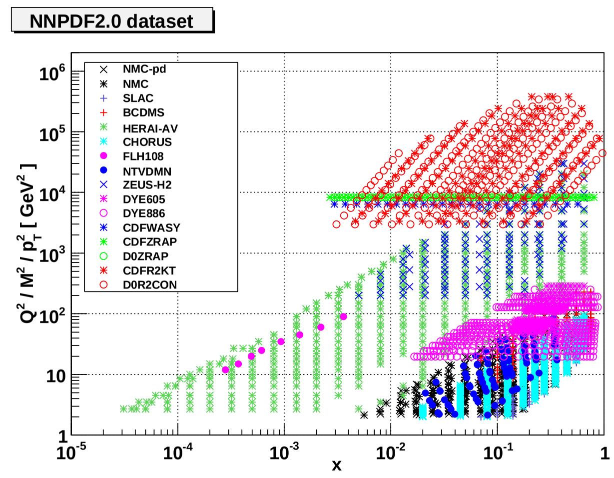


- JETS: $\Rightarrow$ CONSISTENT DATA
- $W^\pm$ CHARGE ASYMMETRIES, D0 INCLUSIVE e DATA $\Rightarrow$ UNCERTAINTIES UNDERESTIMATED BY $\sim 70\%$ (PROB. PEAKS AT $\alpha \sim 1.7$)
- $W^\pm$ CHARGE ASYMMETRIES, D0 e DATA WITH $E_T > 35$ GEV $\Rightarrow$ INCONSISTENT DATA

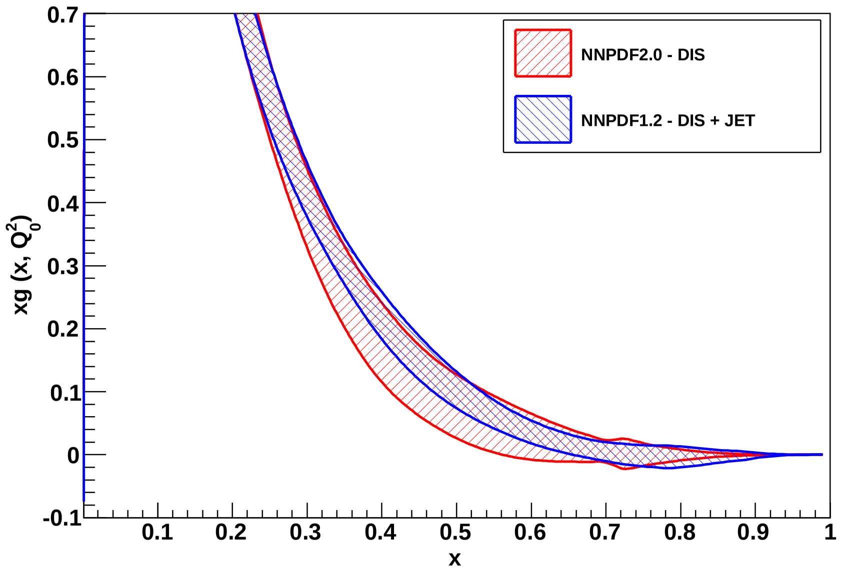
THE IMPACT OF HADRON COLLIDER DATA

JETS AT THE TEVATRON

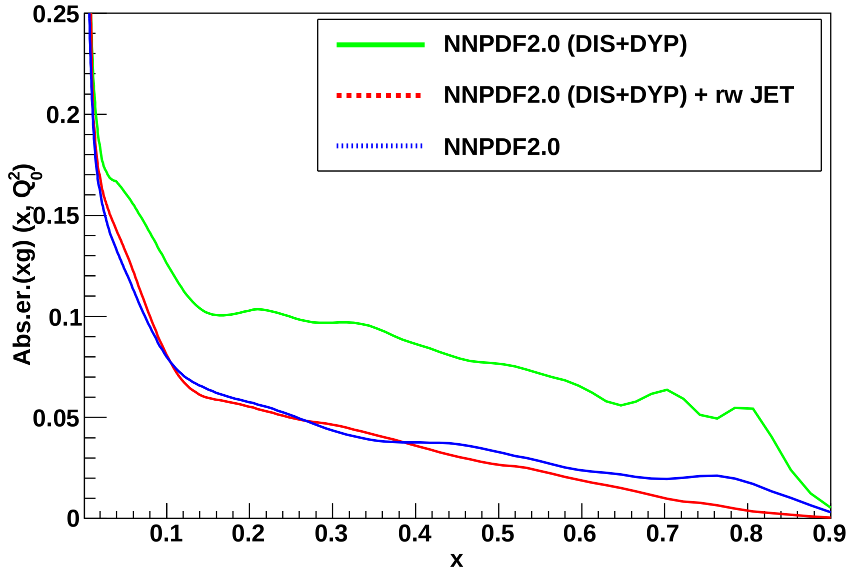
- GOOD CONSISTENCY $\Rightarrow$ ONE-JET INCLUSIVE RUN II DATA WELL REPRODUCED EVEN WHEN NOT FITTED (LARGE x GLUON WELL DETERMINED BY SCALING VIOLATIONS)
- IMPACT ON LARGE x GLUON
- SIGNIFICANT IMPROVEMENT IN GLUON ACCURACY FOR $0.1 \lesssim x \lesssim 0.7$



GLUON

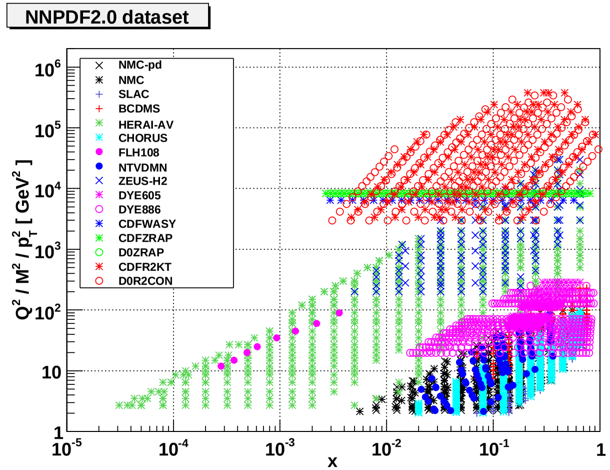


GLUON UNCERTAINTY



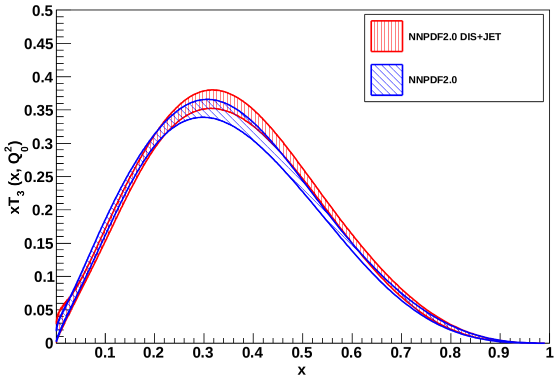
TEVATRON DRELL-YAN+W, Z-PRODUCTION

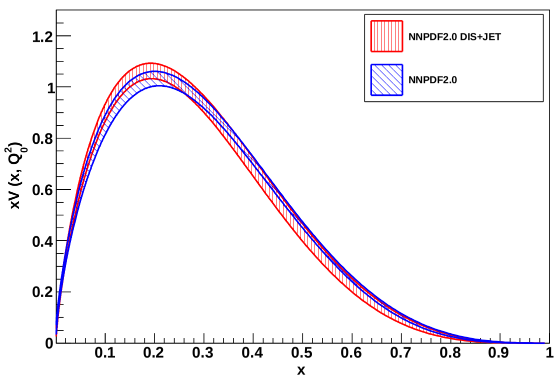
- VERY SUBSTANTIAL IMPROVEMENT IN FIT QUALITY WHEN DATA INCLUDED $\Rightarrow$ SOME PDF COMBINATIONS POORLY DETERMINED WITHOUT THESE DATA
- HUGE IMPROVEMENT IN SEA ASYM $\bar{u}-\bar{d}$ & STRANGENESS $s-\bar{s}$
- SIGNIFICANT IMPROVEMENT IN TOTAL VALENCE $(\sum_i (q_i - \bar{q}_i))$ & ISOTRIPLET $(u+\bar{u}-(d+\bar{d}))$



ISOTRIPLET

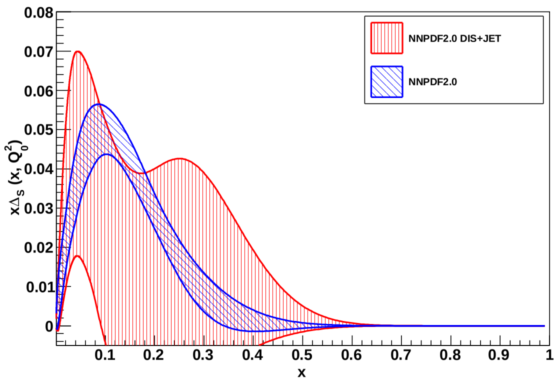
VALENCE

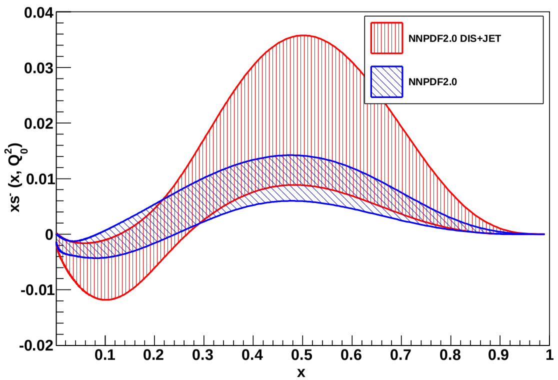




$\bar{d}-\bar{u}$

$s-\bar{s}$

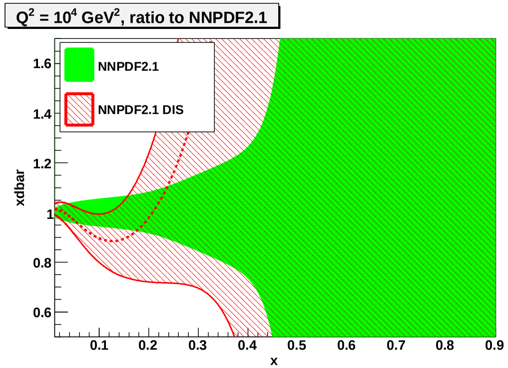




TEVATRON VS DIS: CONSISTENCY ISSUES?

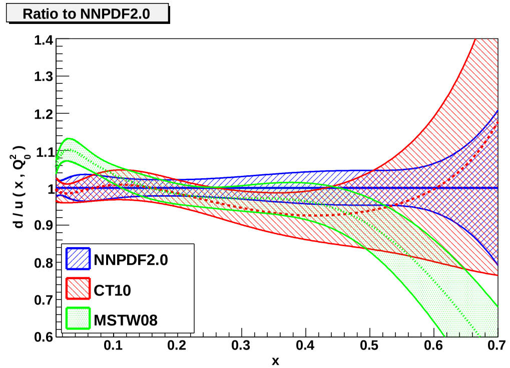
DIS VS DIS+DY:

DOWN



COMPARISON BETWEEN SETS:

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

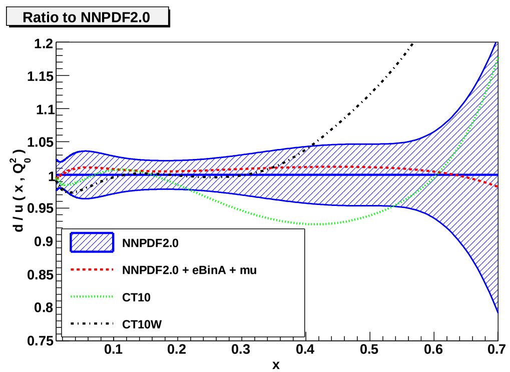


- DOWN AND UP/DOWN: TENSION BETWEEN DIS AND DY DATA
- PERHAPS NUCLEAR CORRECTIONS & HIGHER TWISTS PLAY A ROLE IN NMC, BCDMS DIS
- DIFFERENT PDF SETS IN PARTIAL DISAGREEMENT

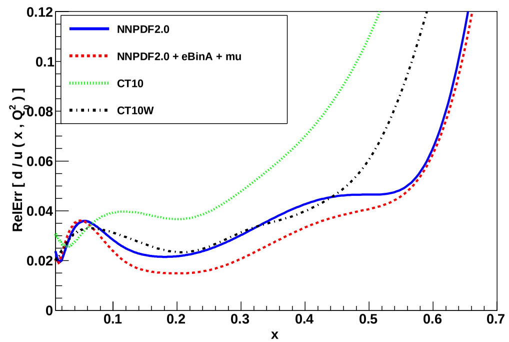
TEVATRON RUN II W ASYMMETRY DATA

- CTEQ: NO GLOBALLY CONSISTENT FIT FOUND: CT10W WITH W DATA, CT10 WITHOUT
- NNPDF: SIZABLE REDUCTION IN UNCERTAINTY

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



RELATIVE UNCERTAINTY



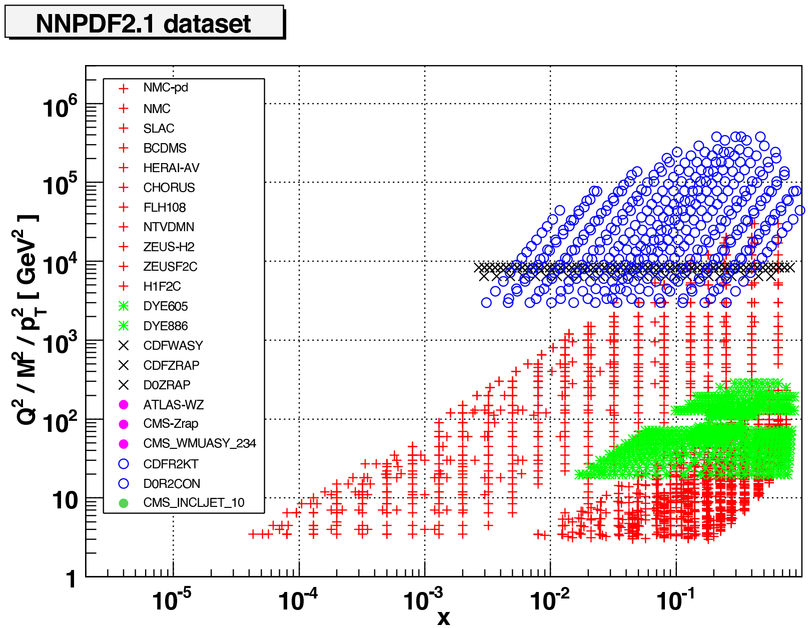
- W ASYMMETRY DATA DETERMINE THE u/d RATIO ACCURATELY, TENSION WITH DIS DATA
- BETTER TO DO WITHOUT NMC/BCDMS DATA (NO LOW Q^2 DATA, NO NUCLEAR TARGETS)

PRECISION PHYSICS AT THE LHC

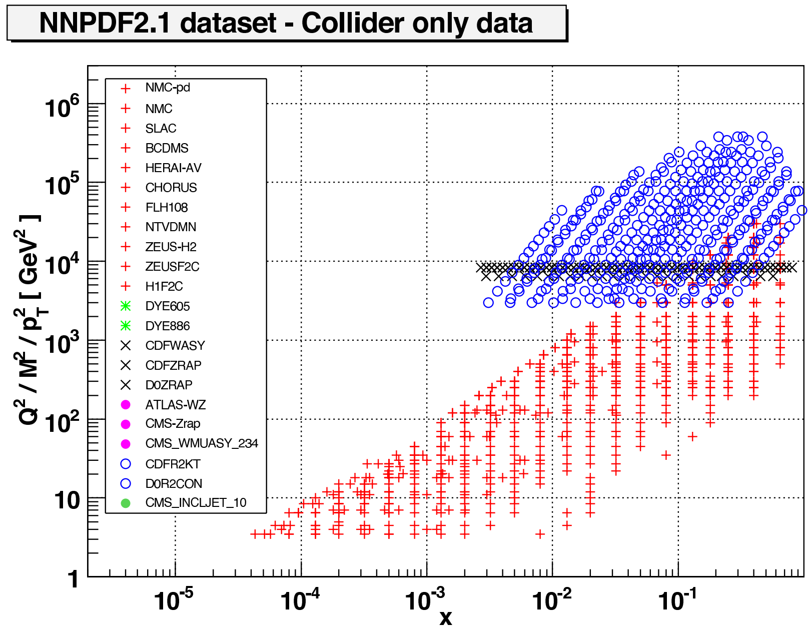
COLLIDER ONLY FITS?

NO FIXED TARGET DATA $\Leftrightarrow$ NO LOW-ENERGY TROUBLE

THE NNPDF2.1 DATASET



NNPDF2.1 - COLLIDER ONLY

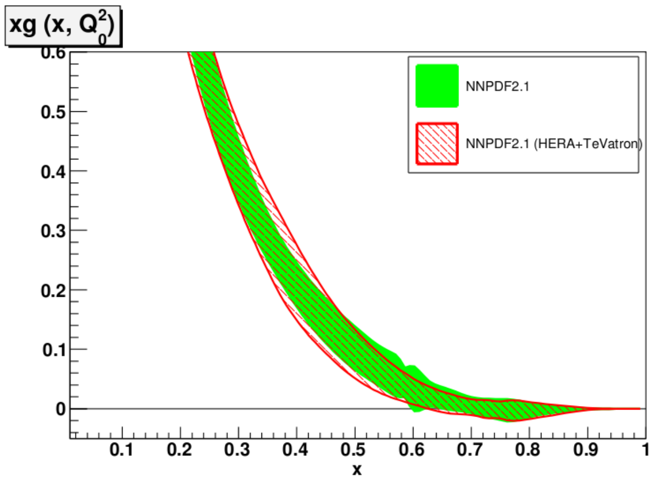


PDFs FROM HERA+TEVATRON DATA?

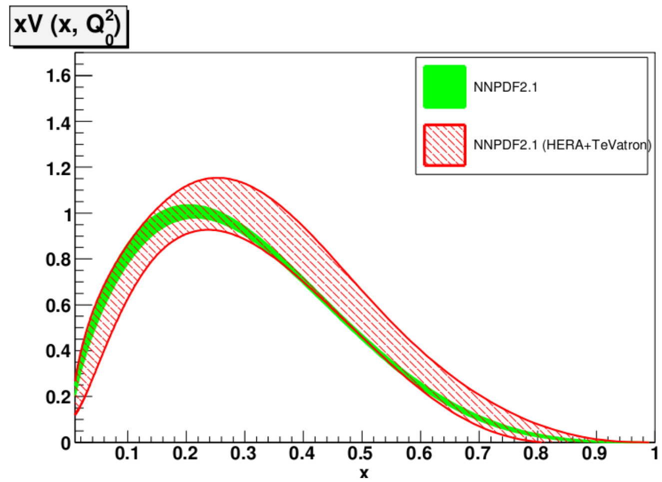
COLLIDER ONLY FITS?

THE PDFs

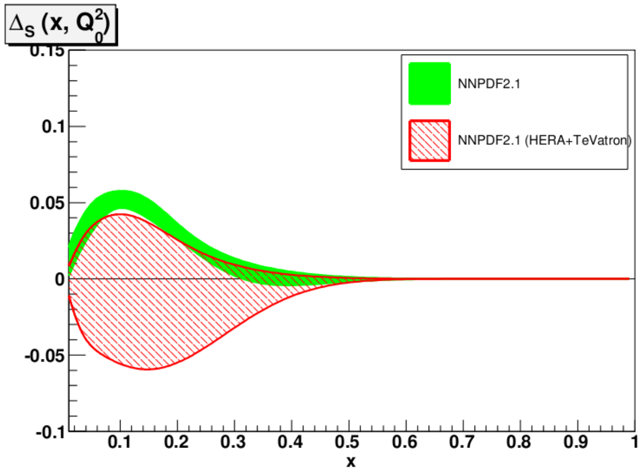
GLUON



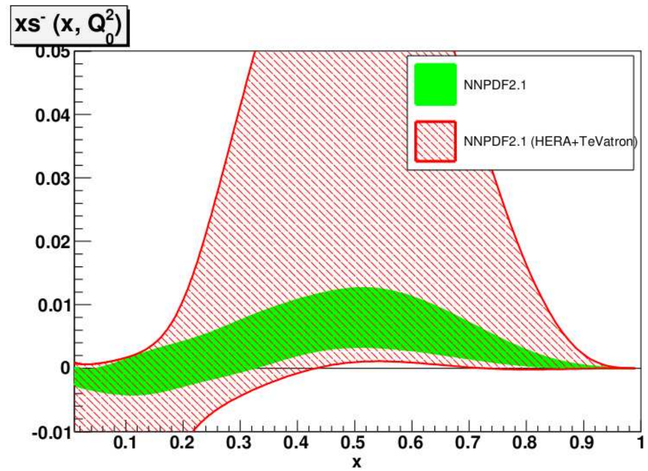
VALENCE



SEA ASYM: $\bar{u} - \bar{d}$



STRANGE: $s - \bar{s}$

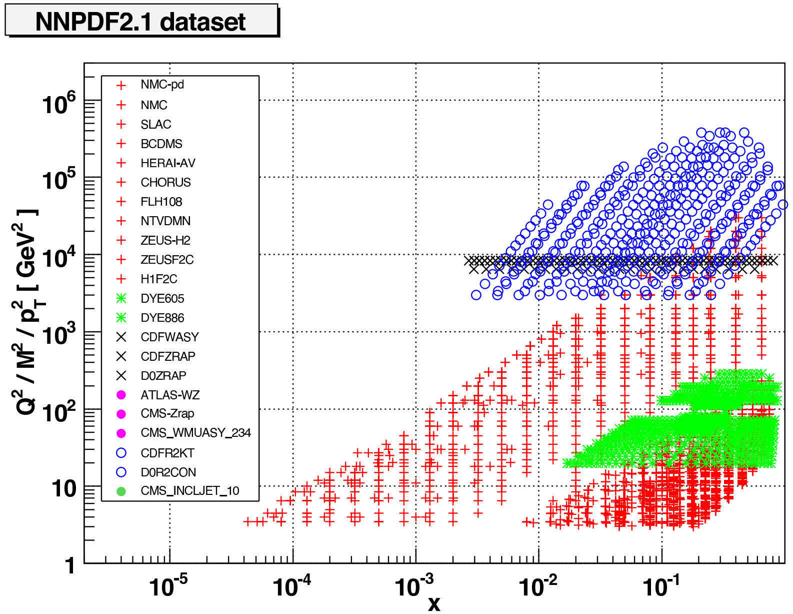


- GOOD ACCURACY FOR GLUON
- GREAT LOSS OF ACCURACY FOR FLAVOR SEPARATION

COLLIDER ONLY FITS?

NO FIXED TARGET DATA $\Leftrightarrow$ NO LOW-ENERGY TROUBLE

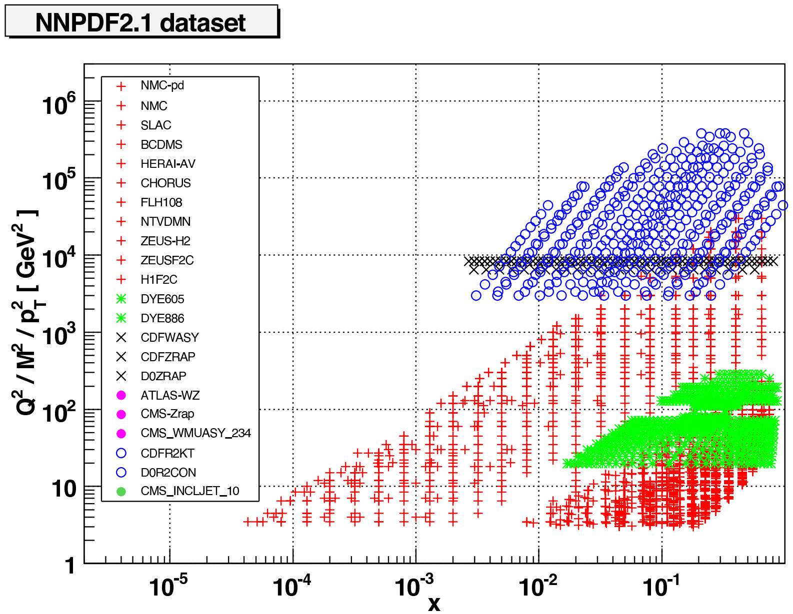
THE NNPDF2.1 DATASET



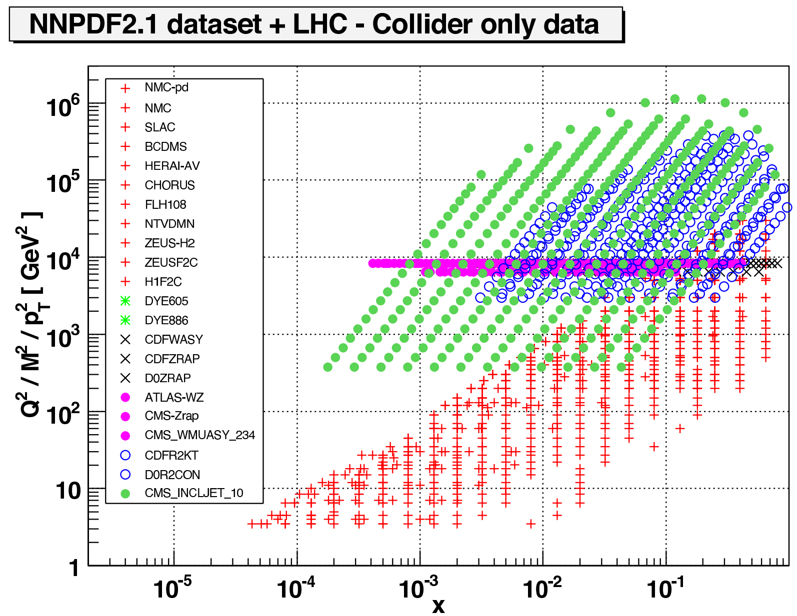
COLLIDER ONLY FITS?

NO FIXED TARGET DATA $\Leftrightarrow$ NO LOW-ENERGY TROUBLE

THE NNPDF2.1 DATASET



NNPDF2.1 COLL. ONLY+LHC



LHC DATA CAN PROVIDE THE MISSING INFORMATION!

W AND Z PRODUCTION AT THE LHC

- $W^{\pm}$, Z RAPIDITY DISTRIBUTIONS MEASURED BY ATLAS WITH $|\eta| < 2.4$, 3.2, 36 PB^{-1}, FULLY CORRELATED SYSTEMATICS
- CMS, SAME KIN., SYST. YET UNAVAILABLE, MUON ASYMMETRIES W. 234 PB^{-1}
- LHCB, $2 < \eta| < 4.5$, 36 PB^{-1}, SYST. YET UNAVAILABLE

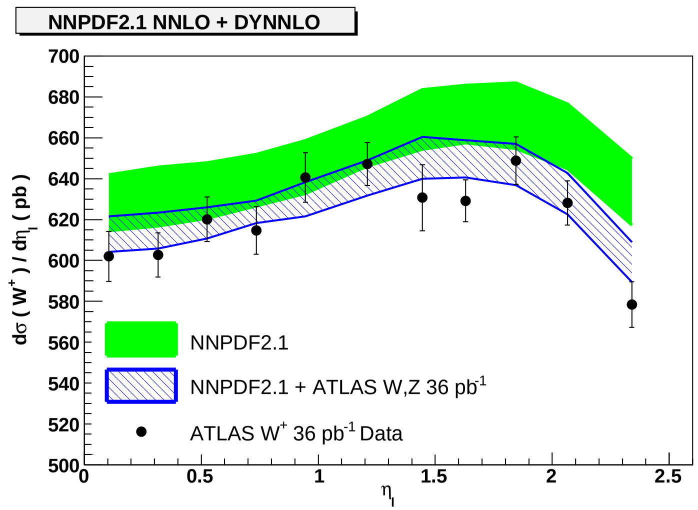
FIT QUALITY: PDF SETS

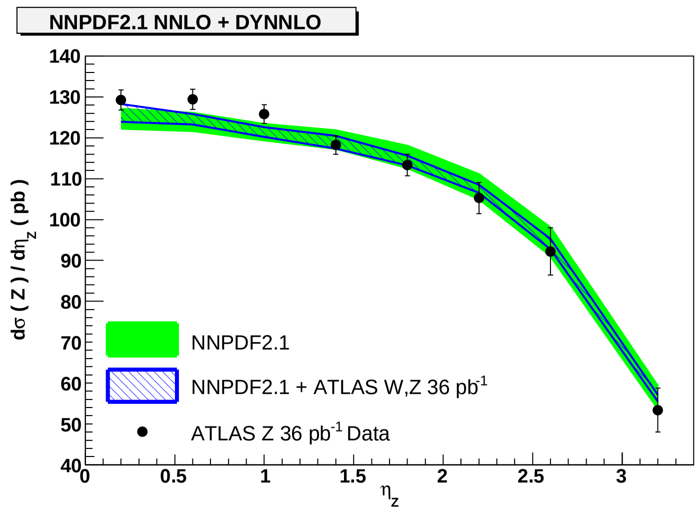
DATASET	χ^2 NNPDF2.1	χ^2 MSTW08	χ^2 HERAPDF1.5
ATLAS W^+ 36 PB^{-1}	5.7	6.5	5.3
ATLAS W^- 36 PB^{-1}	2.5	4.1	6.4
ATLAS Z 36 PB^{-1}	1.8	3.7	2.9
CMS Z RAPIDITY 36 PB^{-1}	1.9	2.9	3.0
CMS MUON ASYMMETRY 234 PB^{-1}	2.0	3.4	2.1
LHCB Z RAPIDITY 36 PB^{-1}	1.1	0.7	0.8
LHCB W LEPTON ASYMMETRY 36 PB^{-1}	0.5	0.6	0.5

- CONSIDERABLE DISCRIMINATING POWER (BIG DIFFERENCES BETWEEN SETS)
- CORRELATED SYSTEMATICS CRUCIAL (OTHERWISE WASHES OUT DIFFERENCES)

W AND Z PRODUCTION AT THE LHC

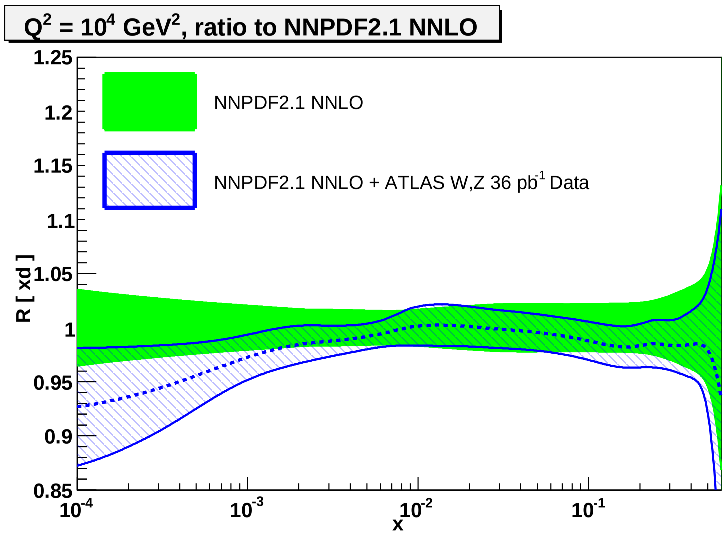
FITTING THE ATLAS DATA



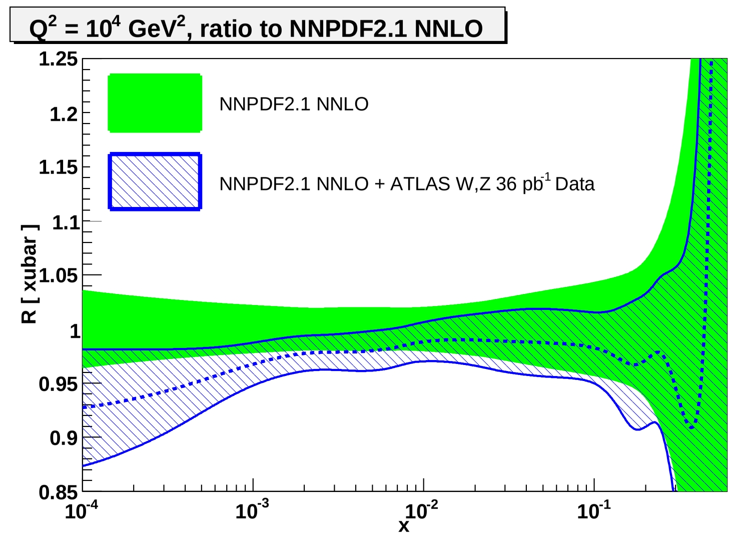


IMPACT ON LIGHT QUARK PDFs

DOWN



ANTIUP

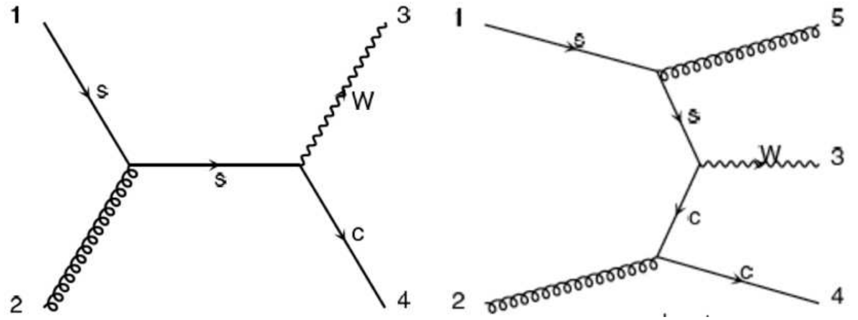


(NNPDF, 2012 prelim.)

- SIGNIFICANT IMPACT ON THE SHAPE
- SIZABLE REDUCTION OF THE UNCERTAINTY

$W + c$ PRODUCTION AT THE LHC

- STRANGENESS PROBED DIRECTLY
- MIGHT REPLACE NEUTRINO DATA



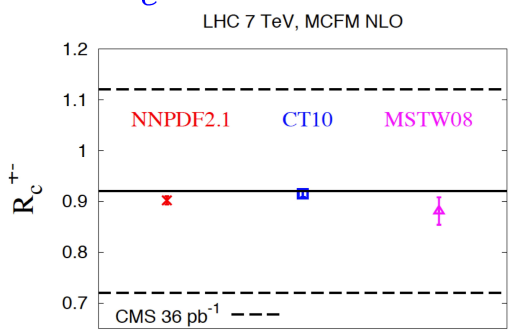
THE CMS MEASUREMENT

$$R_c^{\pm} \equiv \frac{\sigma(W^+ + \bar{c})}{\sigma(W^- + c)}; \quad R_c \equiv \frac{\sigma(W+c)}{\sigma(W+\text{jets})}$$

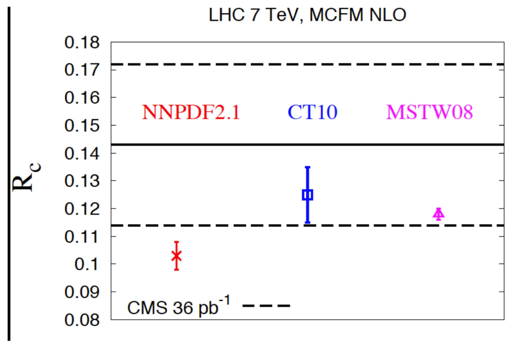
$$R_c^{\pm} = 0.92 \pm 0.19\text{stat.} \pm 0.04\text{syst.}$$
$$R_c = 0.143 \pm 0.015\text{stat.} \pm 0.024\text{syst.}$$

Ratio	MCFM (CT10)	MCFM (MSTW08)	MCFM (NNPDF21)
$R_c^{\pm}$	$0.915^{+0.006}_{-0.006}$	$0.881^{+0.022}_{-0.032}$	0.902 ± 0.008
R_c	$0.125^{+0.013}_{-0.007}$	$0.118^{+0.002}_{-0.002}$	0.103 ± 0.005

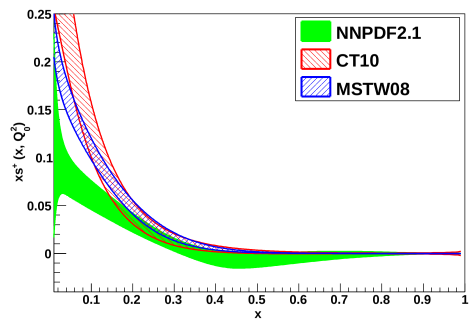
$R_c^{\pm}$ DATA VS TH.



R_c DATA VS TH.



TOTAL STRANGENESS



- DIFFERENCES BETWEEN AVAILABLE SETS SIZABLE
- DATA STILL TOO IMPRECISE, BUT...

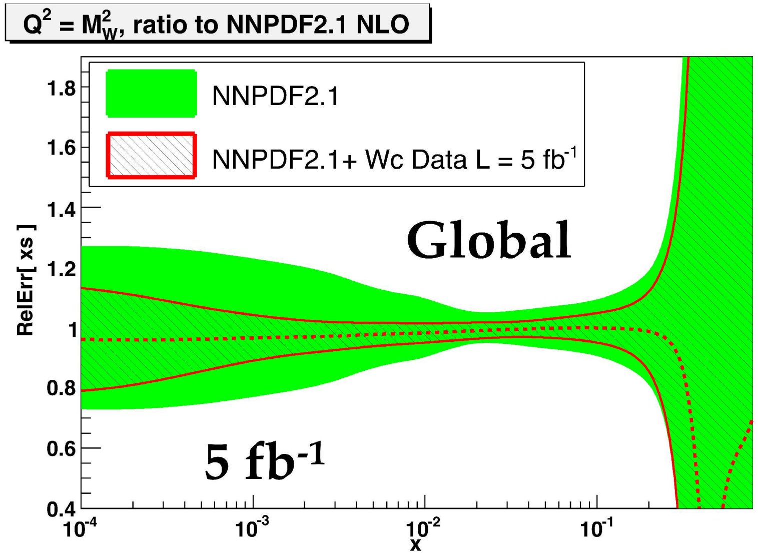
$W + c$ PRODUCTION AT THE LHC

- SIMULATED MEASUREMENT OF c RAPIDITY DISTRIBUTION WITH AMC@NLO

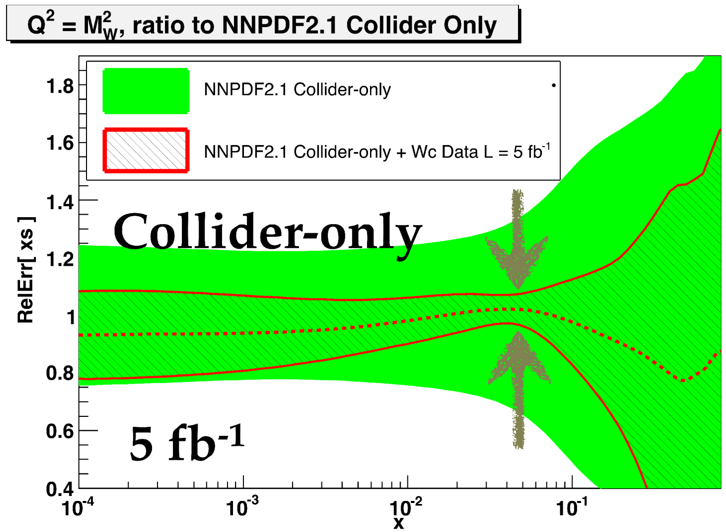
  (J. Rojo, S. Frixione, M. Mangano, 2012, prelim.)
- CMS KINEMATICS $p_T^{\rm jet} > 20$ GeV, $p_T^{\mu} > 25$ GeV $\eta^{\rm jet}$, $\eta^{\mu} < 2.1$
- 15% CHARM TAGGING EFFICIENCY (CMS)
- CURRENTLY 36 PB^{-1}, BUT 5 FB^{-1} SUFFICIENT

THE IMPACT ON STRANGENESS

IN THE NNPDF2.1 FIT

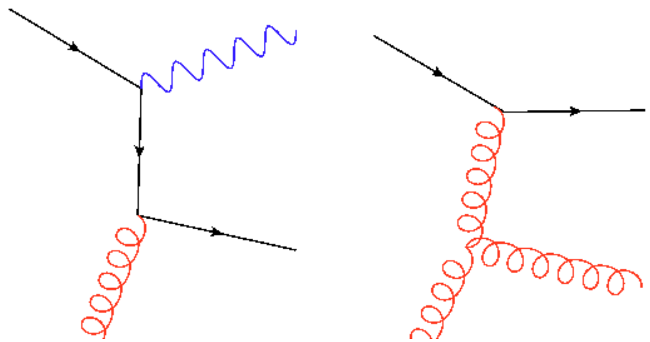


IN THE COLLIDER-ONLY FIT



- IMPACT SIGNIFICANT EVEN ON CURRENT GLOBAL FIT
- COULD DO WITHOUT NEUTRINO DATA

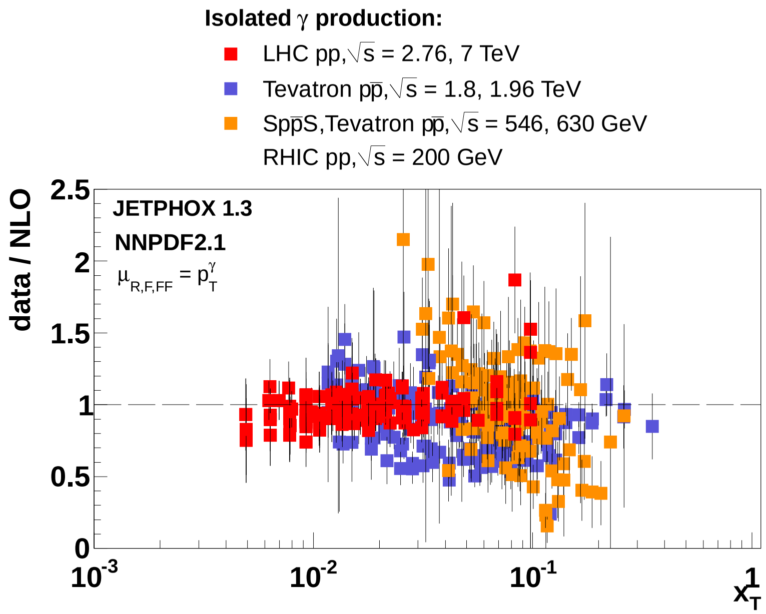
PROMPT PHOTON PRODUCTION



- DIRECT PROBE OF THE qg LUMINOSITY
- MEASURED BY CMS+ATLAS

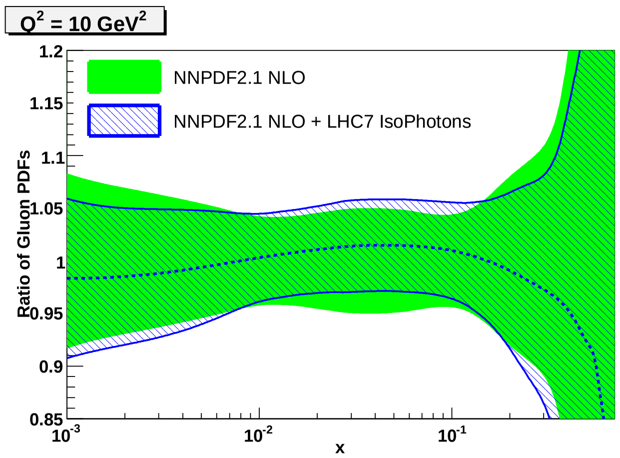
(D. d'Enterria, J. Rojo, 2012, prelim.)

$x_t = x_1 x_2$ RANGE

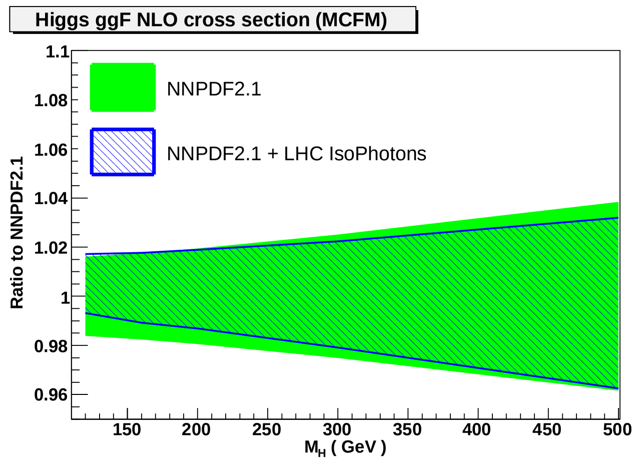


THE IMPACT OF LHC PROMPT PHOTON DATA

GLUON



HIGGS FROM GLUON FUSION



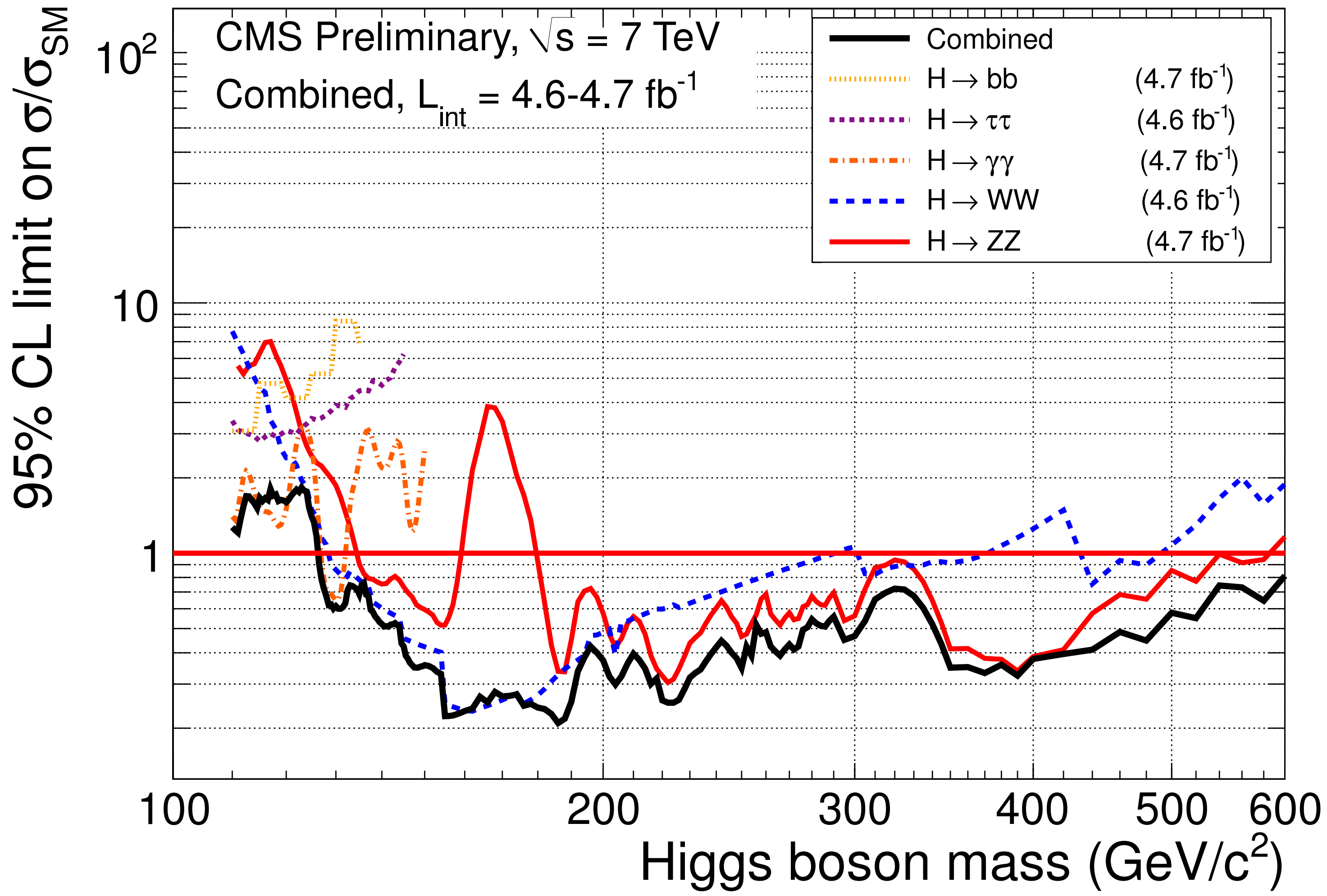
- MODERATE IMPACT ON GLOBAL FIT (BUT COULD RESOLVE DISCREPANCIES)
- SUFFICIENT TO AFFECT HIGGS CROSS SECTION

OUTLOOK

WISHLIST $\Leftrightarrow$ ROADMAP

- PRECISION LHC PHYSICS MUST USE PRECISE LHC DATA
- A "COLLIDER ONLY" HERA+LHC PDF FIT
  (TEVATRON DATA MIGHT BE SUPERFLUOUS)
  - MEDIUM & LARGE x GLUON
    - PROMPT PHOTONS
    - (PRECISION) JETS
    - INCLUSIVE W p_T DISTRIBUTIONS
  - LIGHT FLAVOR SEPARATION
    - LOW–MASS DRELL-YAN
    - HIGH–MASS DRELL-YAN
    - Z RAPIDITY DISTRIBUTIONS
    - W ASYMMETRIES
  - STRANGENES & HEAVY FLAVORS
    - STRANGENESS $\Rightarrow W + c$
    - CHARM $\Rightarrow Z + c$, $\gamma + c$
    - BOTTOM $Z + b$
- MOST OF THESE ALREADY AVAILABLE (AT LEAST IN PRELIM. FORM)
- ATLAS JETS AND W, Z RAP. DISTNS. FULLY AVAILABLE (36 PB^{-1})

WE NEED THIS FOR THE SIGNAL TO RAISE ABOVE THRESHOLD!

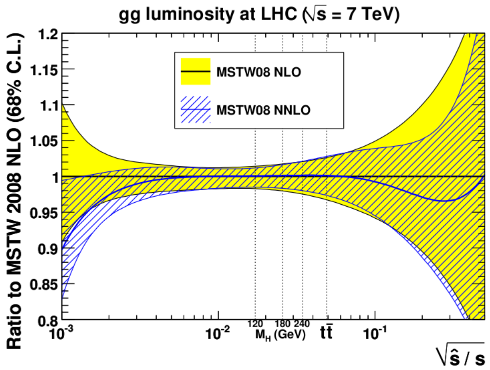


EXTRAS

NNLO STATUS

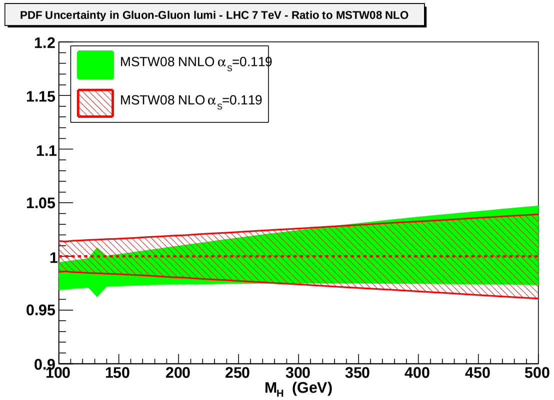
GLUON-GLUON PARTON LUMINOSITY

MSTW NNLO/NLO
best fit α_s



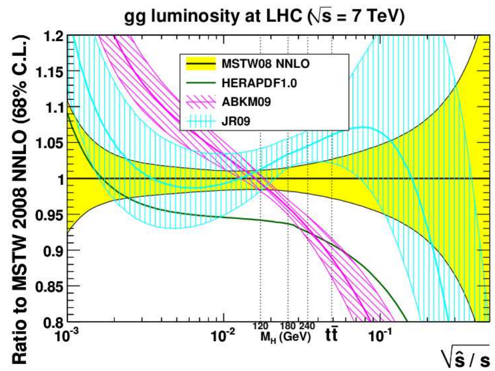
(G. Watt, 2011)

MSTW NNLO/NLO
fixed α_s



(J. Rojo, 2011)

LESS GLOBAL VS MSTW



(G. Watt, 2011)

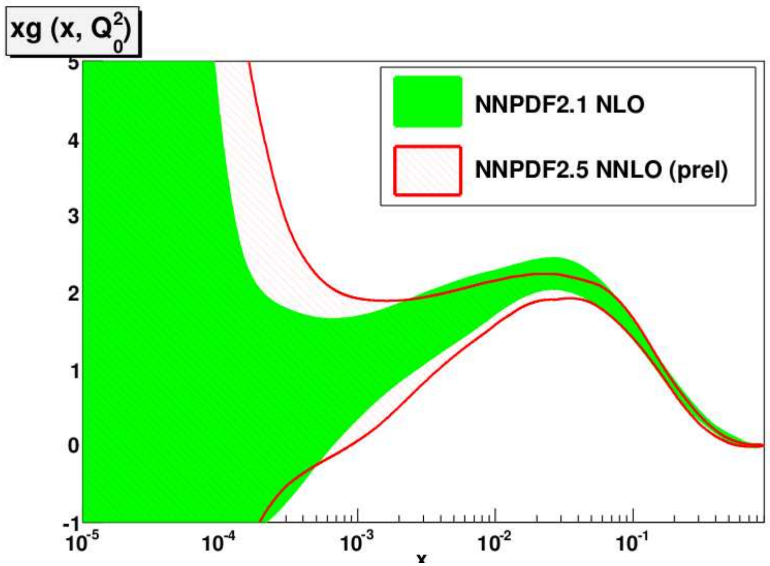
- NLO VS NNLO MSTW LUMINOSITIES QUITE CLOSE
  ... BUT PARTLY DUE TO LOWER NNLO α_s (0.117 VS 0.120)
- DIFFERENCES AS DATASED VARIED MUCH LARGER
  ... BUT IN SOME CASE ALSO α_s QUITE DIFFERENT (ABKM: 0.113)

NNLO PROGRESS

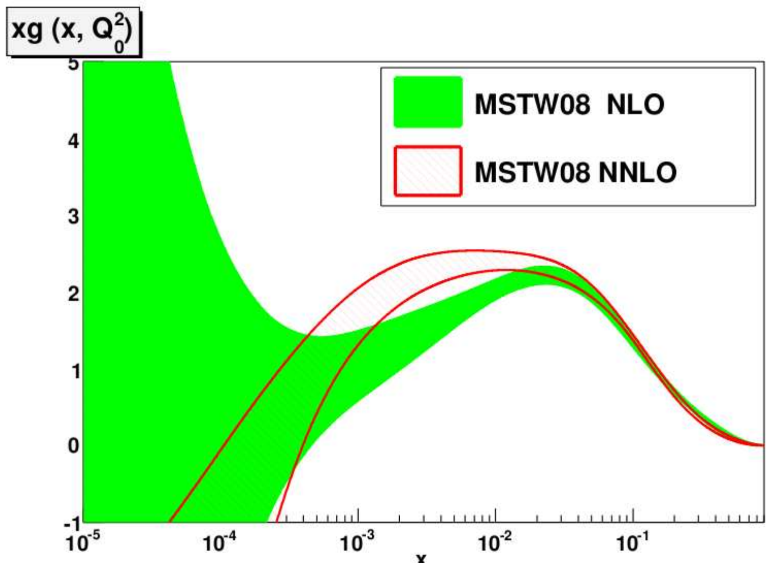
- CTEQ NNLO IN PREPARATION
- NNPDF NNLO PRESENTED IN PRELIM. FORM (DIS ONLY AT NNLO, FULL NNLO IN PREPARATION)
  - GLUON SIMILAR TO MSTW, BUT NO SMALL x INSTABILITY
  -

GLUON PDF: NLO VS NNLO

NNPDF



MSTW

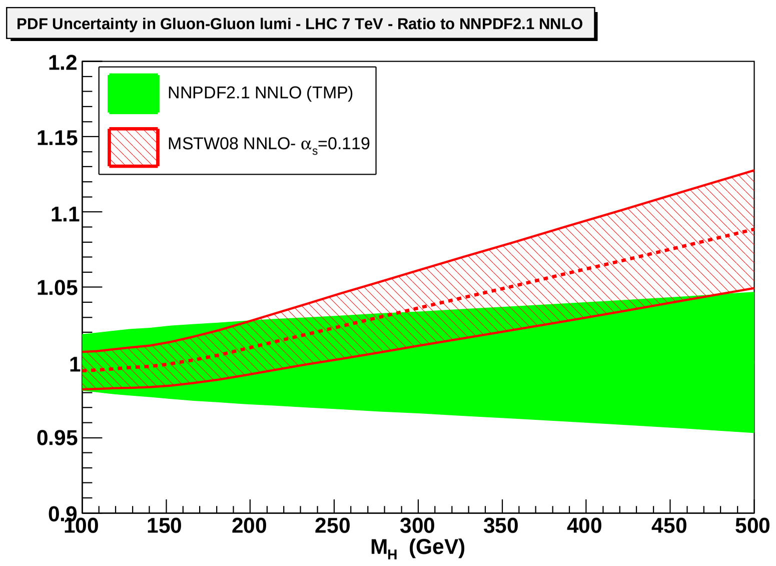


NNLO PROGRESS

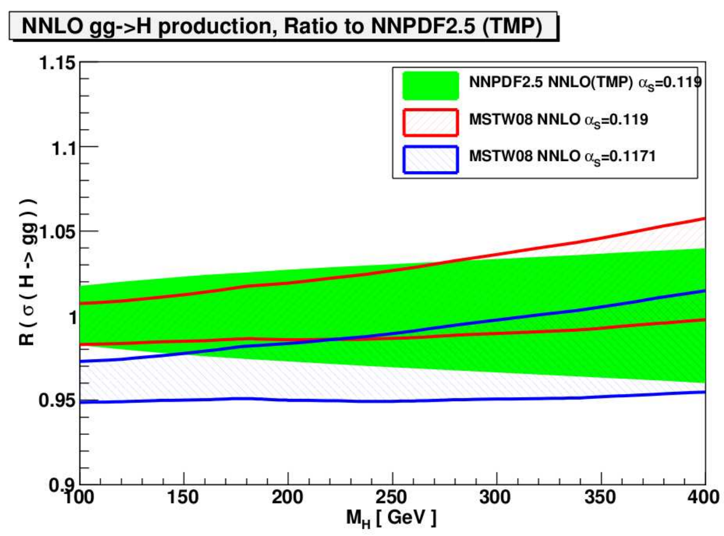
- CTEQ NNLO IN PREPARATION
- NNPDF NNLO PRESENTED IN PRELIM. FORM (DIS ONLY AT NNLO, FULL NNLO IN PREPARATION)
  - GLUON SIMILAR TO MSTW, BUT NO SMALL x INSTABILITY
  - LUMINOSITY & HIGGS XSECT QUITE CLOSE TO MSTW
    ... PROVIDED SAME α_s USED

GLUON LUMI & HIGGS XSECT: NNPDF VS MSTW

GLUON LUMI



HIGGS XSECT.



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)
  $\widetilde{\sigma}(x, y, Q^2) = F_2(x, Q^2) \left(2 - 2y + y^2/[1 + R(x, Q^2)]\right)$

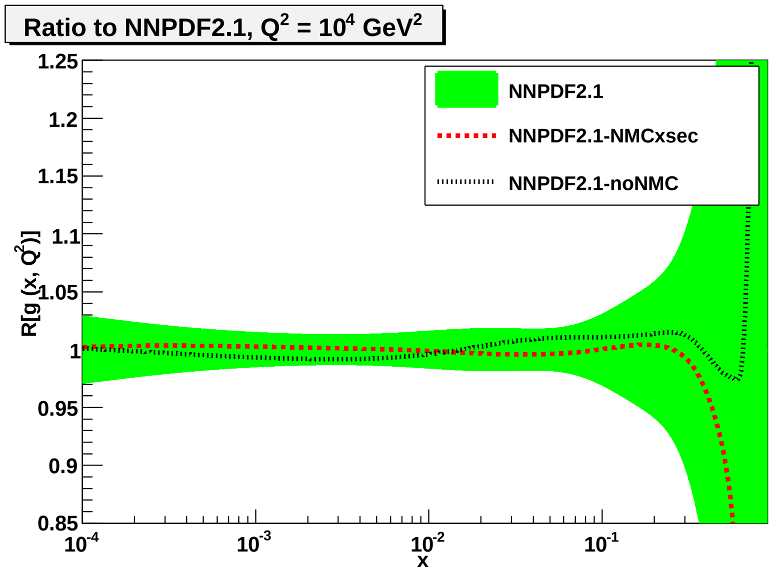
NMC DATA: A PROBLEM?

- 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)
  $\widetilde{\sigma}(x, y, Q^2) = F_2(x, Q^2)\left(2 - 2y + y^2/[1 + R(x, Q^2)]\right)$

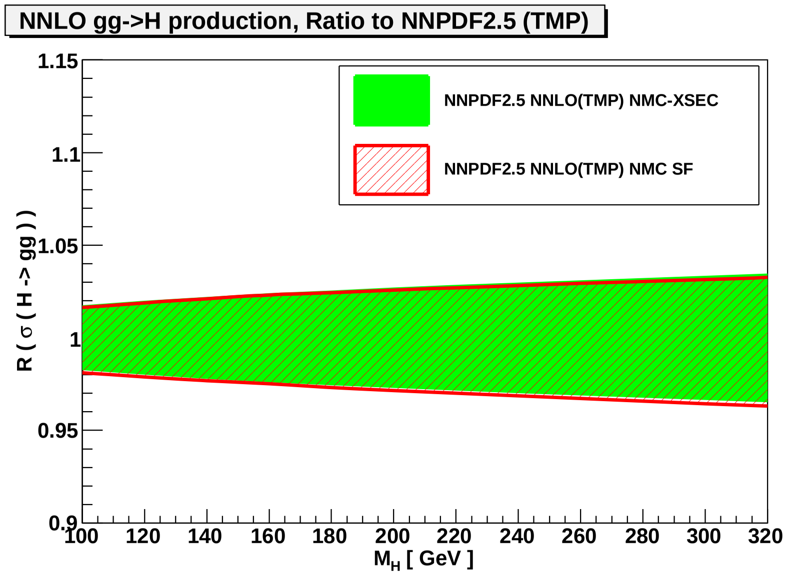
NNPDF FIT REPEATED WITH NMC XSECT OR NO NMC

(NLO: arXiv:1102.3182; NNLO: preliminary)

EFFECT ON GLUON



EFFECT ON HIGGS XSECT



NO SIGNIFICANT CHANGE OBSERVED

NMC DATA: A PROBLEM?

A CLOSER LOOK

- ALEKHIN ET AL. SEE LARGE SHIFT IN BEST-FIT α_s: FROM 0.1170 (NMC F_2) TO 0.1135 (NMC α_s)
- HOW MUCH OF THEIR CHANGE IN HIGGS XSECT ($\sim 8\%$) DUE TO CHANGE IN α_s?
- BASED ON POWER COUNTING, EXPECT $2.8\Delta\alpha_s = 8\%$ (!)
- EITHER REMOVE NMC DATA, OR CHANGE VALUE OF α_s IN MSTW FIT (Thorne, presented by Watt at PDF4LHC 2011):

NMC DATA: A PROBLEM?

A CLOSER LOOK

- ALEKHIN ET AL. SEE LARGE SHIFT IN BEST-FIT α_s: FROM 0.1170 (NMC F_2) TO 0.1135 (NMC α_s)
- HOW MUCH OF THEIR CHANGE IN HIGGS XSECT ($\sim 8\%$) DUE TO CHANGE IN α_s?
- BASED ON POWER COUNTING, EXPECT $2.8\Delta\alpha_s = 8\%$ (!)
- EITHER REMOVE NMC DATA, OR CHANGE VALUE OF α_s IN MSTW FIT (Thorne, presented by Watt at PDF4LHC 2011):
  - IMPACT OF NMC DATA NEGLIGIBLE
  - IMPACT OF α_s VALUE COMPARABLE (SOMEWHAT SMALLER) TO WHOLE ABKM EFFECT
  - MSTW AND ABKM RESULTS STILL QUITE DIFFERENT EVEN FOR SAME $\alpha_s = 0.113$

NNLO PDF	$\alpha_S(M_Z^2)$	σ_H at Tevatron	σ_H at 7 TeV LHC
MSTW08	**0.1171**	**0.342 pb**	**7.91 pb**
Use R_{1990} for NMC	0.1167	-0.7%	-0.9%
Cut NMC ($x < 0.1$)	0.1162	-1.2%	-2.1%
Cut all NMC data	0.1158	-0.7%	-2.1%
Fix $\alpha_S(M_Z^2)$	0.1130	-11%	-7.6%
ABKM09	0.1135	-26%	-11%

THE VALUE OF α_s

- DEDICATED MUNICH MEETING (FEB 2011):
  S. BETHKE PROPOSES TWO UPDATED VALUES:
  - (1) $\alpha_s = 0.1174 \pm 0.0011$
  - (2) $\alpha_s = 0.1187 \pm 0.0006$
  - BOTH INCLUDE NEW VALUE FROM τ DECAYS $\alpha_s = 0.1213 \pm 0.0014$ (WAS $\alpha_s = 0.1197 \pm 0.0016$)
  - VALUE (1) ALSO INCLUDES NEW SCET VALUE FROM e^+e^- THRUST (Abbate et al., 2010) $\alpha_s = 0.1135 \pm 0.0010$, BUT ALL UNCERTAINTIES RESCALED BY FACTOR 2
  - VALUE (2) EXCLUDES IT

- AVERAGING THE TWO MOST RELIABLE VALUES (GLOBAL EW FIT & τ, BOTH N^{3}LO, NO DEP. ON HADRON STRUCTURE) GIVES
  $\alpha_s = 0.1209 \pm 0.0013$

THE VALUE OF α_s

- DEDICATED MUNICH MEETING (FEB 2011):
  S. BETHKE PROPOSES TWO UPDATED VALUES:
  - (1) $\alpha_s = 0.1174 \pm 0.0011$
  - (2) $\alpha_s = 0.1187 \pm 0.0006$
  - BOTH INCLUDE NEW VALUE FROM τ DECAYS $\alpha_s = 0.1213 \pm 0.0014$
    (WAS $\alpha_s = 0.1197 \pm 0.0016$)
  - VALUE (1) ALSO INCLUDES NEW SCET VALUE FROM e^+e^- THRUST
    (Abbate et al., 2010)
    $\alpha_s = 0.1135 \pm 0.0010$, BUT ALL UNCERTAINTIES RESCALED BY FACTOR 2
  - VALUE (2) EXCLUDES IT

- AVERAGING THE TWO MOST RELIABLE VALUES (GLOBAL EW FIT & τ, BOTH N^3LO, NO DEP. ON HADRON STRUCTURE) GIVES
  $\alpha_s = 0.1209 \pm 0.0013$

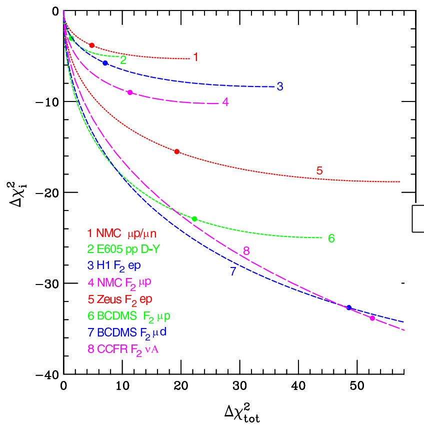
- UNCERTAINTY LARGER THAN 0.002 QUITE UNLIKELY

METHODOLOGY: "STANDARD" (HESSIAN) APPROACH

THE TOLERANCE PROBLEM

- CHOOSE A FIXED FUNCTIONAL FORM ($\sim 25 \div 30$ PARMS. GLOBALLY) AND FIT PARMS TO DATA
- UNCERTAINTIES DETERMINED FROM HESSIAN MATRIX AT MINIMUM
- STANDARD $\Delta\chi^2 = 1$ BANDS TOO NARROW $\Rightarrow$ LARGE DISCREPANCIES FOR INDIVIDUAL EXPERIMENTS
- TOLERANCE $\Rightarrow$ ENVELOPE OF UNCERTAINTIES OF EXPERIMENTS
- DYNAMICAL $\Rightarrow$ SEPARATELY DETERMINED FOR EACH HESSIAN EIGENVECTOR

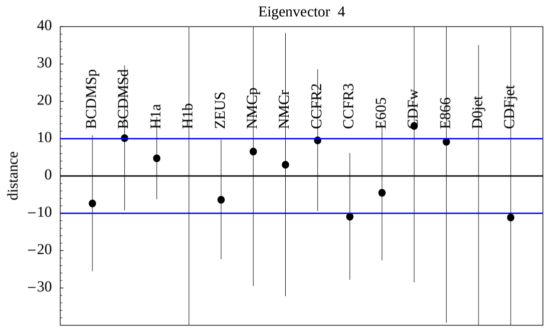
MINIMUM χ_i^2 VS GLOBAL χ^2



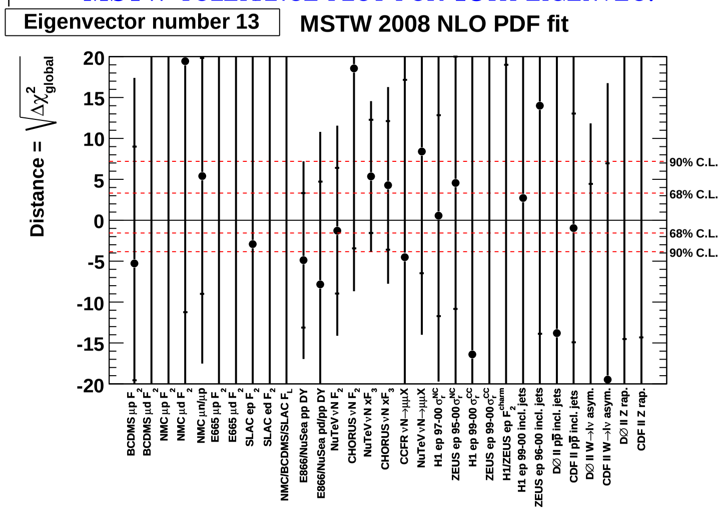
Collins, Pumplin

2001

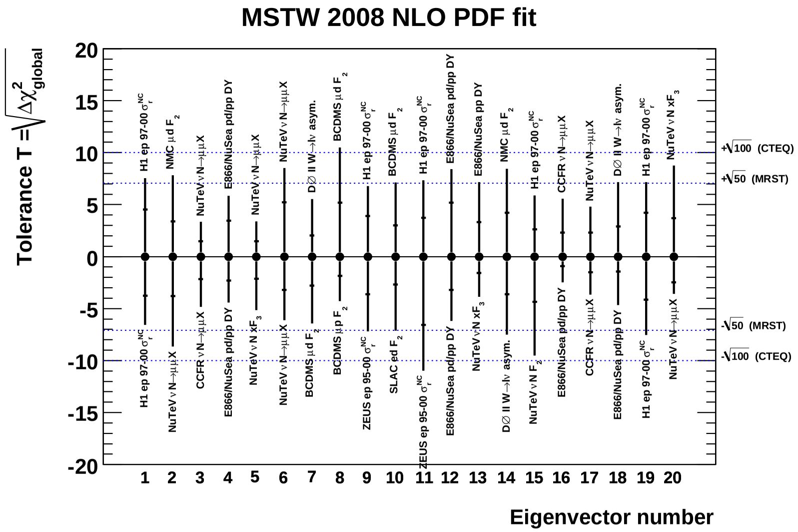
CTEQ TOLERANCE PLOT FOR 4TH EIGENVEC.



MSTW TOLERANCE PLOT FOR 13TH EIGENVEC.



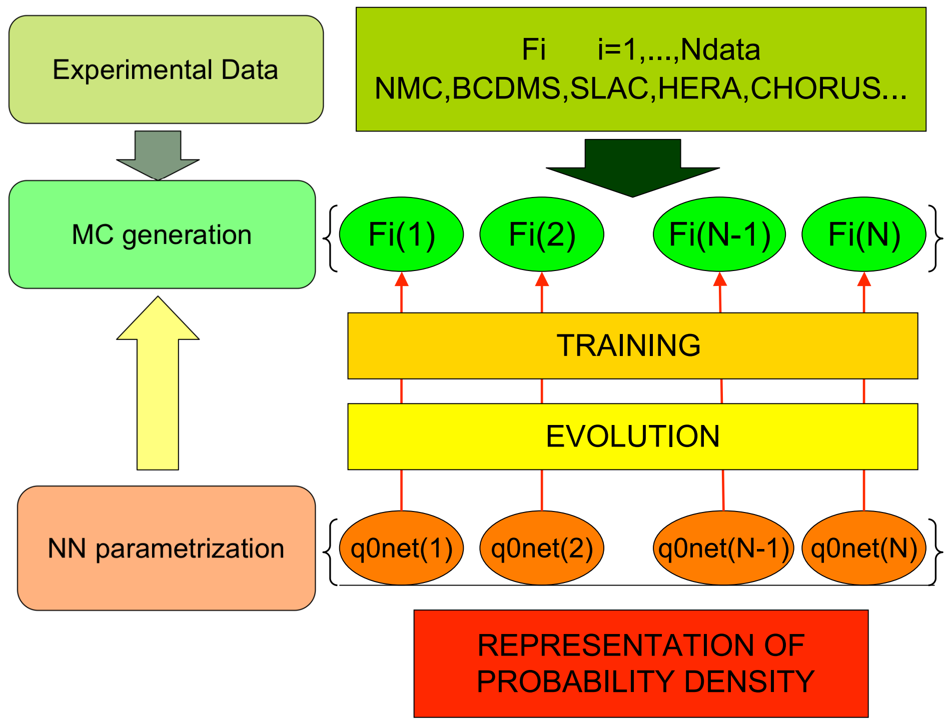
GLOBAL MSTW TOLERANCE



METHODOLOGY: NNPDF APPROACH

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

- START FROM MONTE CARLO SAMPLING OF DATA SPACE
- EACH PDF$\leftrightarrow$ NEURAL NETWORK PARAMETRIZED BY 37 PARAMETERS (NNPDF2.0: $37 \otimes 7 = 259$ PARMS)
  "INFINITE" NUMBER OF PARAMETERS$\Rightarrow$ CAN REPRESENT ANY FUNCTION
- FIT STOPS WHEN QUALITY OF FIT TO RANDOMLY SELECTED "VALIDATION" DATA (NOT FITTED) STOPS IMPROVING

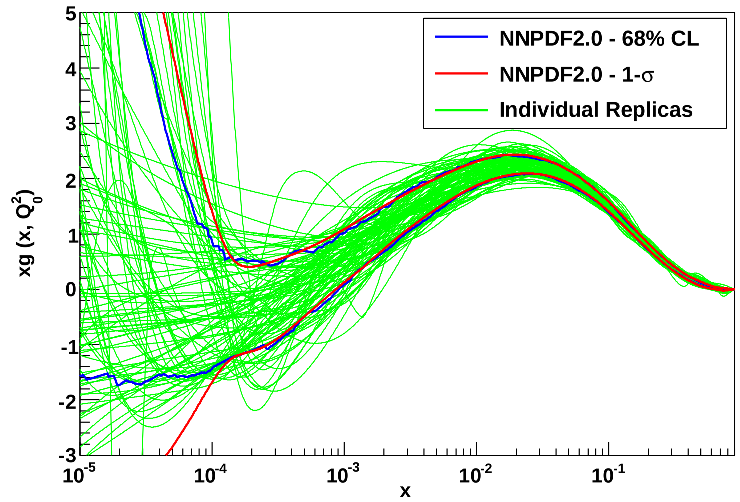


METHODOLOGY: NNPDF APPROACH

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

- START FROM MONTE CARLO SAMPLING OF DATA SPACE
- EACH PDF$\leftrightarrow$ NEURAL NETWORK PARAMETRIZED BY 37 PARAMETERS (NNPDF2.0: $37 \otimes 7 = 259$ PARMS)
  "INFINITE" NUMBER OF PARAMETERS$\Rightarrow$ CAN REPRESENT ANY FUNCTION
- FIT STOPS WHEN QUALITY OF FIT TO RANDOMLY SELECTED "VALIDATION" DATA (NOT FITTED) STOPS IMPROVING

CAN DETERMINE BOTH 68C.L.& 1-σ

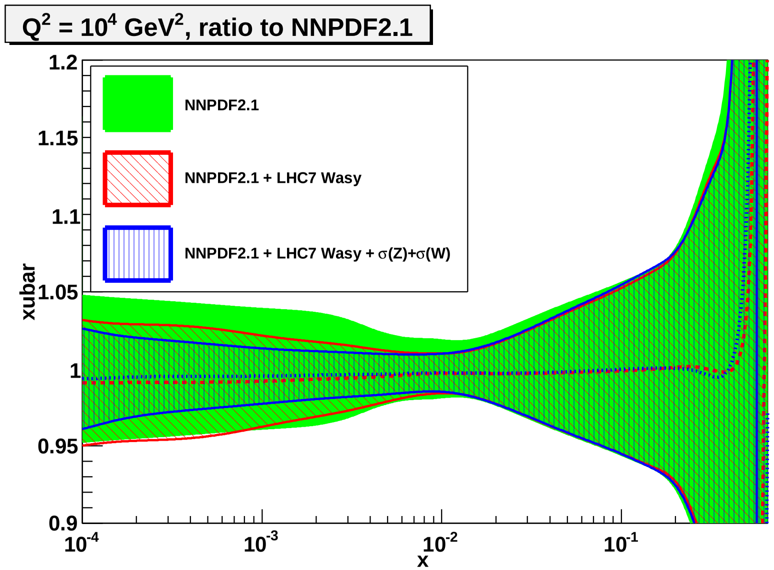


FUTURE LHC DATA

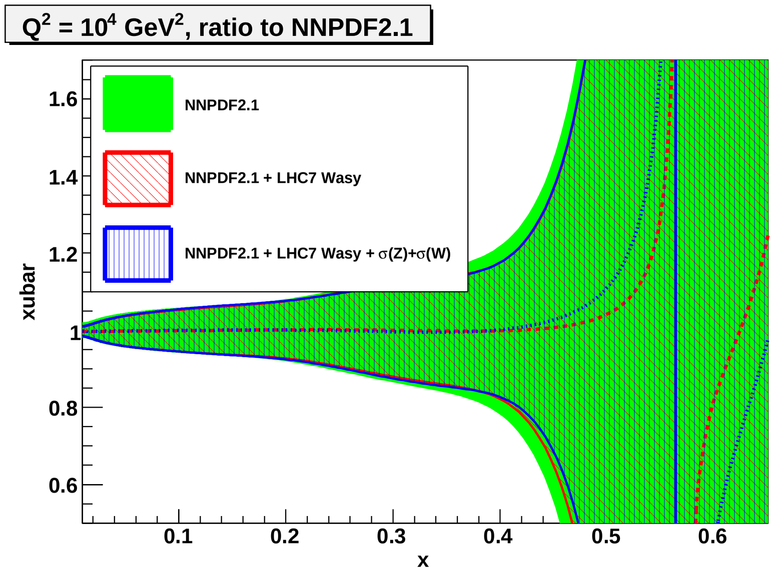
TEST THE IMPACT OF LHC 7TEV W AND Z DATA USING REWEIGHTING IN THE FOLLOWING "TOY" SCENARIOS:

- 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

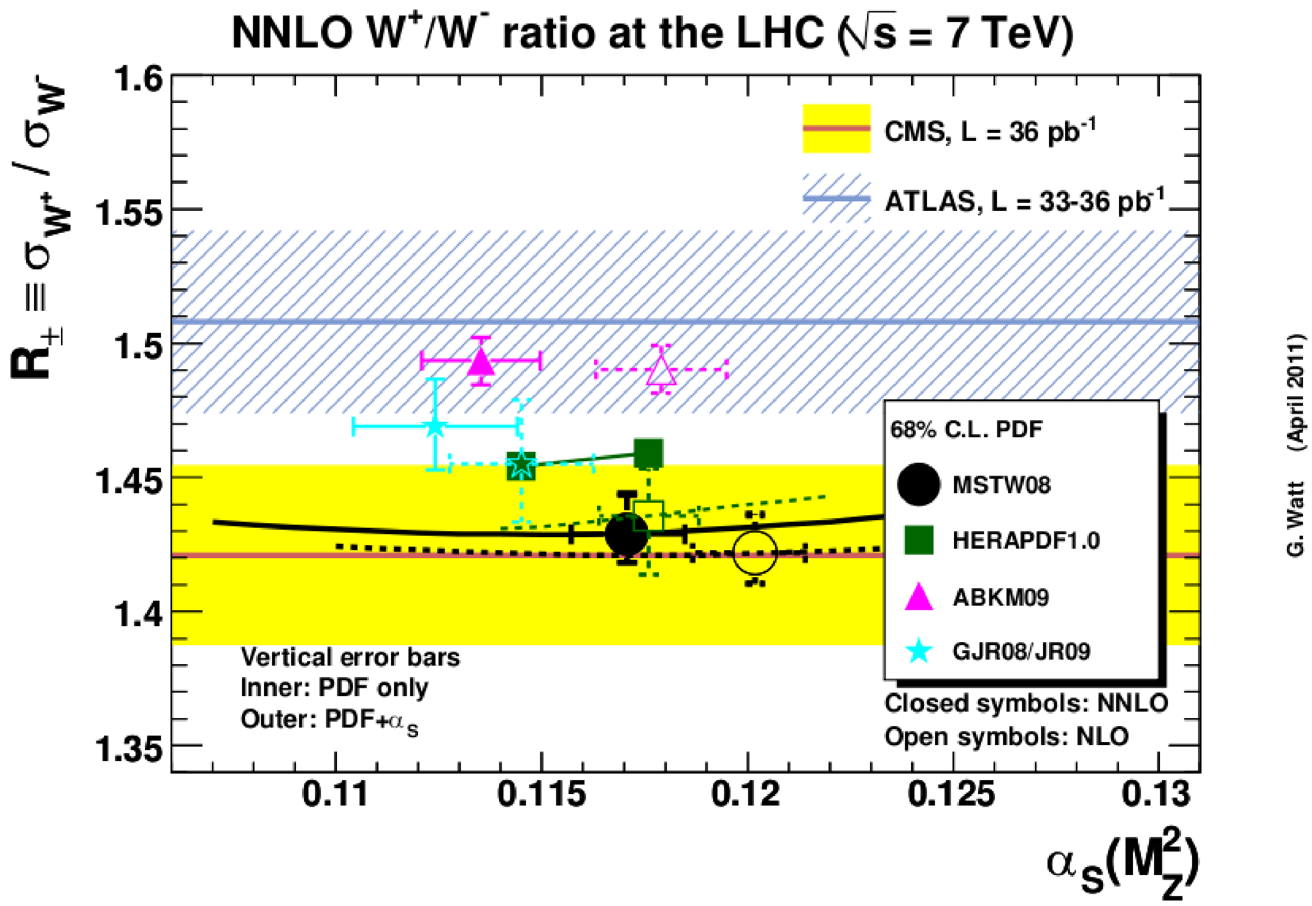


PERCENTAGE ACCURACY ON W, Z CROSS SECTIONS?

- NNLO PROBABLY NECESSARY
- REQUIRES INDEPENDENT ACCURATE MEASUREMENT OF LUMINOSITY
- ACCURATE ACCEPTANCE CALCULATIONS
- EVEN FOR W^+/W^- RATIO NOT EASY
  $\Rightarrow$

PERCENTAGE ACCURACY ON W, Z CROSS SECTIONS?

- NNLO PROBABLY NECESSARY
- REQUIRES INDEPENDENT ACCURATE MEASUREMENT OF LUMINOSITY
- ACCURATE ACCEPTANCE CALCULATIONS
- EVEN FOR W^+/W^- RATIO NOT EASY
  $\Rightarrow$CURRENT CMS-ATLAS DISAGREEMENT (?)



ACCEPTANCE: PROBLEMS?

- CMS [arXiv:1012.2466, CMS PAS EWK-10-005] use POWHEG (with CTEQ6.6/CT10, MSTW08NLO, NNPDF2.0), check with FEWZ.

$$\Rightarrow \frac{\sigma^{\rm tot}_{W^+}}{\sigma^{\rm tot}_{W^-}} = \begin{cases} 1.418 \pm 0.008({\rm stat.}) \pm 0.022({\rm syst.}) \pm 0.029({\rm th.}) & (e) \\ 1.423 \pm 0.008({\rm stat.}) \pm 0.019({\rm syst.}) \pm 0.031({\rm th.}) & (\mu) \\ 1.421 \pm 0.006({\rm stat.}) \pm 0.014({\rm syst.}) \pm 0.030({\rm th.}) & (e+\mu) \end{cases}$$

- ATLAS [arXiv:1010.2130, ATLAS-CONF-2011-041] use PYTHIA (with MRSTLO*) $\Rightarrow \sigma^{\rm tot}_{W^+}/\sigma^{\rm tot}_{W^-} = 1.508 \pm 0.034$ $(e+\mu)$.

MC+PDF	A_{W^+}/A_{W^-}		$\sigma^{\rm tot}_{W^+}/\sigma^{\rm tot}_{W^-}$	
	e	μ	e	μ
PYTHIA+MRSTLO*	1.02	1.02	1.50	1.51
PYTHIA+CTEQ6.6	1.05	1.05	1.46	1.47
PYTHIA+HERAPDF1.0	1.03	1.04	1.48	1.49
MC@NLO+HERAPDF1.0	1.05	1.05	1.46	1.47
MC@NLO+CTEQ6.6	1.06	1.06	1.45	1.46

$$\frac{\sigma^{\rm tot}_{W^+}}{\sigma^{\rm tot}_{W^-}} = \frac{\sigma^{\rm fid}_{W^+}}{\sigma^{\rm fid}_{W^-}} \cdot \frac{A_{W^-}}{A_{W^+}}$$

Alternative acceptance calculations move ATLAS $\sigma^{\rm tot}_{W^+}/\sigma^{\rm tot}_{W^-}$ much closer to CMS.

- Precision requires better acceptance calculations and/or data to theory comparisons at level of fiducial cross section.

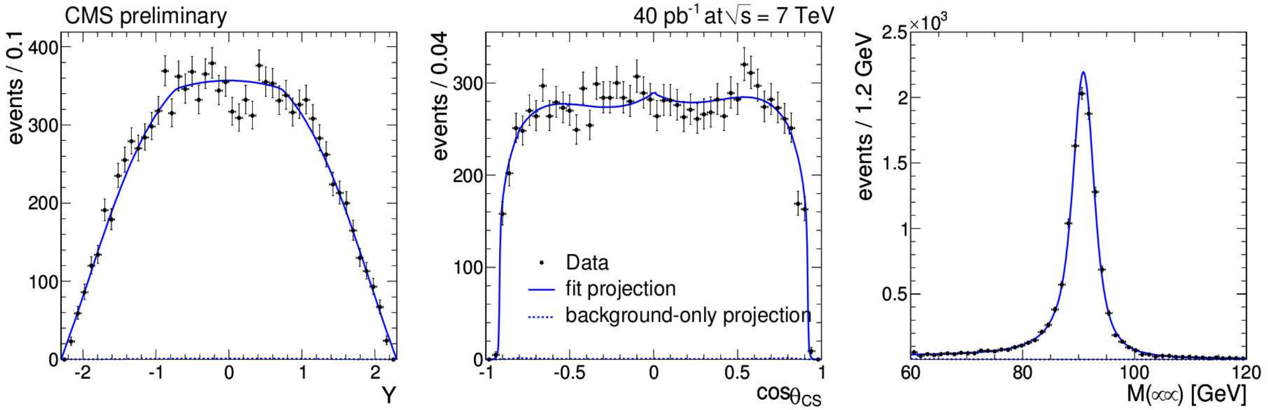
(G.Watt, 2011)

$\sin^2\theta_W$ FROM A_{FB} AT THE LHC?

- FORWARD/BACKWARD ASYMMETRY OF LEPTONS PRODUCED BY Z DECAY $\Rightarrow$ ELECTROWEAK MIXING ANGLE
- CURRENTLY $\sin^2\theta_W = 0.23153 \pm 0.00016$ FROM GLOBAL FIT BUT TWO MOST PRECISE MEASUREMENTS $\sin^2\theta_W = 0.23098 \pm 0.00026$ (SLD) & $\sin^2\theta_W = 0.23221 \pm 0.00029$ (B) DISAGREE BY THREE SIGMA
- $\rightarrow$ AIM AT 0.0001 LEVEL ACCURACY!
- IS IT POSSIBLE AT HADRON COLLIDER WHERE PDF UNCERTAINTIES PLAY A ROLE?

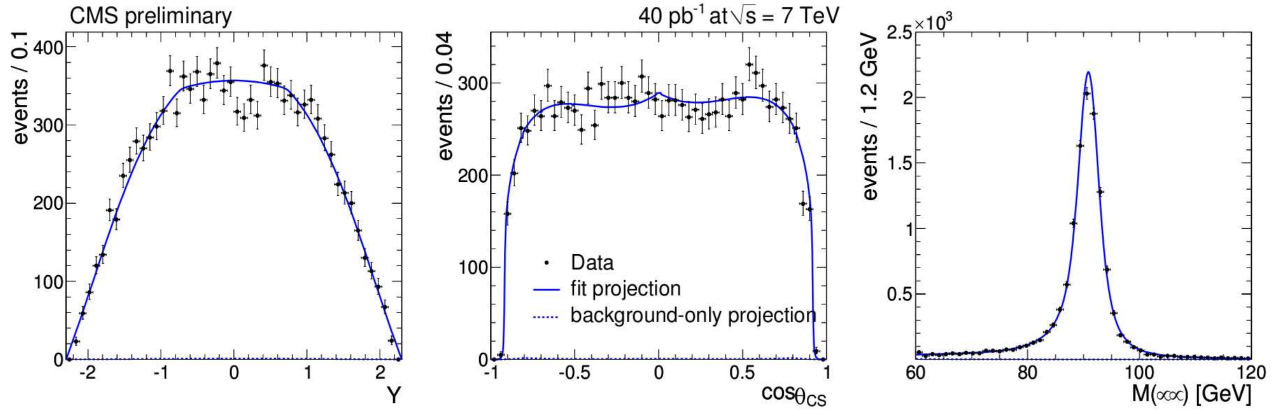
$\sin^2\theta_W$ FROM A_{FB} AT THE LHC?

- FORWARD/BACKWARD ASYMMETRY OF LEPTONS PRODUCED BY Z DECAY $\Rightarrow$ ELECTROWEAK MIXING ANGLE
- CURRENTLY $\sin^2\theta_W = 0.23153 \pm 0.00016$ FROM GLOBAL FIT BUT TWO MOST PRECISE MEASUREMENTS $\sin^2\theta_W = 0.23098 \pm 0.00026$ (SLD) & $\sin^2\theta_W = 0.23221 \pm 0.00029$ (B) DISAGREE BY THREE SIGMA
- $\rightarrow$ AIM AT 0.0001 LEVEL ACCURACY!
- IS IT POSSIBLE AT HADRON COLLIDER WHERE PDF UNCERTAINTIES PLAY A ROLE?
- MEASUREMENT PERFORMED BY CMS WITH 40 pb^{-1} $\Rightarrow$
  $\sin^2\theta_W = 0.2287 \pm 0.0077$ (STAT.) ± 0.0036 (SYSR.)

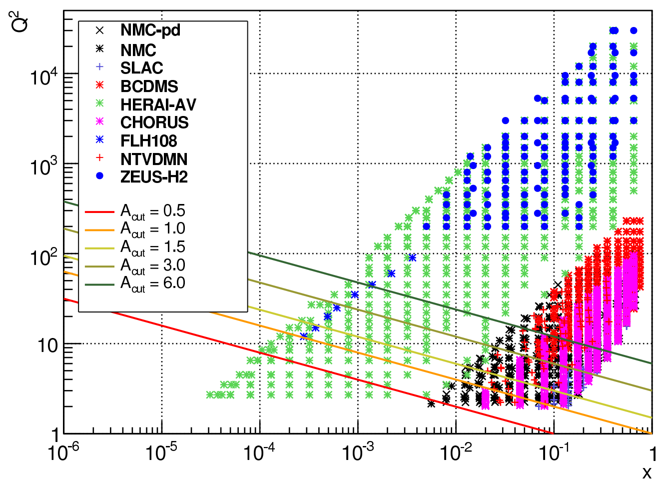


$\sin^2\theta_W$ FROM A_{FB} AT THE LHC?

- FORWARD/BACKWARD ASYMMETRY OF LEPTONS PRODUCED BY Z DECAY $\Rightarrow$ ELECTROWEAK MIXING ANGLE
- CURRENTLY $\sin^2\theta_W = 0.23153 \pm 0.00016$ FROM GLOBAL FIT BUT TWO MOST PRECISE MEASUREMENTS $\sin^2\theta_W = 0.23098 \pm 0.00026$ (SLD) & $\sin^2\theta_W = 0.23221 \pm 0.00029$ (B) DISAGREE BY THREE SIGMA
- $\rightarrow$ AIM AT 0.0001 LEVEL ACCURACY!
- IS IT POSSIBLE AT HADRON COLLIDER WHERE PDF UNCERTAINTIES PLAY A ROLE?
- MEASUREMENT PERFORMED BY CMS WITH 40 pb^{-1} $\Rightarrow$ $\sin^2\theta_W = 0.2287 \pm 0.0077$ (STAT.) ± 0.0036 (SYSR.)
- GREATER SENSITIVITY AT LARGE RAPIDITY $\Rightarrow$ LHCB PROPOSES MEASUREMENT FOR $2 \lesssim y \lesssim 4$ @ Z PEAK
- 0.0001 STAT. ACCURACY CAN BE ATTAINED WITH 50 fb^{-1}, BUT LARGE & SMALL x PDFS INVOLVED
- PDF+ THEORY (RESUMMATION) UNCERTAINTIES HARD!



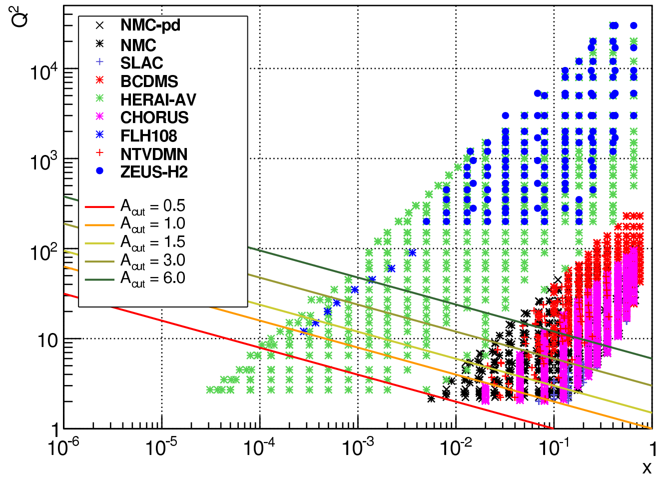
BEYOND DGLAP?
DISCOVERING A NEW QCD EFFECT IN HERA DATA



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

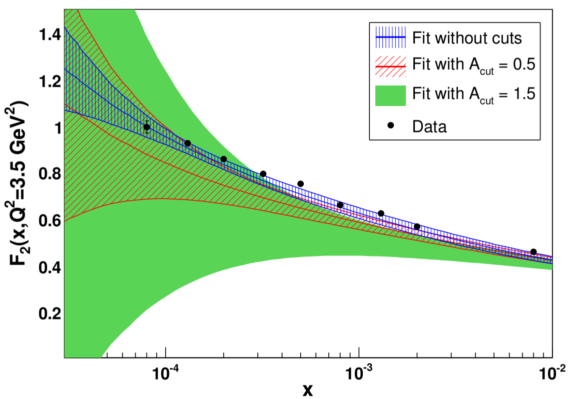


IDEA: (Gélis, 2008, $\Rightarrow$ Caola, s.f. ,Rojo 2010)

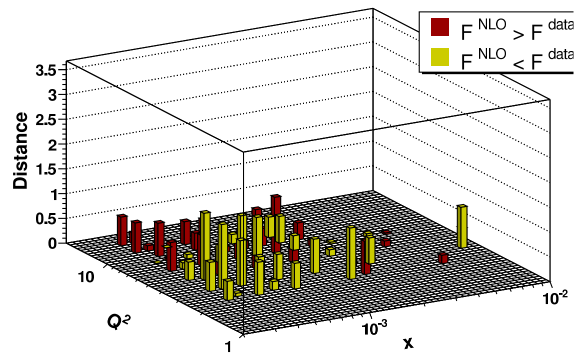
- CUT OUT DATA IN THE "DANGEROUS" (SMALL x) REGION
- DETERMINE PDFs IN THE "SAFE" (LARGE x AND Q^2) REGION
- EVOLVE BACKWARDS AND COMPARE TO DATA

OLD HERA DATA

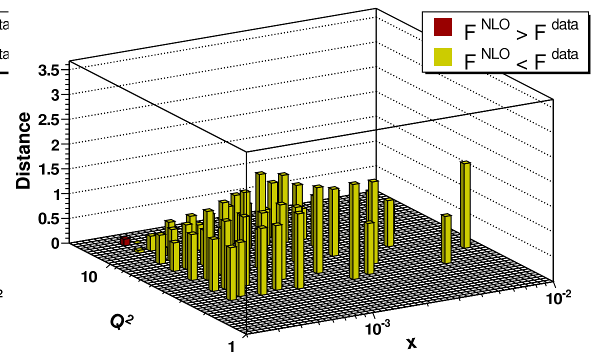
BACKWARD EV. VS DATA



DAT/TH DIST: NO CUT



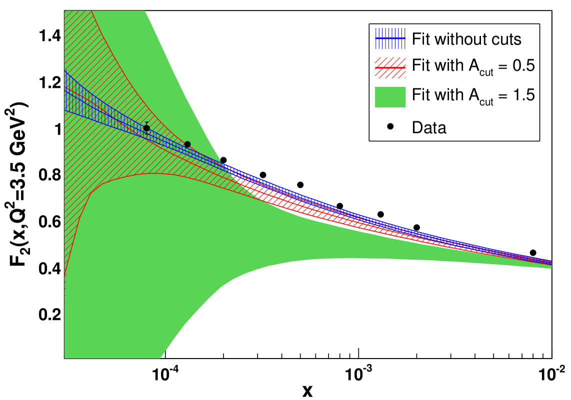
DAT/TH DIST: CUT



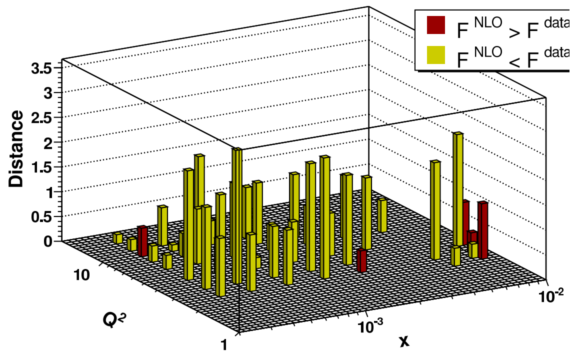
- BACKWARD EVOLVED FIT LIES SYSTEMATICALLY BELOW DATA
- DATA AT LOW x AND Q^2 SHOW LESS EVOLUTION THAN PREDICTED BY NLO DGLAP
- IF LOW x AND Q^2 DATA INCLUDED, THE FIT COMPENSATES READJUSTING PDFs

NEW COMBINED HERA DATA

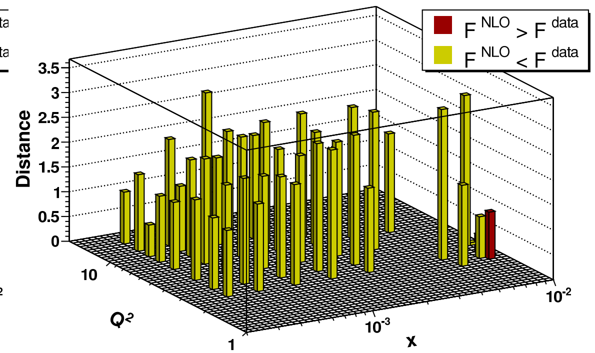
BACKWARD EV. VS DATA



DAT/TH DIST: NO CUT

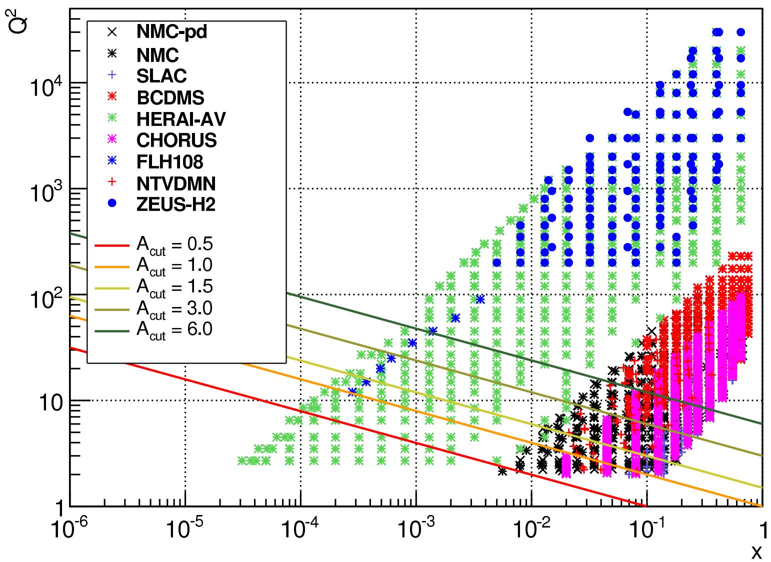


DAT/TH DIST: CUT

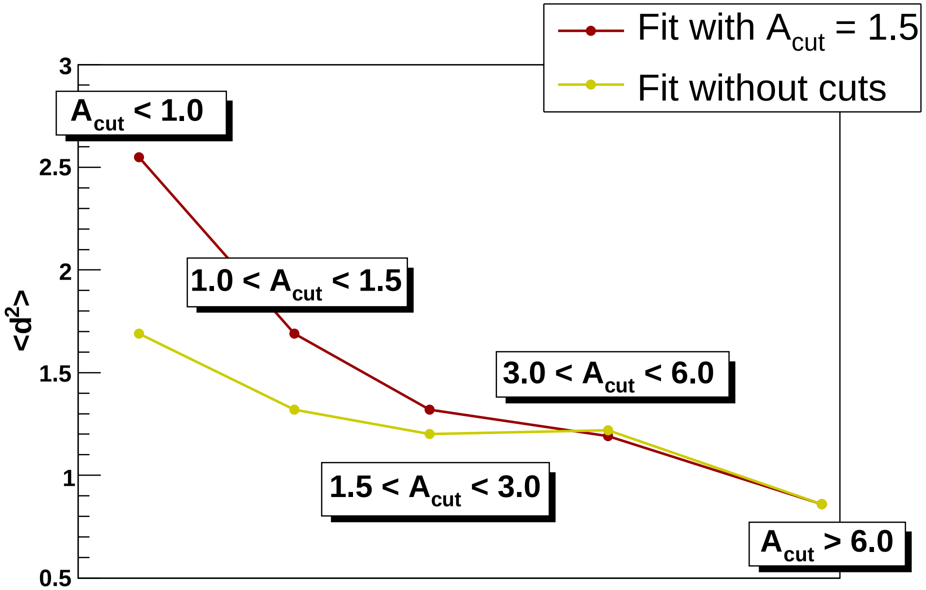


- DATA AT LOW x AND Q^2 SHOW LESS EVOLUTION THAN PREDICTED BY NLO DGLAP
- BACKWARD EVOLVED FIT LIES SYSTEMATICALLY BELOW DATA
- WITH MORE PRECISE DATA, THE FIT NO LONGER MANAGES TO COMPENSATE BY READJUSTING THE PDFs: EVEN FULL FIT LIES BELOW DATA

DETERIORATION IN FIT QUALITY:

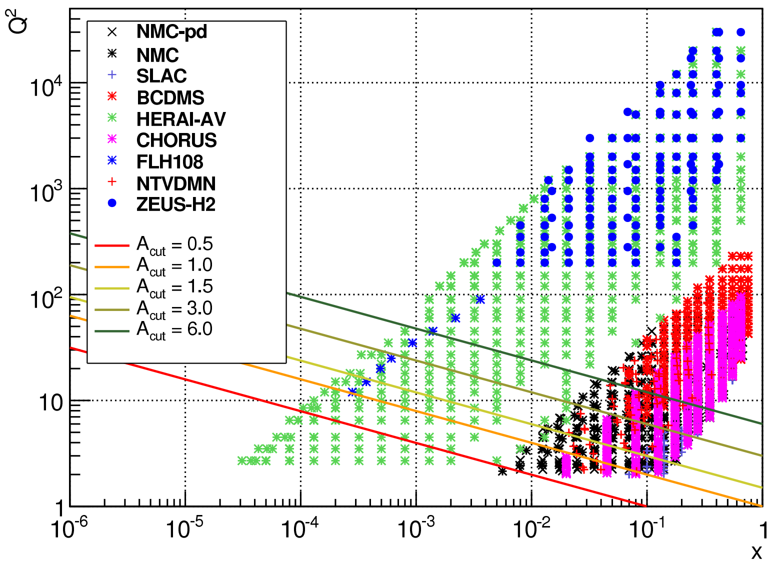


$\langle d^2 \rangle$ VS x SLICES

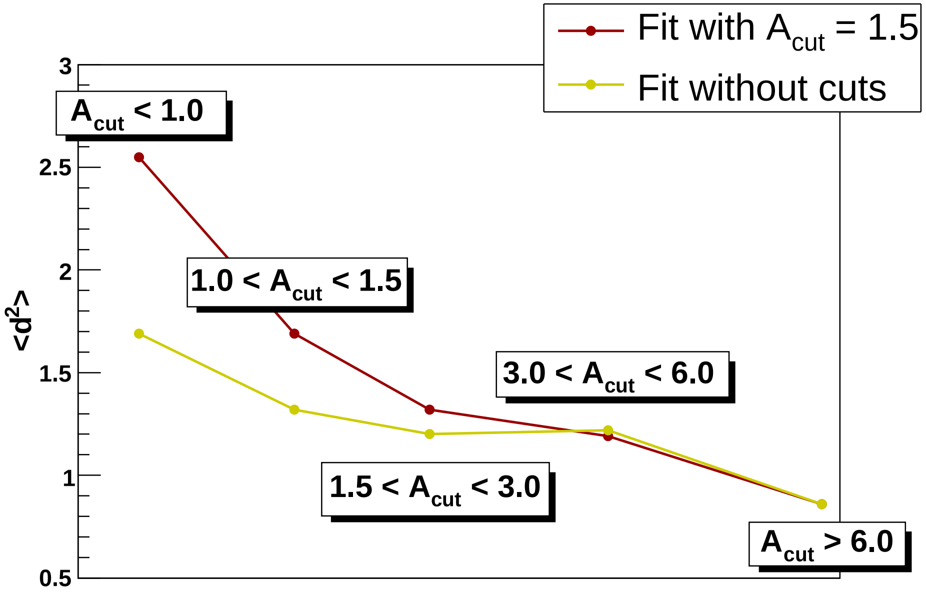


- QUALITY OF UNCUT FIT DETERIORATES IN LOW x REGIONS
- QUALITY OF CUT FIT INCREASINGLY POOR AS x DECREASES
- DISTANCE RISES DESPITE HUGE INCREASE IN UNCERTAINTY

DETERIORATION IN FIT QUALITY:

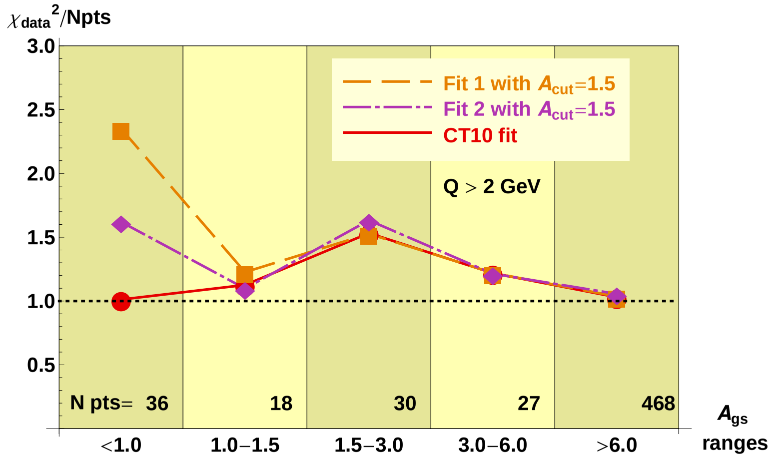


$\langle d^2 \rangle$ VS x SLICES



- 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



THE STANDARD (HESSIAN) SOLUTION

(MSTW, CTEQ):

FUNCTIONAL PARTON FITTING

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

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

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

  - CTEQ/TEA: 22$\rightarrow$ 26 PARMS.

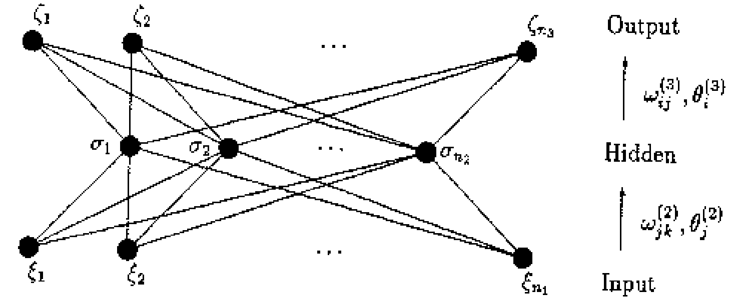
$$x\,f(x,Q_0) = a_0\,x^{a_1}\,(1-x)^{a_2}\,\exp\left(a_3x + a_4x^2 + a_5\sqrt{x} + a_6x^{-a_7}\right) \text{ (7 indep. functions)}$$

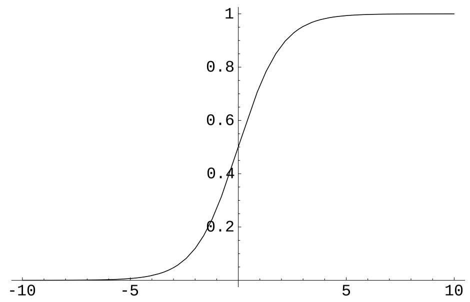
a_6, a_7 ONLY USED FOR GLUON

  - 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')

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

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

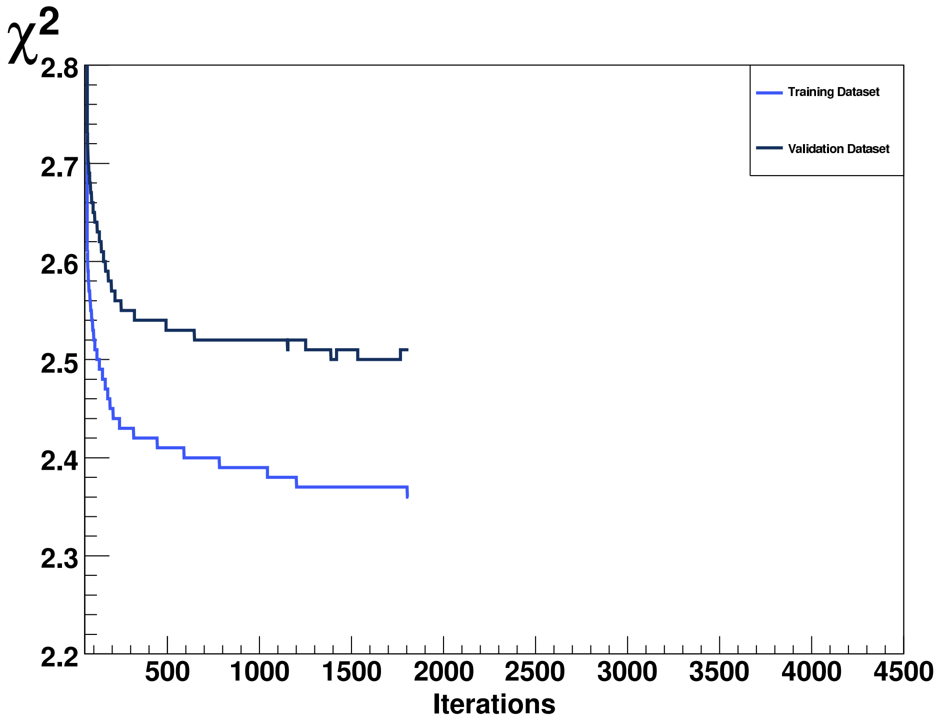
- ANY FUNCTION CAN BE REPRESENTED BY A SUFFICIENTLY BIG NEURAL NETWORK $\Rightarrow$ CAN CHECK INDEPENDECE OF SIZE
- MINIMZATION PERFORMED BY GENETIC ALGORITHM
- INFORMATION CONTAINED ("COARSENESS") IN NN INCREASES DURING MINIMIZATION

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

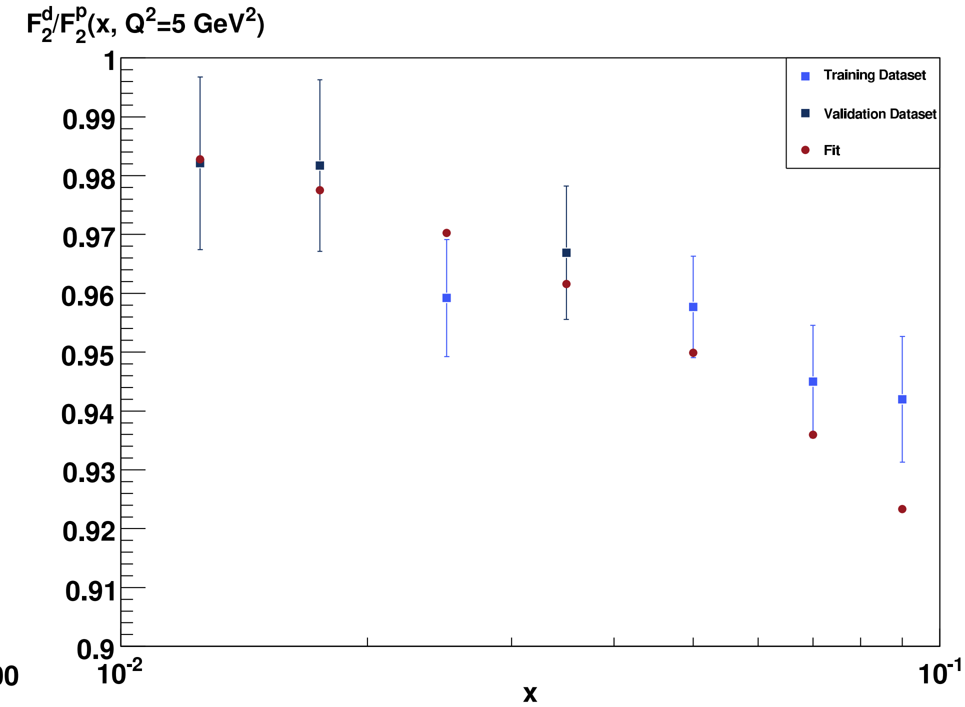
- REPLICAS ARE FITTED TO A DATA SUBSET
- A DIFFERENT SUBSET OF DATA USE FOR EACH REPLICA
- OPTIMAL FIT WHEN FIT TO VALIDATION (CONTROL) DATA STOPS IMPROVING
-

OPTIMAL FITTING

χ^2



FIT TO DATA

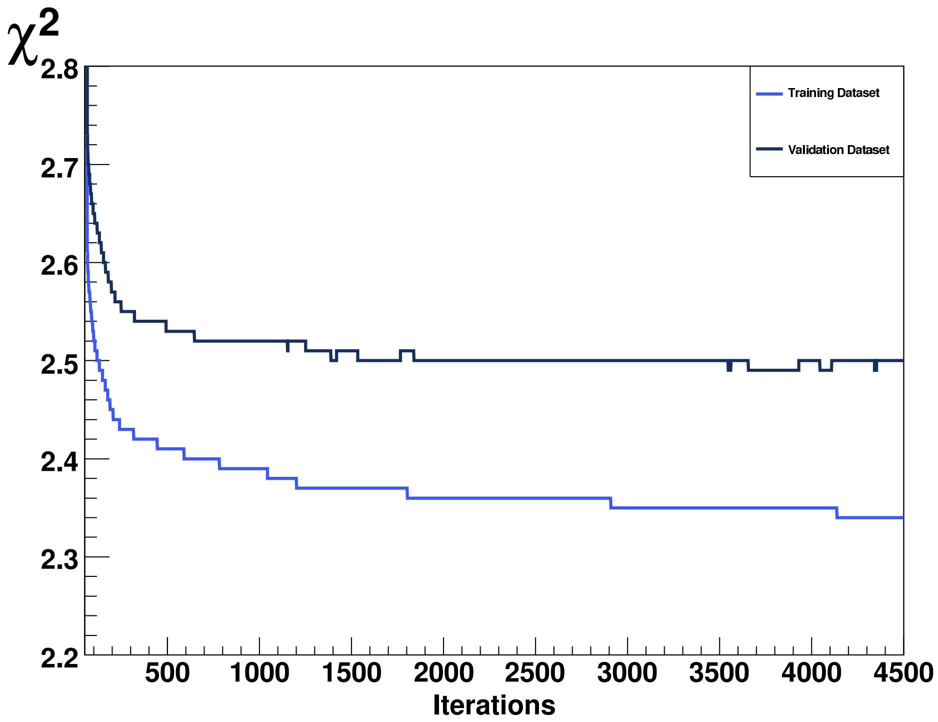


DATA MONTE CARLO $\Rightarrow$ PDF MONTE CARLO
CROSS-VALIDATION

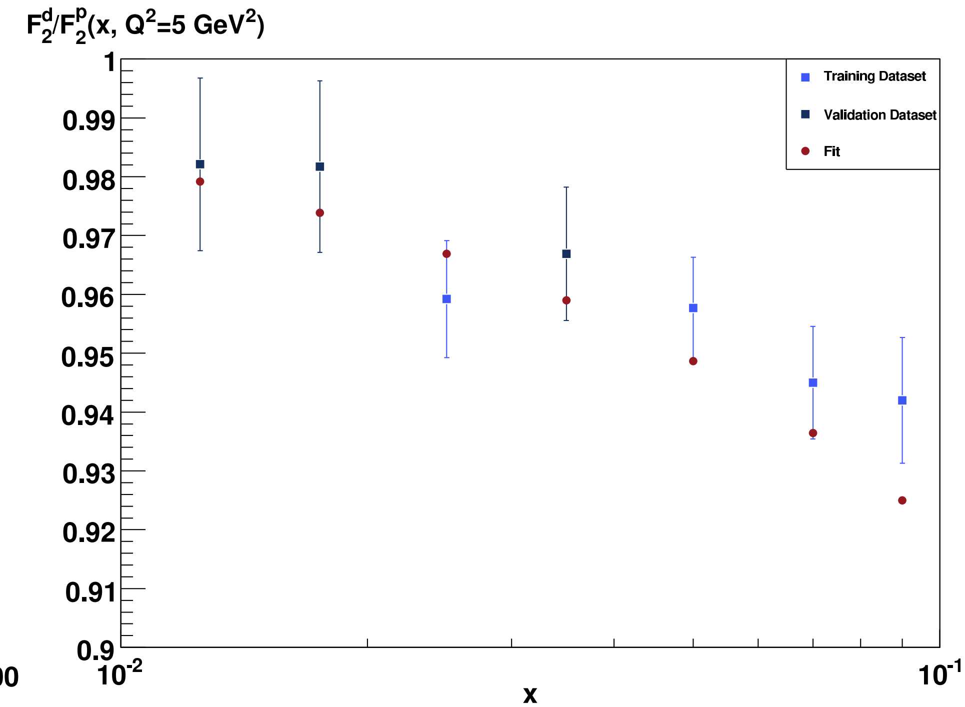
- REPLICAS ARE FITTED TO A DATA SUBSET
- A DIFFERENT SUBSET OF DATA USE FOR EACH REPLICA
- OPTIMAL FIT WHEN FIT TO VALIDATION (CONTROL) DATA STOPS IMPROVING
- THE BEST FIT IS NOT AT THE MINIMUM OF THE χ^2

OVERFITTING

χ^2



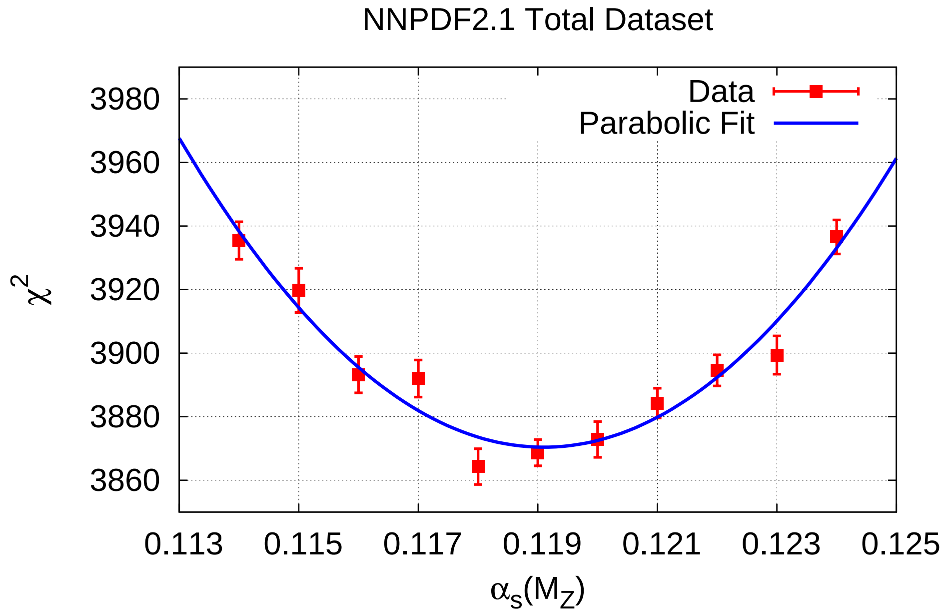
FIT TO DATA



α_s FROM THE NNPDF2.1 GLOBAL FIT

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

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



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

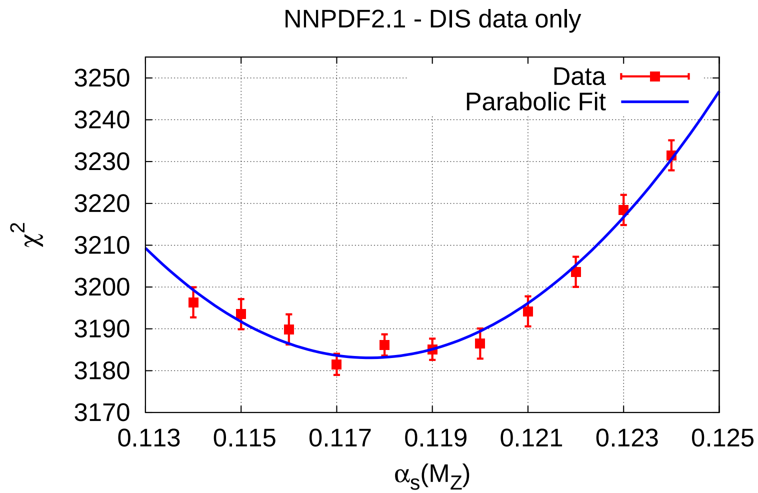
DIS ONLY

DO DIS DATA PREFER A SMALLER VALUE?

NLO NNPDF2.1 DEEP-INELASTIC DATA

$N_{\rm rep} = 500$ PER VALUE OF α_s

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

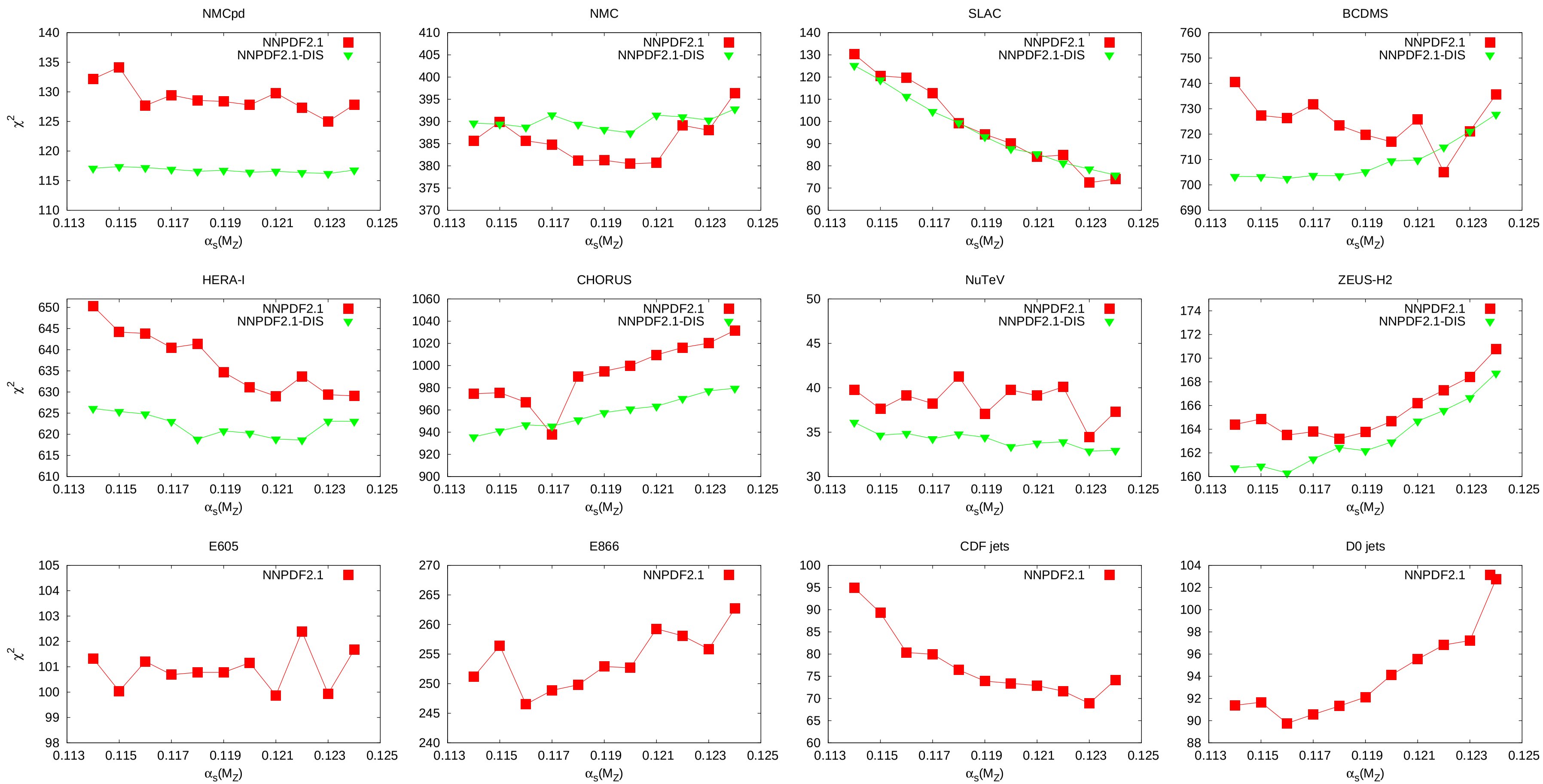


$$\chi^2/\text{d.o.f.} = 0.8 \text{ for the parabolic fit}$$

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

INDIVIDUAL EXPERIMENTS: WHAT'S GOING ON?

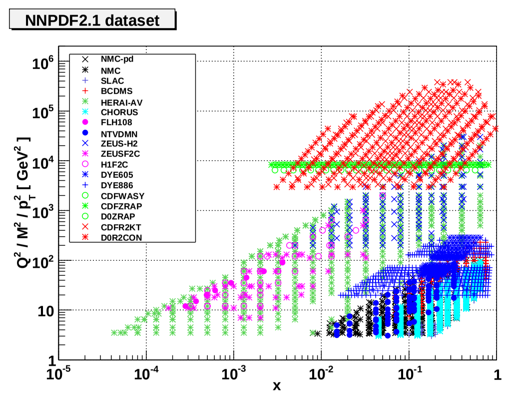
χ^2 PROFILE FOR INDIVIDUAL EXPERIMENTS IN GLOBAL & DIS FITS



- 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

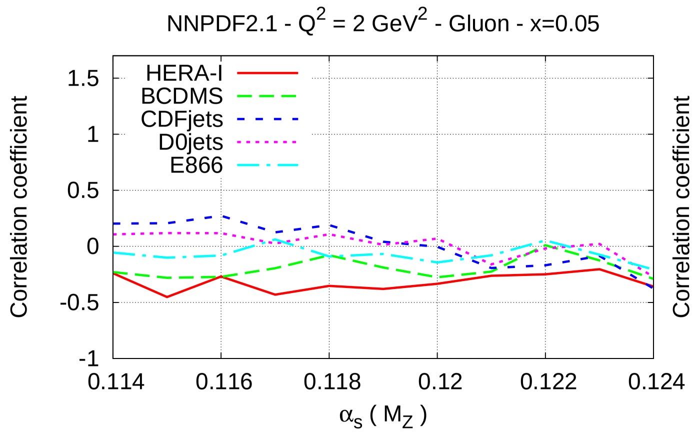
IS THERE A PROBLEM WITH DIS?

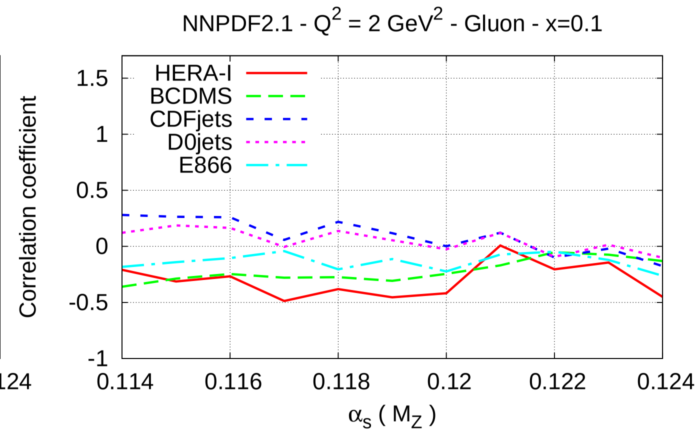
CONJECTURE



- 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

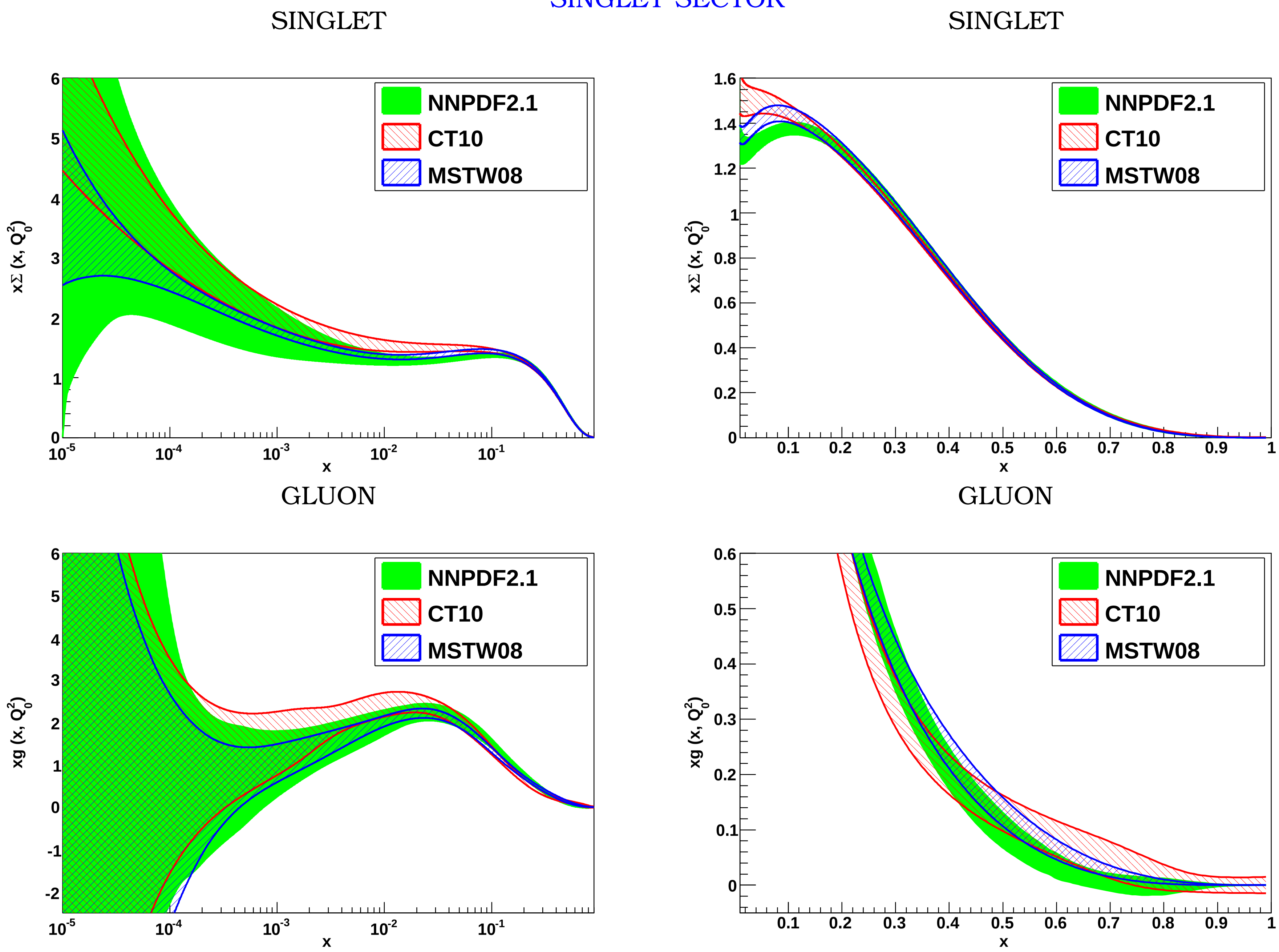




- COMPUTE THE CORRELATION BETWEEN χ^2 & PDFs AT THE GLOBAL BEST FIT LARGE (POSITIVE OR NEGATIVE) CORRN. $\Leftrightarrow$ RUNAWAY DIRECTION
- OPPOSITE SIGN CORRELATIONS $\Leftrightarrow$ DATA PULLING IN OPPOSITE DIRECTIONS
- AT LOW $\alpha_s = 0.114$ DIS FIT HAS A RUNAWAY DIRECTION STABILIZED BY JETS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

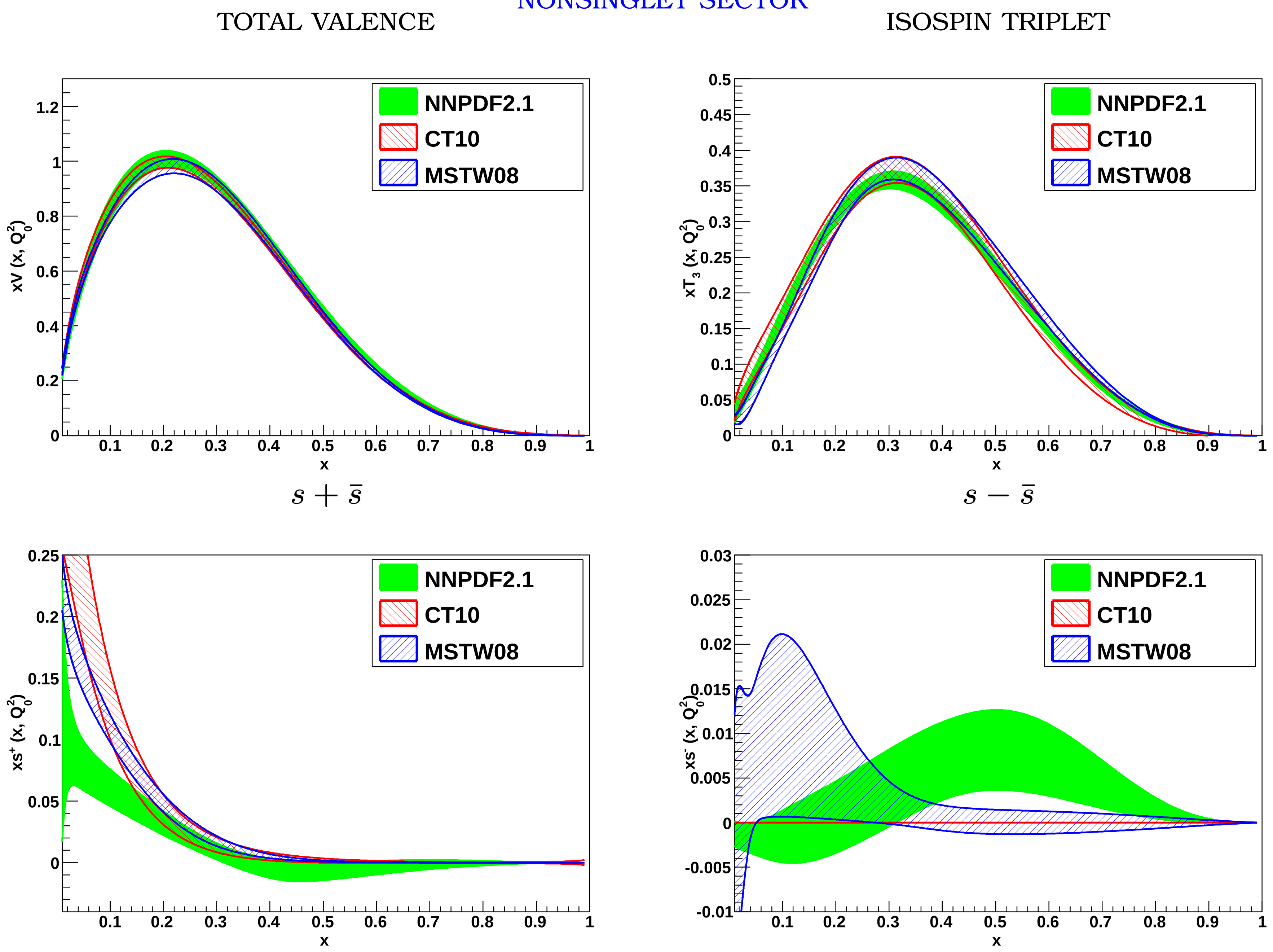
NNPDF2.1, MSTW08, CT10 COMPARISON

SINGLET SECTOR



NNPDF2.1, MSTW08, CT10 COMPARISON

NONSINGLET SECTOR



MONTE CARLO DATA GENERATION

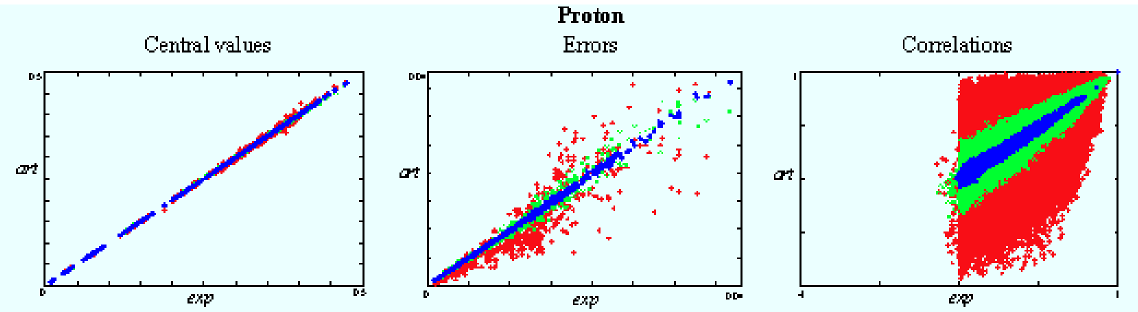
- BCDMS+ NMC PROTON & DEUTERON F_2 DATA (FULL CORRELATED SYSTEMATICS AVAILABLE), TAKEN AT 4 BEAM ENERGIES
- ON TOP OF STAT. ERRORS, 4 SYSTEMATICS + 1 NORMALIZATION (NMC) OR 6 SYSTEMATICS + 1 ABSOLUTE & 2 RELATIVE NORMALIZATIONS (BCDMS), WITH VARIOUS FORMS OF CORRELATION (FULL, OR FOR EACH TARGET, OR 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

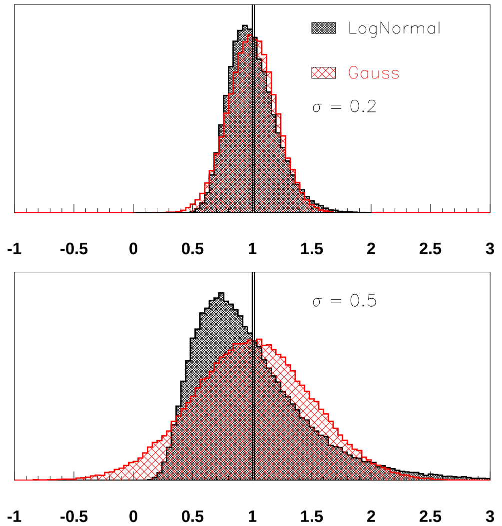


SCATTER PLOT ART. VS. EXP. FOR 10 (RED) 100 (GREEN) AND 1000 (BLUE) REPLICAS

NEED 1000 REPLICAS TO REPRODUCE CORRELATIONS TO PERCENT ACCURACY

PARAMETRIZATION UNCERTAINTIES?
NONGAUSSIAN BEHAVIOUR?

LOGNORMAL VS. GAUSSIAN



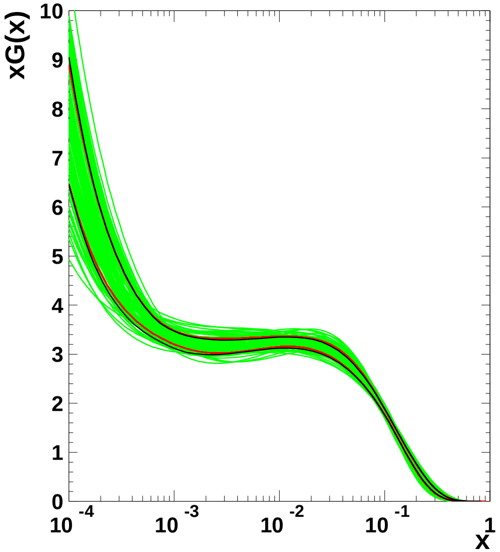
THE HERALHC BENCHMARK

(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

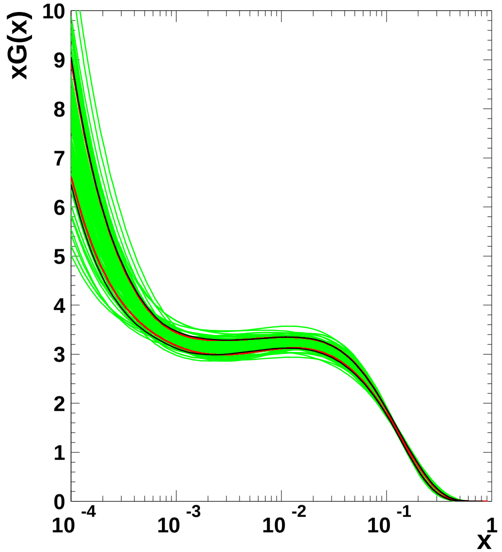
LOGNORMAL: HESS. VS MC

Fit vs H1PDF2000, Q^2 = 4. GeV^2



GAUSSIAN: HESS. VS MC

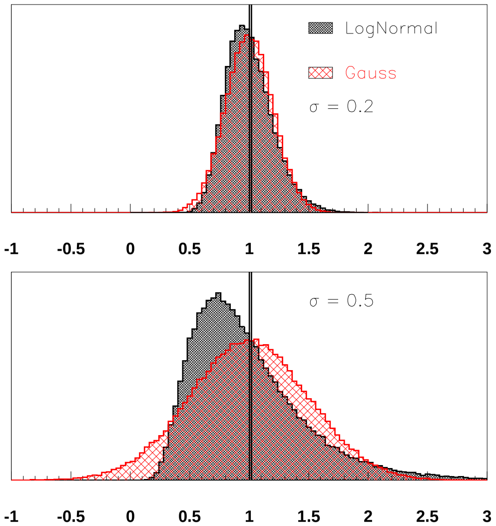
Fit vs H1PDF2000, Q^2 = 4. GeV^2



- NO DIFFERENCE BETWEEN LOGNORMAL, GAUSSIAN, MC, HESSIAN

PARAMETRIZATION UNCERTAINTIES?
NONGAUSSIAN BEHAVIOUR?

LOGNORMAL VS. GAUSSIAN

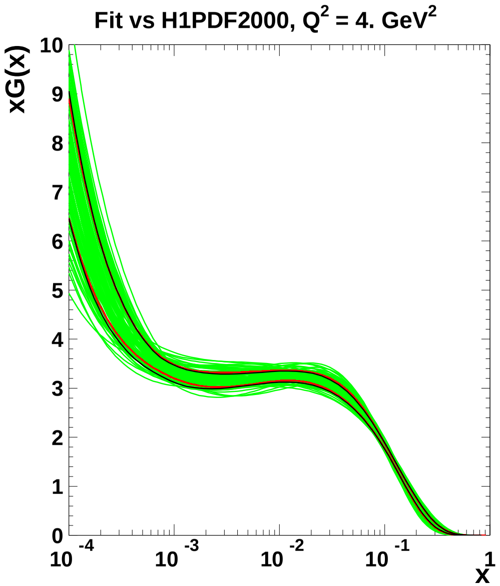


THE HERALHC BENCHMARK

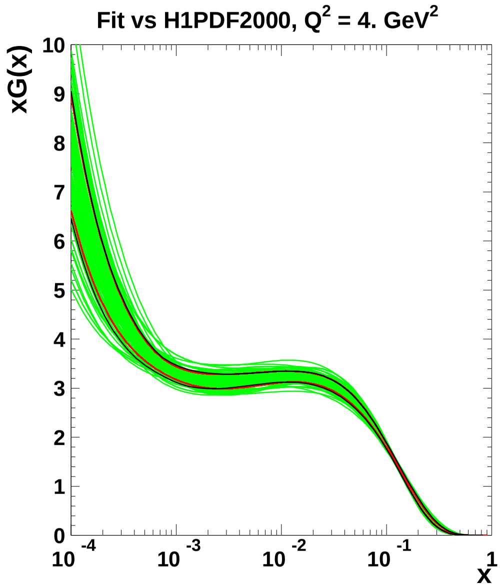
(Feltesse, Glazov, Radescu + NNPDF 2008)

- 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

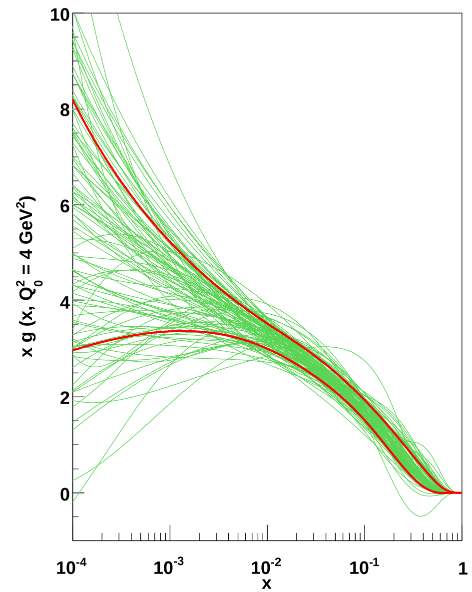
LOGNORMAL: HESS. VS MC



GAUSSIAN: HESS. VS MC



NNPDF

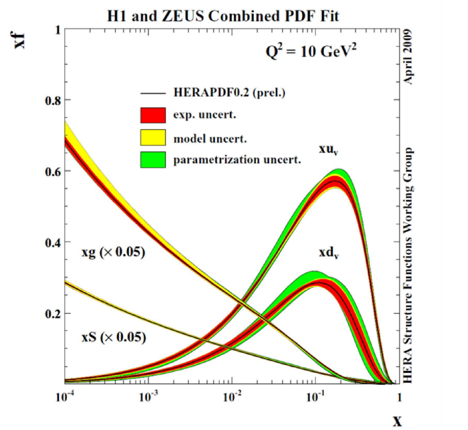


- NO DIFFERENCE BETWEEN LOGNORMAL, GAUSSIAN, MC, HESSIAN
- SIZABLE DIFFERENCE WR TO FLEXIBLE NNPDF PARAMETRIZATION

PARAMETRIZATION UNCERTAINTIES?

EXPLORING THE SPACE OF PARAMETERS: HESSIAN APPROACH

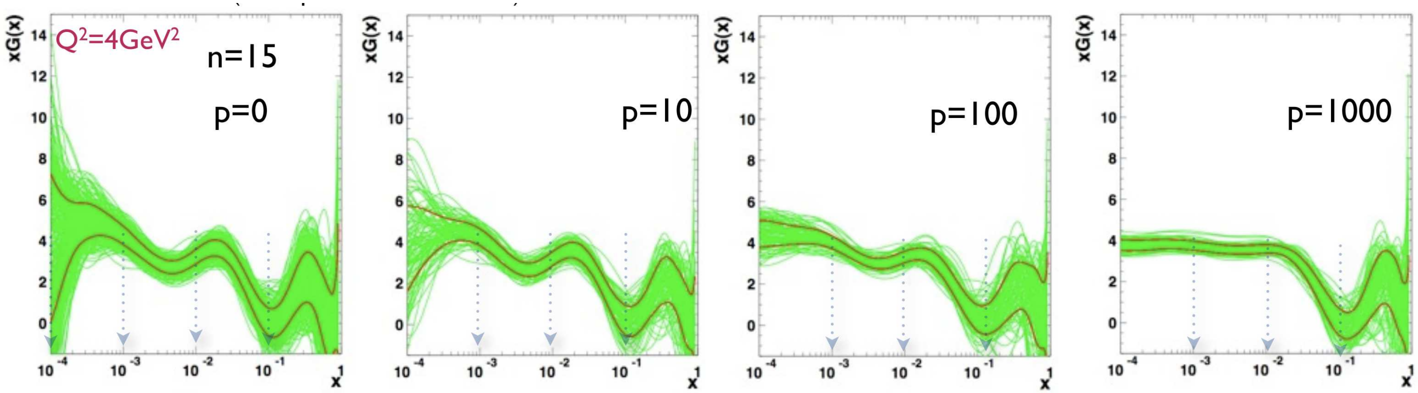
- IN HESSIAN APPROACH CAN VARY THE FUNCTIONAL FORM, ASSUMPTIONS, STARTING SCALE
- DONE IN THE HERAPDF1.0 FIT: VARIATION OF STRANGENESS FRACTION, LARGE x BEHAVIOUR, HIGHER ORDER POLYNOMIAL TERMS
- NO TOLERANCE ($\Delta\chi^2 = 1$), UNCERTAINTY DOUBLED



- 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



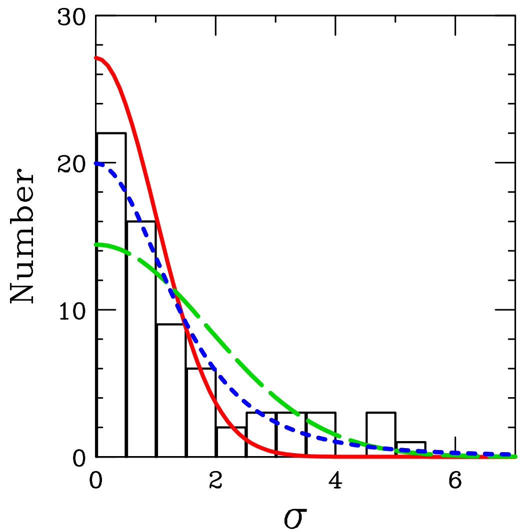
(Glazov, Radescu, 2009)

WHERE IS THE UNCERTAINTY COMING FROM?

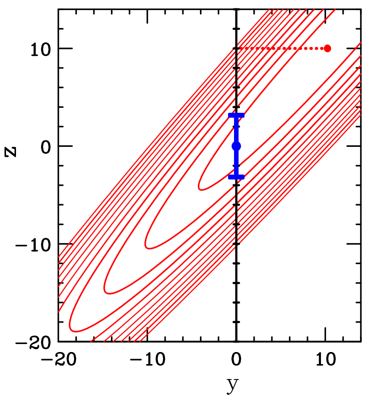
WHY DOES ONE NEED LARGE TOLERANCES?

DATA INCOMPATIBILITY (Pumplin, 2009)

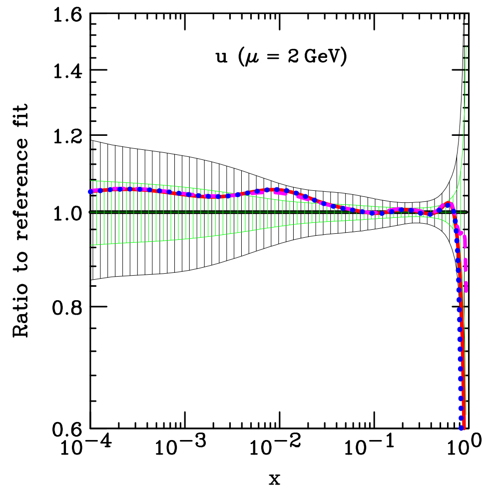
- CAN "REDIAGONALIZE": DIAGONALIZE SIMULTANEOUSLY χ^2 FOR TOTAL AND i–TH EXPT $\Rightarrow$ COMPATIBILITY OF EACH EXPT WITH GLOBAL FIT
- STUDY DISTRIBUTION OF DISCREPANCIES
- APPROX. GAUSSIAN WITH UNCERTAINTIES RESCALED BY 2 $\Rightarrow$ $\Delta\chi^2 \sim 10$ FOR 90%C.L.



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.



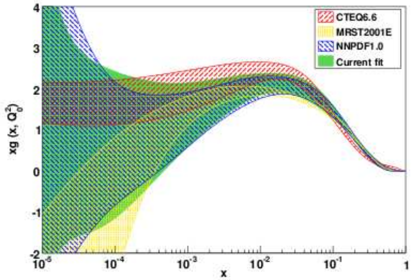
WHERE IS THE UNCERTAINTY COMING FROM?

FIT TO REPLICAS VS RANDOM SUBSET OF CENTRAL VAL.S

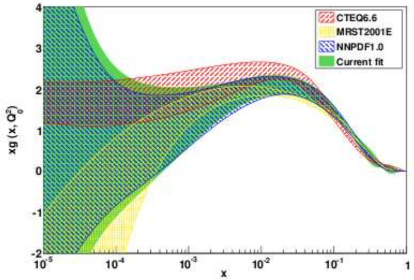
	REPLICAS	CENTRAL V.
χ^2	1.32	1.32
$\langle\chi^2\rangle_{\rm rep}$	2.79 ± 0.24	1.65 ± 0.20
$\langle\sigma^{\rm dat}\rangle$	0.039	0.035

GLUE

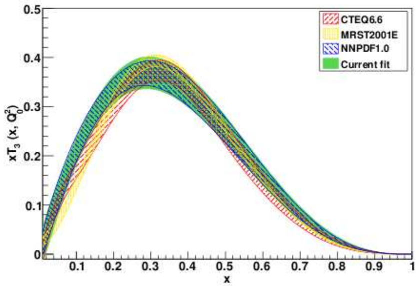
replicas

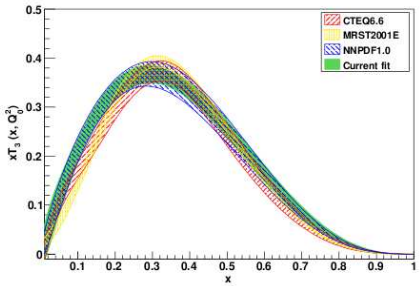


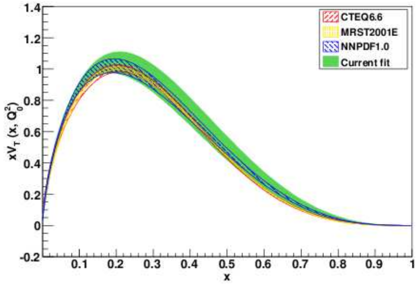
c. vals.

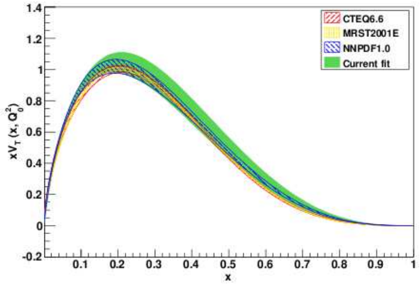


LIGHT QUARKS

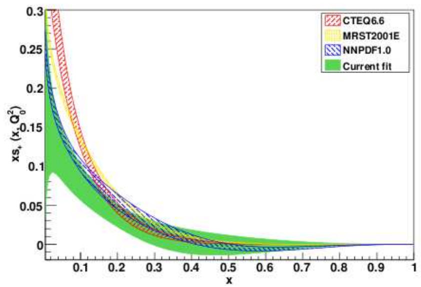


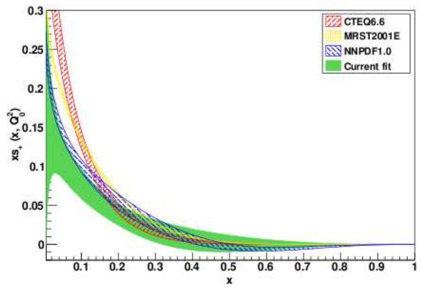






STRANGE





- 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

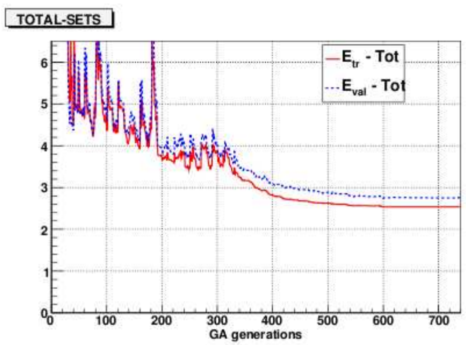
WHERE IS THE UNCERTAINTY COMING FROM?

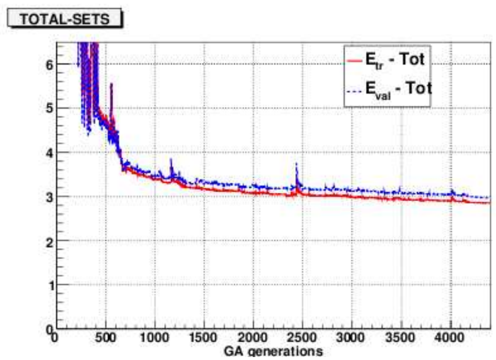
CENTRAL VALUES: VARYING PARTITION VS FIXED PARTITION

	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

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

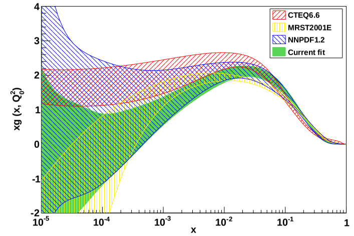




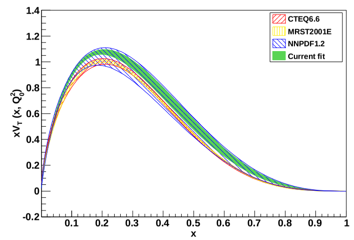
FUNCTIONAL UNCERTAINTY

- 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.

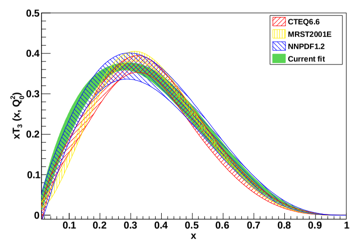
GLUE



VALENCE



TRIPLET



STRANGE