NNPDF

CLOSURE TESTS FOR PARTON DISTRIBUTIONS

STEFANO FORTE

UNIVERSITÀ DI MILANO & INFN

FOR THE COLLABORATION: R. D. BALL, V. BERTONE, S. CARRAZZA, C. DEANS, L. DEL DEBBIO, S.F., A. GUFFANTI, N. HARTLAND J. I. LATORRE, J. ROJO, M. UBIALI

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

INFN Istituto Nazionale di Fisica Nucleare

PDF4LHC

CERN, DECEMBER 13, 2013

TOWARDS NNPDF3.0

OPTMIZATION

- FULL MIGRATION OF THE CODE TO C++
- STREAMLINING, OPTIMIZATION AND DEBUGGING OF THE WHOLE CODE INCLUDING GENETIC ALGORITHM, NUMERICAL METHODS, ETC
- FAST INTERFACES `FastKernel + APPLGRID/FASTNLO` USED SYSTEMATICALLY
- $\Rightarrow$ MORE DETAILED MINIMIZATION
  TYPICAL FIT: $\sim$ 4000 DATAPOINTS FOR 50000 ITERATIONS OF A GA WITH 80 MUTANTS:
  $\sim 10^{10}$ PREDICTIONS COMPUTED FOR EACH REPLICA
  (50H COMPUTING PER REPLICA ON CERN `lxplus`)

NEW DATA:

ALL LHC DATA WITH INFO ON SYSTEMATICS

- ATLAS: HIGH-MASS DRELL-YAN (2011); JETS 2.76 TEV
- CMS: W μ ASYM 5 fb^{-1}; CMS W+CHARM 5 fb^{-1}; DOUBLE-DIFFERENTIAL DRELL-YAN; INCLUSIVE JETS 5 FB^{-1}
- LHCB $Z \to e^+e^-$ RAPIDITY DISTN. (2011)
- HERA II COMBINED F_2^c
- H1 HERA-II INCLUSIVE F_2
- ZEUS HERA-II INCLUSIVE F_2
- ATLAS W p_T DISTN; ATLAS PROMPT PHOTON; LHCB $Z \to \mu^+\mu^-$ RAPIDITY DIST.; ATLAS+CMS TOP RAPIDITY DISTN.: UNDER CONSIDERATION (PRELIMINARY DATA, INTERFACES BEING DEVELOPED &C...)

MINIMIZATION STRATEGY BASED ON A CLOSURE TEST

CLOSURE TESTS

WHAT IS A CLOSURE TEST?

- ASSUME UNDERLYING PDFS KNOWN
- GENERATE DATA WITH GIVEN STATISTICAL AND CORRELATED SYSTEMATICS
- PERFORM A FIT & COMPARED TO "TRUTH"
- PREVIOUS STUDIES BY THORNE & WATT (2012) ALONG SIMILAR LINES

LEVELS

- DATA ARE GENERATED FOR THE SAME KINEMATICS OF ALL DATA IN NNPDF2.3 USING UNDERLYING MSTW08 PDFS (CT10 ALSO TRIED)
- LEVEL 0:
  - EACH DATAPOINT EQUAL TO THE MSTW "TRUE VALUE"; UNCERTAINTY ASSUMED TO COINCIDE WITH THE EXPERIMENTAL ONE
  - FIT → MUST FIND $\chi^2 = 0$ (GET BACK MSTW "TRUTH")
- LEVEL 2:
  - EACH DATAPOINT IS OBTAINED AS A RANDOM FLUCTUATION WITH GIVEN COVARIANCE MATRIX ABOUT MSTW "TRUTH"
  - GENERATE PSEUDODATA REPLICAS OF THESE "DATA"
  - THEN FIT PDF REPLICAS TO PSEUDODATA REPLICAS
  - FIT MUST FIND (PER DATAPOINT) $\chi^2 = 1$ (best-fit to data); $\langle E \rangle = 2$ (fit of each replica to data replica); $\langle \chi^{2\,(1)} \rangle = 1$ (fit of each replica to data)
  - MUST FIND THAT (PREDICTION)-(THEORY) IS COMPATIBLE WITH ZERO WITHIN ERRORS
  - MUST FIND THAT MSTW "TRUE PDFS" IS WITHIN ONE σ BAND IN 68% OF CASES

(LEVEL 1: SAME AS LEVEL 2, BUT WITHOUT PSEUDODATA REPLICAS)

STOPPING VS. WEIGHT PENALTY

- NNPDF OPTIMAL FIT CURRENTLY DETERMINED BY CROSS-VALIDATION: DATA RANDOMLY DIVIDED IN TWO SETS, χ^2 OF FITTED (TRAINING) DATASET KEEPS DECREASING BUT χ^2 OF NON-FITTED (VALIDATION) DATASET STARTS INCREASING
- MUST INTRODUCE THRESHOLDS FOR INCREASE & DECREASE BASED ON TYPICAL χ^2 FLUCTUATIONS
- ALTERNATIVE IDEA: INTRODUCE A MEASURE OF THE COMPLEXITY OF THE j-TH NN: $\Delta_j = \sum_{i=1}^{N_w} (w_i^j)^2$
- THEN ADD TO χ^2 A WEIGHT-PENALTY $f(w_i) = \sum_{j=1}^{N_{pdfs}} \alpha_j \Delta_j$ AND MINIMIZE χ^2
- CONSTANTS α_j DETERMINED BY EXPECTED COMPLEXITY OF THE j-TH NETWORK BASED ON PREVIOUS FIT: $\alpha_i = \left[\frac{\langle \Delta_i \rangle}{N_w}\right]^{-1}$
- ITERATE UNTIL CONVERGENCE
- FITS STOPS WHEN NETWORKS FIT THE DATA BUT ARE NOT TOO COMPLEX

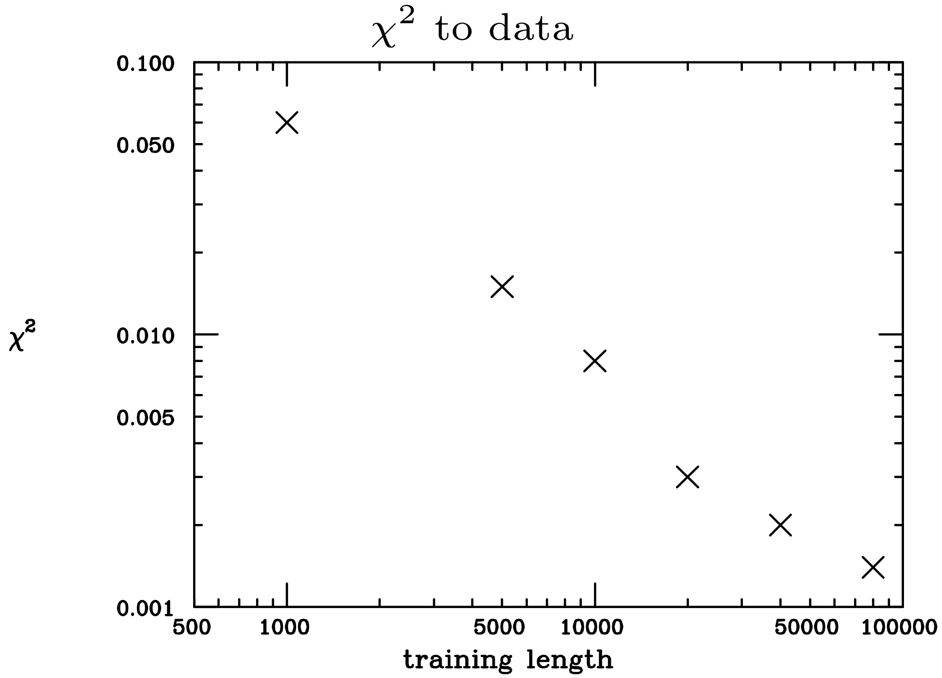
ADVANTAGES

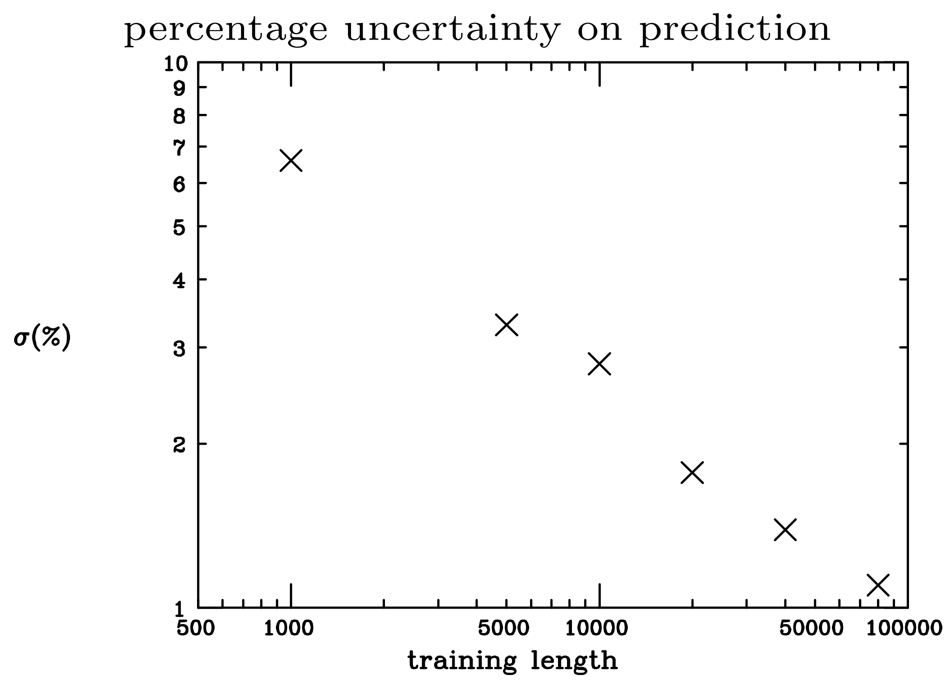
- OPTIMAL WEIGHTS DETERMINED SELF-CONSISTENTLY
- NO OVERLEARNING $\rightarrow$ NO STOPPING CRITERION NEEDED (JUST MAKE FIT LONG ENOUGH)
- NATURALLY SMOOTH PDF SHAPES

LEVEL-0 CLOSURE

- FITS PRODUCED WITH INCREASING (FIXED) TRAINING LENGTH
- ALL FITS WITH SAME DATA AND SAME RANDOM SEED (RANDOM SEED INDEP SEPARATELY TESTED)
- GOODNESS OF FIT TO DATA AND PERCENTAGE UNCERTAINTY ON PREDICTION STUDIED VS. TRAINING LENGTH
- χ^2 MUST GO TO ZERO; σ MUST GO TO ZERO AT DATA LEVEL, NOT AT PDF LEVEL

FIT QUALITY





PERFECT FIT!

LEVEL-2 CLOSURE

FIXED-LENGTH FITS TO 100% OF DATA (NO CROSS-VALIDATION)

FIT QUALITY

- AT 10K GA ITERATIONS, $\chi^2 = 0.96$, $\langle E \rangle = 2.0$ (NOTE $\chi^2_{mstw} = 0.96$)
- CHECKED AGAIN AT 20K, 40K, 80K: SAME PERFECT VALUES (TO TWO DECIMAL PLACES)

AGREEMENT WITH THEORY (DATA LEVEL)

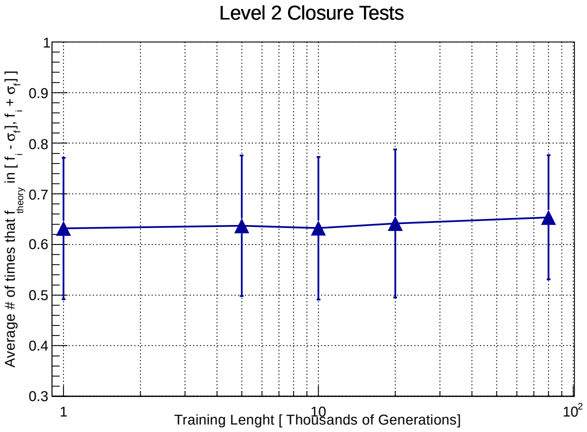
- COMPARE PREDICTION TO THEORY $$\frac{\left(F_i^{(0),\mathrm{nnpdf}} - F_i^{\mathrm{mstw}}\right)^2}{\left(\Delta F_i^{\mathrm{nnpdf}}\right)^2}$$
- AVERAGE OVER DATAPOINTS, DIFFERENT TL & METHODS: PERFECT

	NO WP	WP
10K	0.948 ± 0.854	0.638 ± 0.714
20K	0.966 ± 0.916	0.613 ± 0.732
40K	1.01 ± 1.00	0.622 ± 0.788

AGREEMENT WITH THEORY (PDF LEVEL)

- COMPARE PDF TO THEORY $\left[q_l^{(k)}(x_i) - \Delta q_l^{(k)}(x_i), q_l^{(k)}(x_i) + \Delta q_l^{(k)}(x_i)\right]$ AT $x = 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7$ FOR ALL PDFs
- FRACTION WHICH FALLS WITHIN ONE σ
- NO WP SHOWN, WP SIMILAR: PERFECT

fraction within one σ



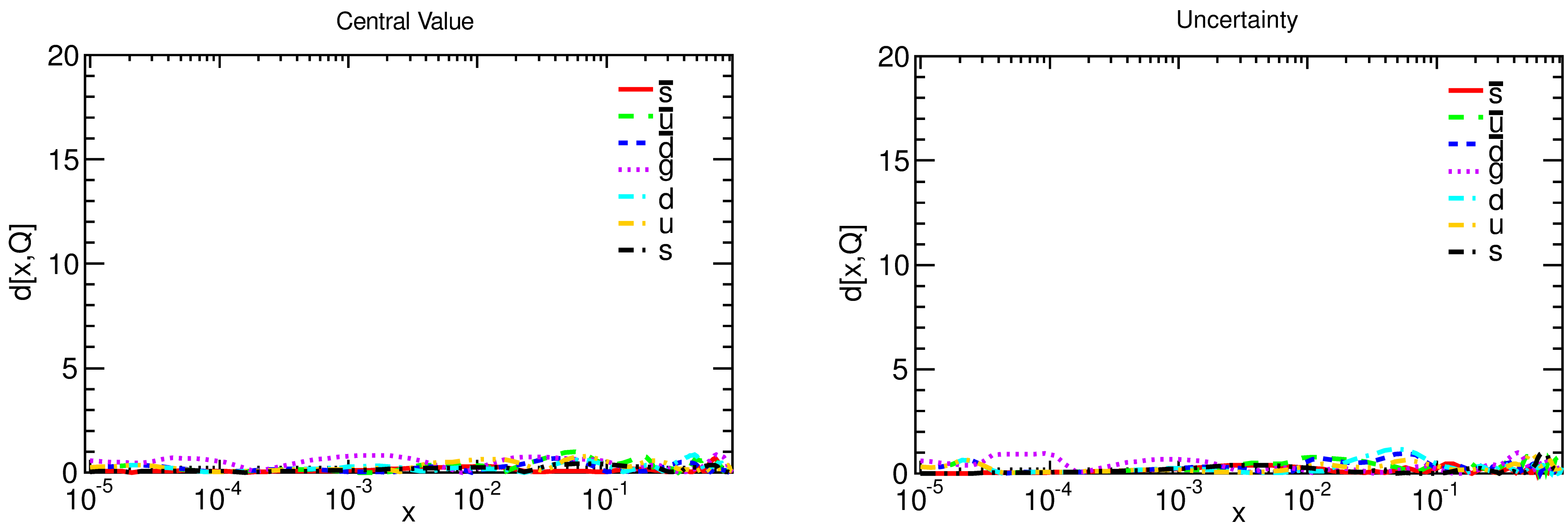
OVERLEARNING

- IT LOOKS LIKE AT 10K A PERFECT FIT IS REACHED
- NO OVERLEARNING → NO STOPPING CRITERION NEEDED (EVEN W/O WP)
- IS IT POSSIBLE?
- PDFS SHOULD STOP CHANGING WITH TL?:

OVERLEARNING

- IT LOOKS LIKE AT 10K A PERFECT FIT IS REACHED
- NO OVERLEARNING → NO STOPPING CRITERION NEEDED (EVEN W/O WP)
- IS IT POSSIBLE?
- PDFs SHOULD STOP CHANGING WITH TL?: THEY DO!

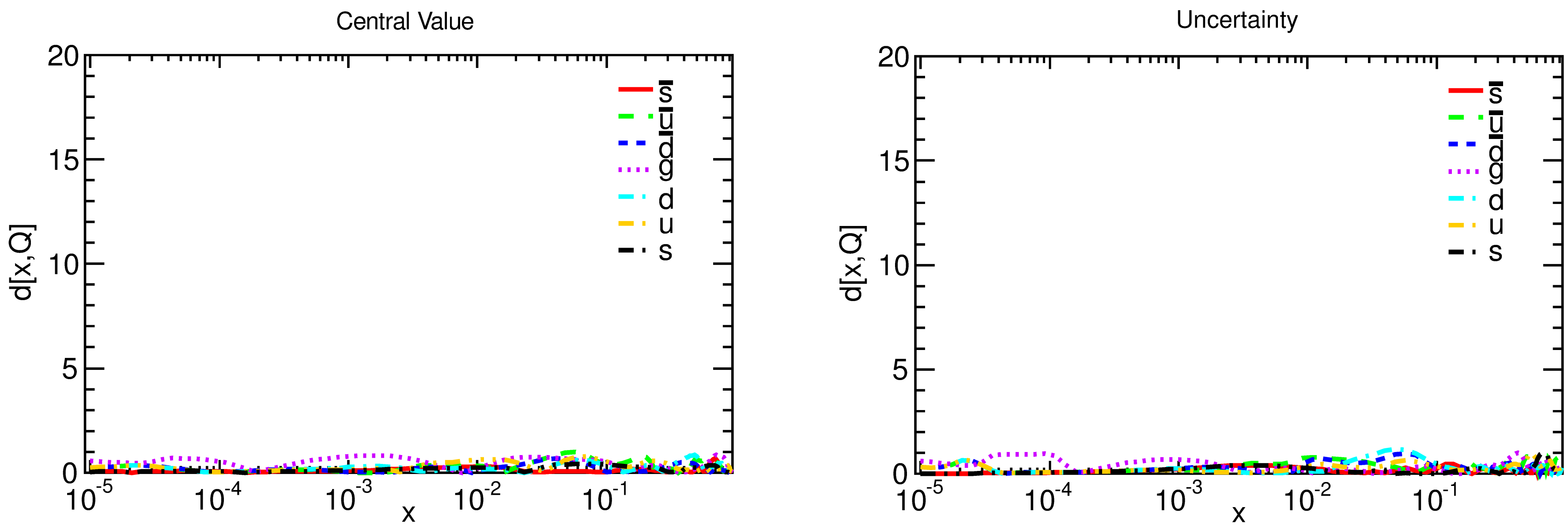
DISTANCES BETWEEN PDFs @ 10K AND 20K



OVERLEARNING

- IT LOOKS LIKE AT 10K A PERFECT FIT IS REACHED
- NO OVERLEARNING $\rightarrow$ NO STOPPING CRITERION NEEDED (EVEN W/O WP)
- IS IT POSSIBLE?
- PDFS SHOULD STOP CHANGING WITH TL?: THEY DO!
- UNCERTAINTIES MUST BE DRIVEN BY DATA FLUCTUATIONS
- TEST: REPEAT FIT WITH RESCALED UNCERTAINTIES
  $\Rightarrow$ FIT SHOULD BE UNCHANGED

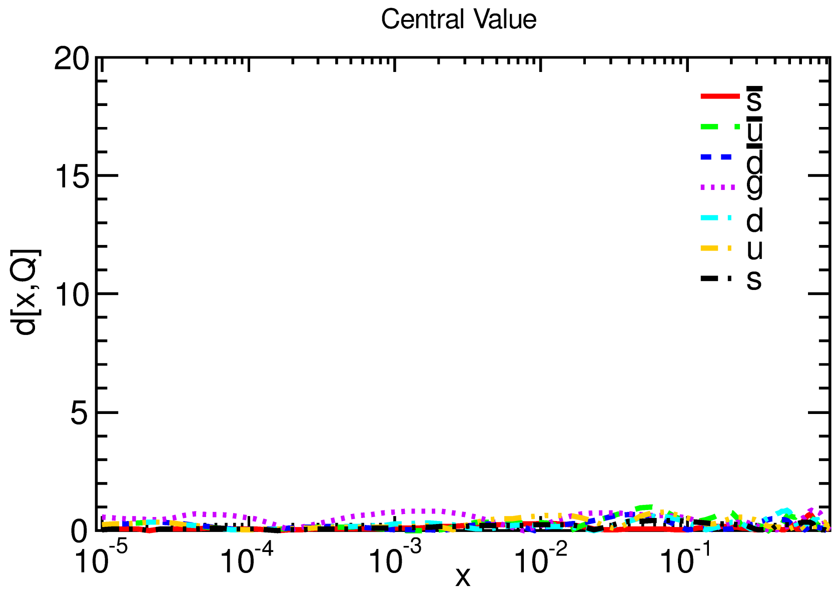
DISTANCES BETWEEN PDFS @ 10K AND 20K

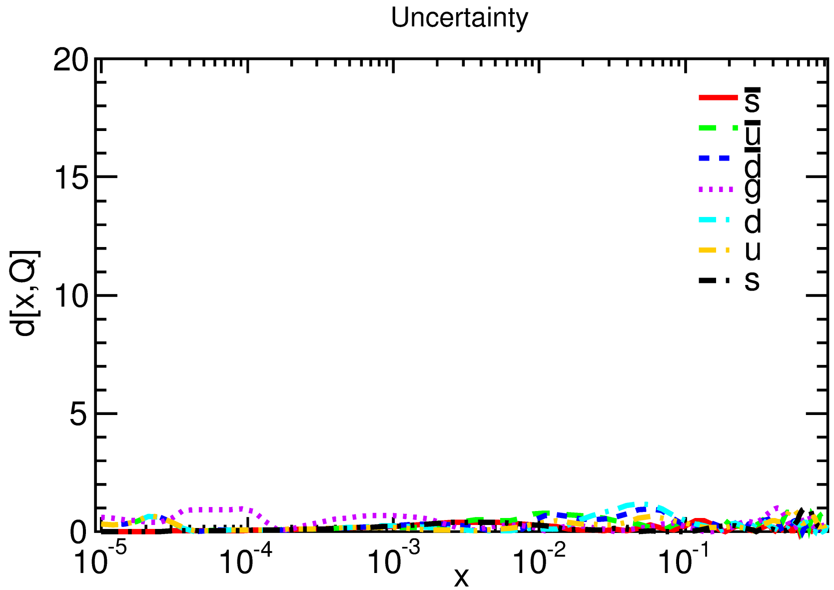


OVERLEARNING

- IT LOOKS LIKE AT 10K A PERFECT FIT IS REACHED
- NO OVERLEARNING $\rightarrow$ NO STOPPING CRITERION NEEDED (EVEN W/O WP)
- IS IT POSSIBLE?
- PDFS SHOULD STOP CHANGING WITH TL?: THEY DO!
- UNCERTAINTIES MUST BE DRIVEN BY DATA FLUCTUATIONS
- TEST: REPEAT FIT WITH RESCALED UNCERTAINTIES
  $\Rightarrow$ FIT SHOULD BE UNCHANGED
- RESULT: UNRESCALED: $\chi^2 = 0.964$;
  RESCALED BY $\times 2$: $\chi^2 = 0.245$ $(0.964/4 = .241)$;
  RESCALED BY $\times 0.65$: $\chi^2 = 2.280$ $(0.964/.65^2 = 2.281)$: IT IS!

DISTANCES BETWEEN PDFS @ 10K AND 20K





WHEN IS OVERLEARNING POSSIBLE?

- FOR ONE DATAPOINT, $\chi^2 = \frac{(t-d)^2}{\sigma^2}$, $\chi^2 = 0$ IF $t = d$
- BUT FOR TWO DATAPOINTS, $\chi^2 = \frac{(t-d_1)^2+(t-d_2)^2}{\sigma^2}$, minimum $\chi^2_{\min} = \frac{(d_1-d_2)^2}{4\sigma^2}$, IF d_i DRAWN FROM A RANDOM SAMPLE, $\chi^2_{\min} = \frac{1}{2}$
- FOR N DATAPOINTS, $\chi^2_{\min} = 1 - \frac{1}{N}$
  IF THERE ARE INFINITELY MANY MEASUREMENTS AT THE SAME POINT $\chi^2_{\min} = 1$
  $\Rightarrow$ NO OVERLEARNING

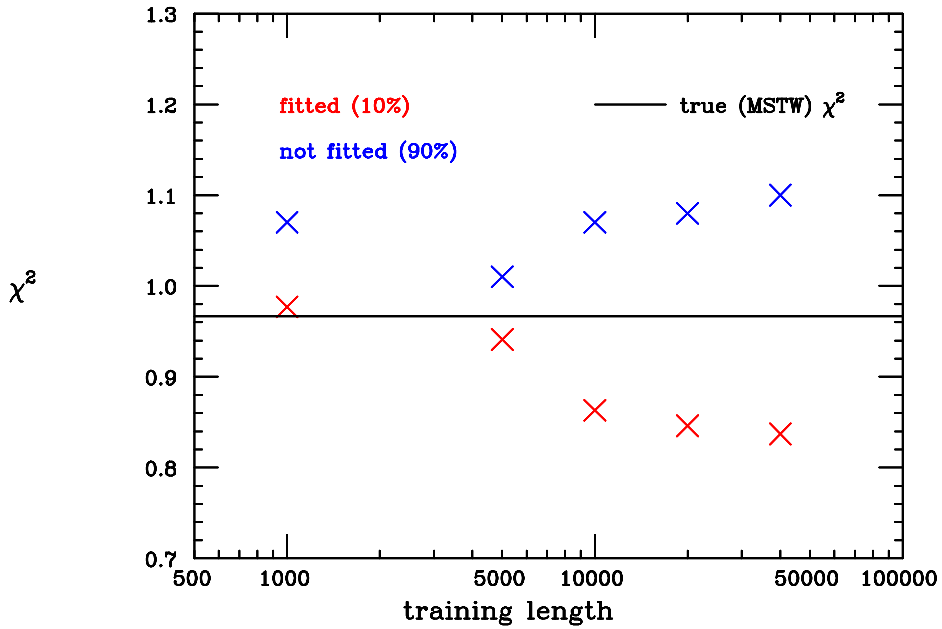
- TEST: REDUCE FRACTION OF FITTED DATA
- COMPUTE χ^2 OF FITTED & NON-FITTED DATA
- STUDY AS FUNCTION OF TRAINING LENGTH

WHEN IS OVERLEARNING POSSIBLE?

- FOR ONE DATAPOINT, $\chi^2 = \frac{(t-d)^2}{\sigma^2}$, $\chi^2 = 0$ IF $t = d$
- BUT FOR TWO DATAPOINTS, $\chi^2 = \frac{(t-d_1)^2+(t-d_2)^2}{\sigma^2}$, minimum $\chi^2_{\rm min} = \frac{(d_1-d_2)^2}{4\sigma^2}$,
  IF d_i DRAWN FROM A RANDOM SAMPLE, $\chi^2_{\rm min} = \frac{1}{2}$
- FOR N DATAPOINTS, $\chi^2_{\rm min} = 1 - \frac{1}{N}$
  IF THERE ARE INFINITELY MANY MEASUREMENTS AT THE SAME POINT $\chi^2_{\rm min} = 1$
  $\Rightarrow$ NO OVERLEARNING

- TEST: REDUCE FRACTION OF FITTED DATA
- COMPUTE χ^2 OF FITTED & NON-FITTED DATA
- STUDY AS FUNCTION OF TRAINING LENGTH
- OVERLEARNING SETS IN AROUND 5000 GA ITERATIONS
- 'DATA REDUNDANCY' OF ORDER ~ 10

FIT TO 10% OF DATA



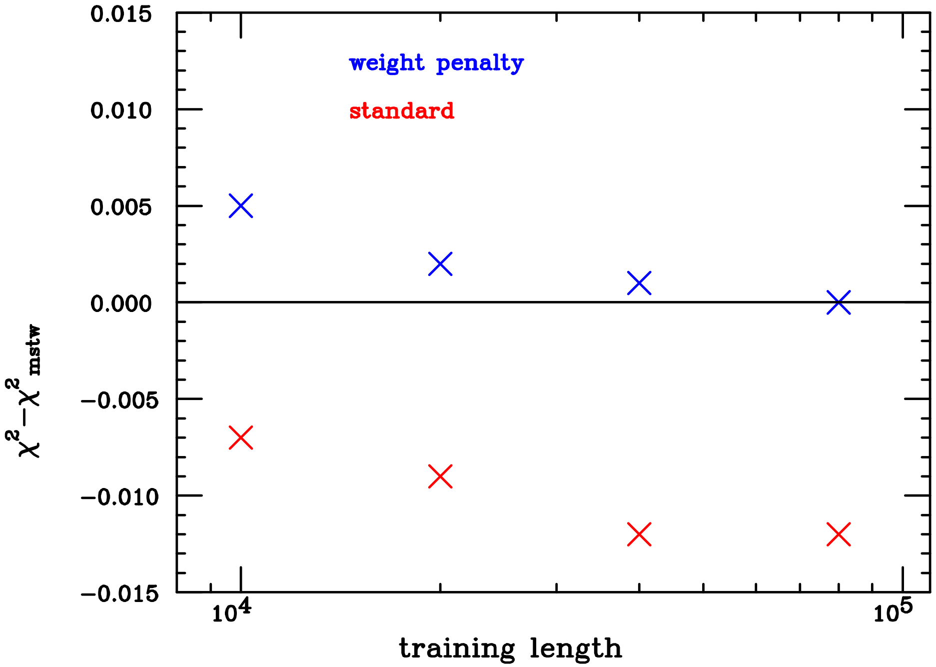
MICRO-OVERLEARNING

- EVEN IF NO OVERLEARNING VISIBLE
(PDFS DO NOT CHANGE IN STATISTICALLY SIGNIFICANT WAY)
CAN EXPLOIT KNOWLEDGE OF "TRUE" UNDERLYING THEORY
- $\Rightarrow$ COMPARE χ^2 OF FIT TO χ^2 OF "TRUTH": $\chi^2 = \chi^2_{\rm mstw}$
(REMEMBER PSEUDODATA FLUCTUATE ABOUT TRUTH, SO $\chi^2_{\rm mstw} \sim 1$)
- WHEN $\chi^2 < \chi^2_{\rm mstw}$ OVERLEARNING SETS IN

MICRO-OVERLEARNING

- EVEN IF NO OVERLEARNING VISIBLE (PDFs DO NOT CHANGE IN STATISTICALLY SIGNIFICANT WAY) CAN EXPLOIT KNOWLEDGE OF "TRUE" UNDERLYING THEORY
- $\Rightarrow$ COMPARE χ^2 OF FIT TO χ^2 OF "TRUTH": $\chi^2 = \chi^2_{\mathrm{mstw}}$ (REMEMBER PSEUDODATA FLUCTUATE ABOUT TRUTH, SO $\chi^2_{\mathrm{mstw}} \sim 1$)
- WHEN $\chi^2 < \chi^2_{\mathrm{mstw}}$ OVERLEARNING SETS IN

$\chi^2 - \chi^2_{\mathrm{mstw}}$ VS. TRAINING LENGTH



- DIFFERENCE COMPUTED BOTH FOR NON-WP & WP MIND SCALE ON y AXIUS!!
- NON-WP MICRO-OVERLEARNS SOMEWHERE AROUND 10K
- WP DOES NOT EVEN MICRO-OVERLEARN UP TO 80K
- MICRO-OVERLEARNING: MUCH SMALLER THAN STAT. FLUCTUATIONS ($\Delta\chi^2 \ll \sigma_{\chi^2}$)

MORE TESTS

- ANALYSIS OF χ^2 PROFILES (INCLUDING FOR INDIVIDUAL EXPERIMENTS)
  $\Rightarrow$ NO STATISTICALLY SIGNIFICANT OVERLEARNING SEEN
- DETAILED COMPARISON BETWEEN WP & NON-WP
- DEPENDENCE ON TRAINING LENGTH
- DEPENDENCE ON RANDOM SEED
- DEPENDENCE ON THE UNDERLYING SET I: CTEQ VS MSTW
- DEPENDENCE ON THE UNDERLYING SET II: SINUSOIDAL OSCILLATION ADDED ON TOP OF MSTW
- FITS WITH HUGE NEURAL NETWORK ARCHITECTURE:
  2-20-15-1, I.E. 391 PARMS PER NETWORK, 2737 IN TOTAL
  $\Rightarrow$ RESULTS ARE NOT DRIVEN BY NEURAL NETWORK SIZE

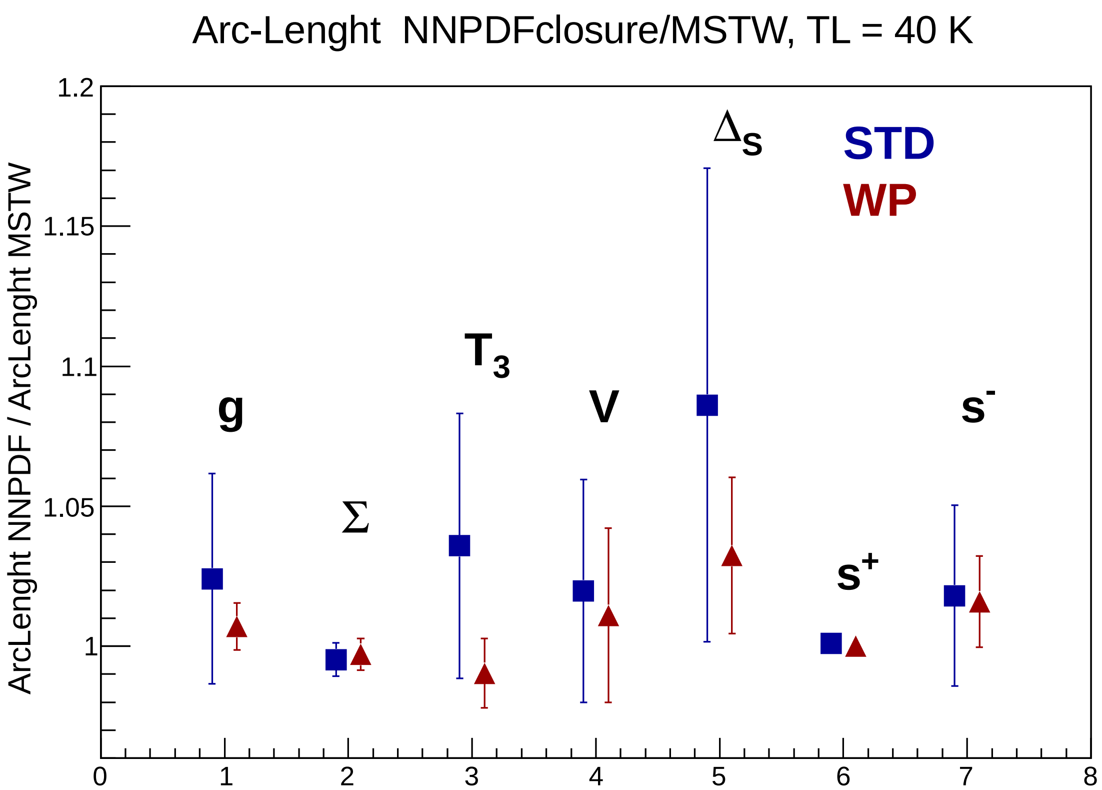
CONCLUSION

- FIXED-LENGTH FIT FULLY ADEQUATE,
- NO OVERLEARNING
- EFFECT OF WP VERY MODERATE

FLUCTUATIONS AND ARC-LENGTH

- DO WE REALLY NEED WEIGHT-PENALTY?
- LOOK AT ARC-LENGTH!: $L = \int_0^1 \sqrt{1 + \left(\frac{df}{dx}\right)^2} dx$
- SUGGESTED BY Moch, Glazov Radescu (2011) AS PENALTY FOR CHEBYSHEV POLY FITS: BUT HOW TO DETERMINE PENALTY? WP SELF-CONSISTENT!
- $\langle L \rangle$ SIMILAR FOR STANDARD AND WP, BUT σ_L RATHER SMALLER FOR WP; (ALSO, MORE STABLE W.R. TO TRAINING LENGTH)
  $\Rightarrow$ WP FIT MORE STABLE, & WITH SMALLER UNCERTAINTIES

ARC-LENGTH NORMALIZED TO "TRUTH"

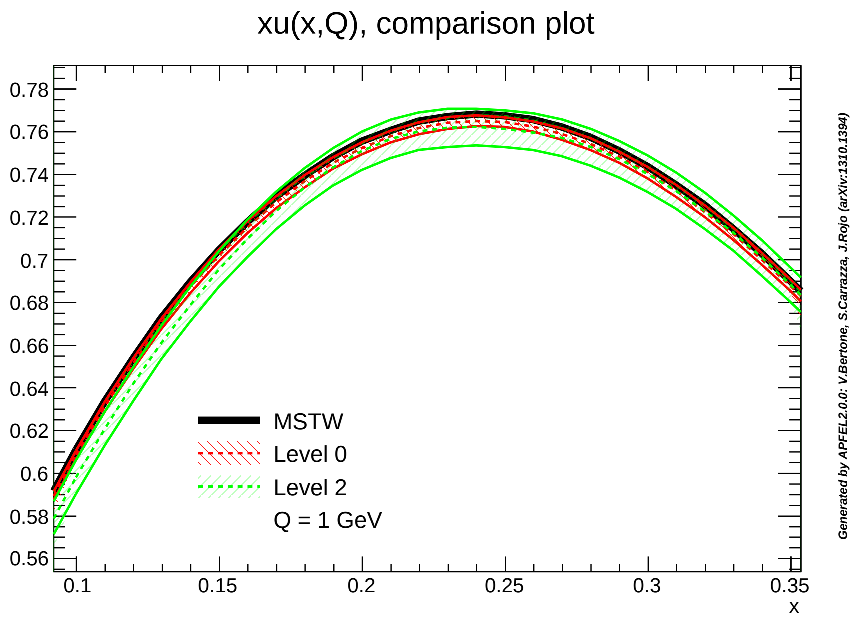


FUNCTIONAL VS. DATA UNCERTAINTY

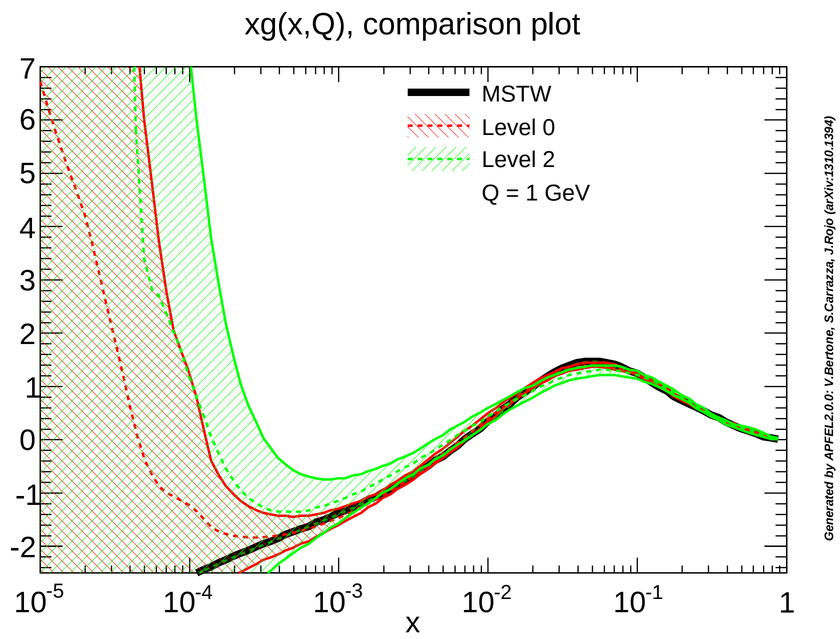
- LEVEL 2 UNCERTAINTY IS FAITHFUL REPRESENTATION OF UNDERLYING DATA UNCERTAINTY
- LEVEL 0 FIT UNCERTAINTY IS MINIMAL UNCERTAINTY WHEN DATA HAVE ZERO UNCERTAINTY → "FUNCTIONAL" UNCERTAINTY
- "TRUTH" (MSTW) IS CONTAINED WITHIN BOTH BANDS
- IN DATA REGION DATA UNC. $\ll$ FUNCTIONAL UNC.
  IN EXTRAPOLATION REGION DATA UNC. $\sim$ FUNCTIONAL UNC.

PDFS AT PARAMETRIZATION SCALE

up



gluon



OUTLOOK

- FURTHER STUDIES POSSIBLE/INTERESTING:
  - INTRODUCE "ARTIFICIAL INCONSISTENCIES" IN DATA (MISS OUT SOME SYSTEMATICS) & SEE HOW FIT BEHAVES
  - STUDY IN CONTROLLED SETTINGS IMPACT OF SPECIFIC DATASETS ON PDF KNOWLEDGE
  - DETERMINE $\Delta\chi^2$ CRITERIA $\Rightarrow$ BENCHMARKING
- CURRENT METHODOLOGY FULLY ADEQUATE FOR NNPDF3.0