PARTON DISTRIBUTIONS AND MACHINE LEARNING FROM THE LHC TO THE EIC

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

Istituto Nazionale di Fisica Nucleare

European Research Council
Established by the European Commission

N3PDF Machine Learning • PDFs • QCD

EIC UG EARLY CAREER WSHOP

WARSAW, JULY 24, 2023

SUMMARY

- LIFE BEFORE THE LHC
  - PARTON DISTRIBUTIONS AND THEIR DETERMINATION
  - THE PROBLEM OF PDF UNCERTAINTIES
- THE PROBLEM OF PDF UNCERTAINTIES
  - TOLERANCE
  - GENERALIZATION
- PDFs AS AN AI PROBLEM
  - THE NEURAL MONTECARLO
  - NEURAL NETWORK ARCHITECTURE AND TRAINING
- FROM AI TO MACHINE LEARNING
  - OPTIMIZATION AND HYPEROPTIMIZATION
  - THE MEANING OF CORRELATIONS
- VALIDATION
  - CLOSURE TESTS AND FUTURE TESTS
  - THE MEANING OF UNCERTAINTIES
- THE STATE OF THE ART
  - DATA AND UNCERTAINTIES
  - THEORETICAL ACCURACY AND INTRINSIC CHARM
- UNDERSTANDING MACHINE LEARNING
  - DISTRIBUTIONS IN FUNCTION SPACE AND IN FEATURE SPACE
  - GENERALIZATION
- THE IMPACT OF THE EIC
  - FROM QCD TO NEW PHYSICS
  - THE EIC AND MACHINE LEARNING

PDFS IN HISTORY

PDFs: THE EARLY DAYS

THE DISCOVERY OF THE W (1984)

THEORETICAL PREDICTION

TABLE 2

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

$\sqrt{S}$ (GeV)	$W^+ + W^-$ GHR	$W^+ + W^-$ DO1	$W^+ + W^-$ DO2	Z^0 GHR	Z^0 DO1	Z^0 DO2	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ GHR	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ DO1	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ 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

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: THE EARLY DAYS

THE DISCOVERY OF THE W (1984)

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

$\sqrt{S}$ (GeV)	$W^+ + W^-$ GHR	$W^+ + W^-$ DO1	$W^+ + W^-$ DO2	Z^0 GHR	Z^0 DO1	Z^0 DO2	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ GHR	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ DO1	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ 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

PDFs IN 1984

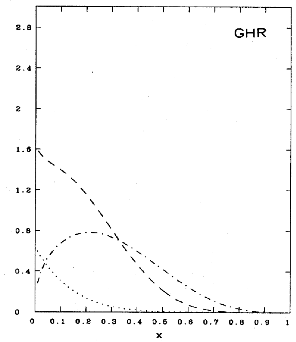


FIG. 25. Parton distributions of Glück, Hoffmann, and Reya (1982), at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and q_v (dotted line).

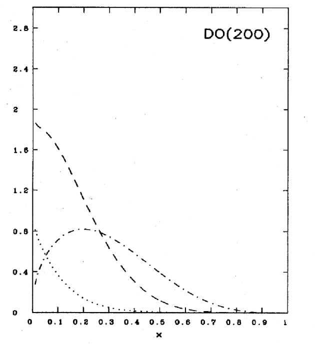


FIG. 27. "Soft-gluon" ($\Lambda=200$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

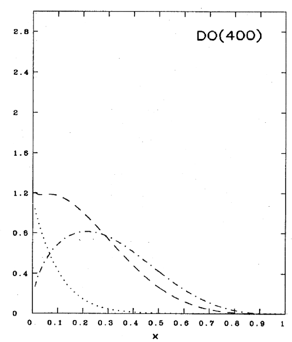


FIG. 26. "Hard-gluon" ($\Lambda=400$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

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

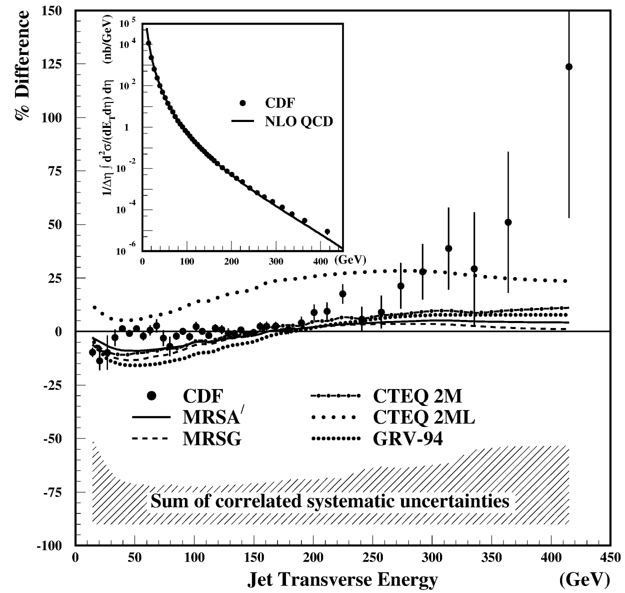
GHR VS DUKE-OWENS

- 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 and DISCOVERY

THE DISCOVERY OF QUARK COMPOSITENESS? (1995)

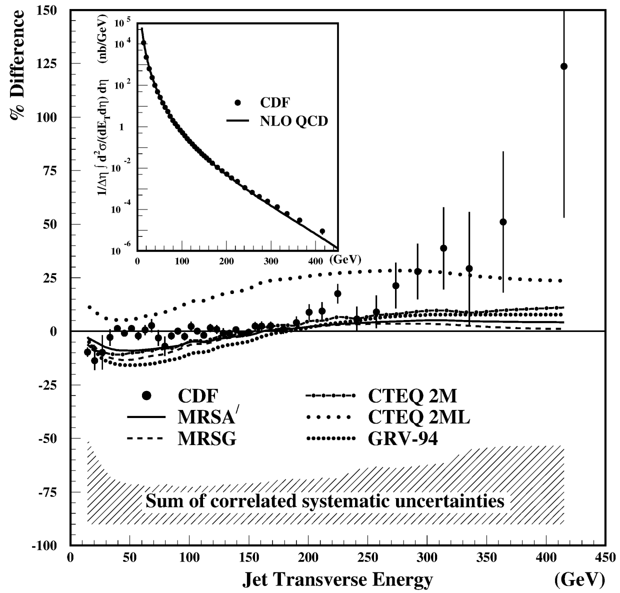
- DISCREPANCY BETWEEN QCD CALCULATION AND CDF JET DATA (1995)
- EVIDENCE FOR QUARK COMPOSITENESS
- .



PDFs AND DISCOVERY

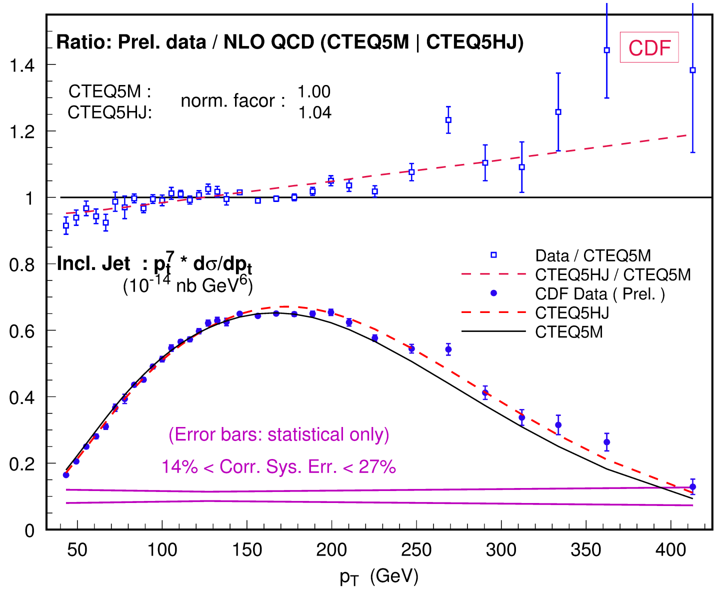
THE DISCOVERY OF THE GLUON (1995)

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

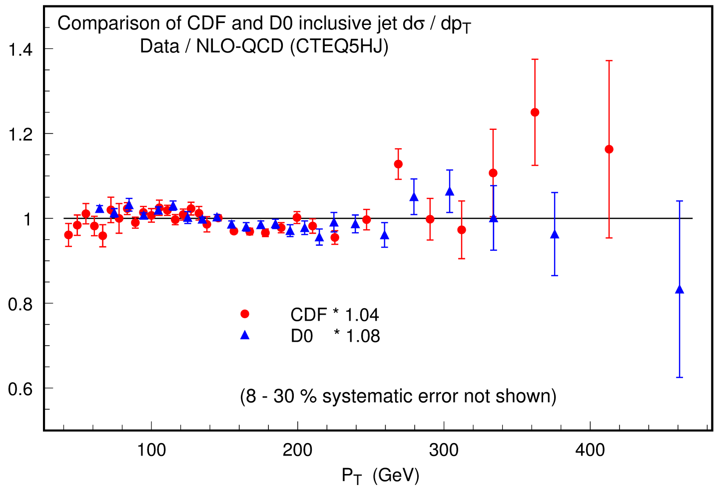


DISCREPANCY REMOVED IF JET DATA INCLUDED IN THE FIT

NEW CTEQ FIT (1996)



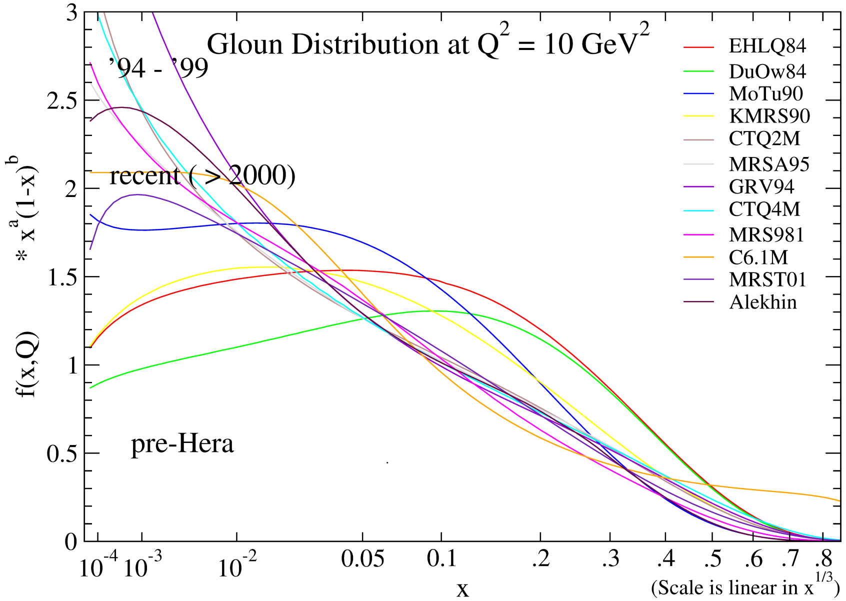
FINAL CTEQ FIT (1998)



PDFs AND DISCOVERY

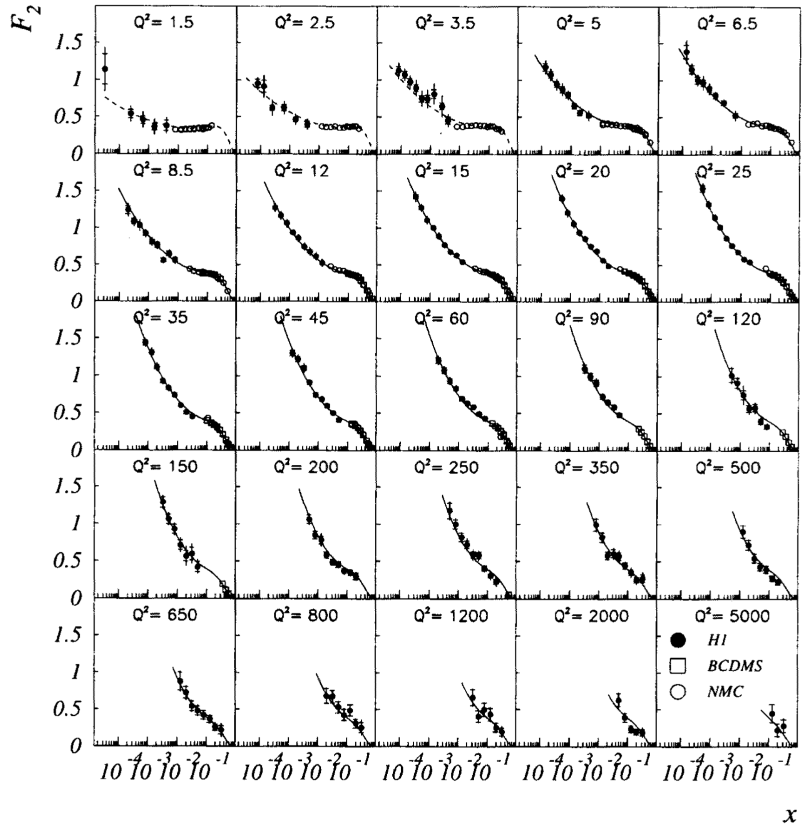
1995: THE RISE OF STRUCTURE FUNCTIONS AT HERA

HISTORICAL COMPILATION OF GLUON PDFs



W.K.Tung, DIS 2004

FIRST HERA DATA VS OLDER DATA



A. de Roeck, Cracow epiphany conf. 1996

- RISE OF F_2 AT HERA CAME $\Rightarrow$ SURPRIZE
- HINTED BY PRE-HERA DATA; VETOED BY THEORETICAL BIAS

DISCOVERING PDFs
WHAT'S THE PROBLEM?

D. Kosower, 1999

- FOR A SINGLE QUANTITY, WE QUOTE 1 SIGMA ERRORS: VALUE$\pm$ 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 A PROBABILITY DENSITY (MEASURE) IN THE SPACE OF FUNCTIONS

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

A SOLUTION? ∼ 2000

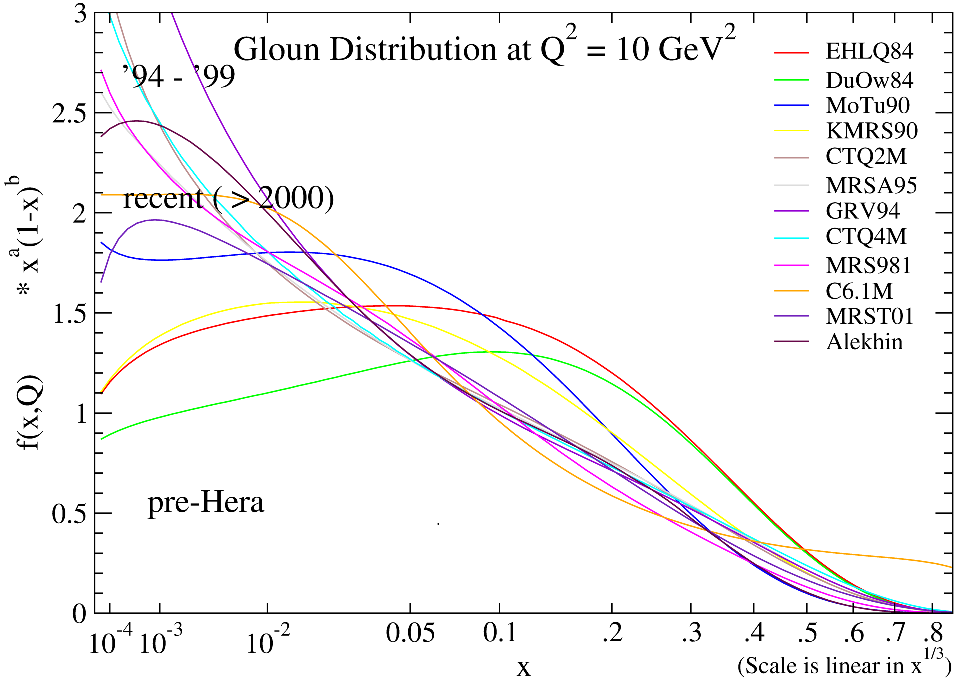
- FIT A MODEL-INSPIRED FUNCTIONAL FORM
- DETERMINE UNCERTAINTY BY STANDARD ERROR PROPAGATION
  $\Rightarrow \Delta\chi^2 = 1$ CONTOUR IN PARAMETER SPACE

gluon parametrization (MRST 2004)

$$xg(x, Q_0^2) = A_g(1-x)^{\eta_g}(1 + \epsilon_g x^{0.5} + \gamma_g x)x^{\delta_g} - A_-(1-x)^{\eta_-}x^{-\delta_-}$$

- PROBLEM REDUCED TO FINITE-DIMENSIONAL
- WHO PICKS THE FUNCTIONAL FORM?

HISTORICAL COMPILATION OF GLUON PDFs

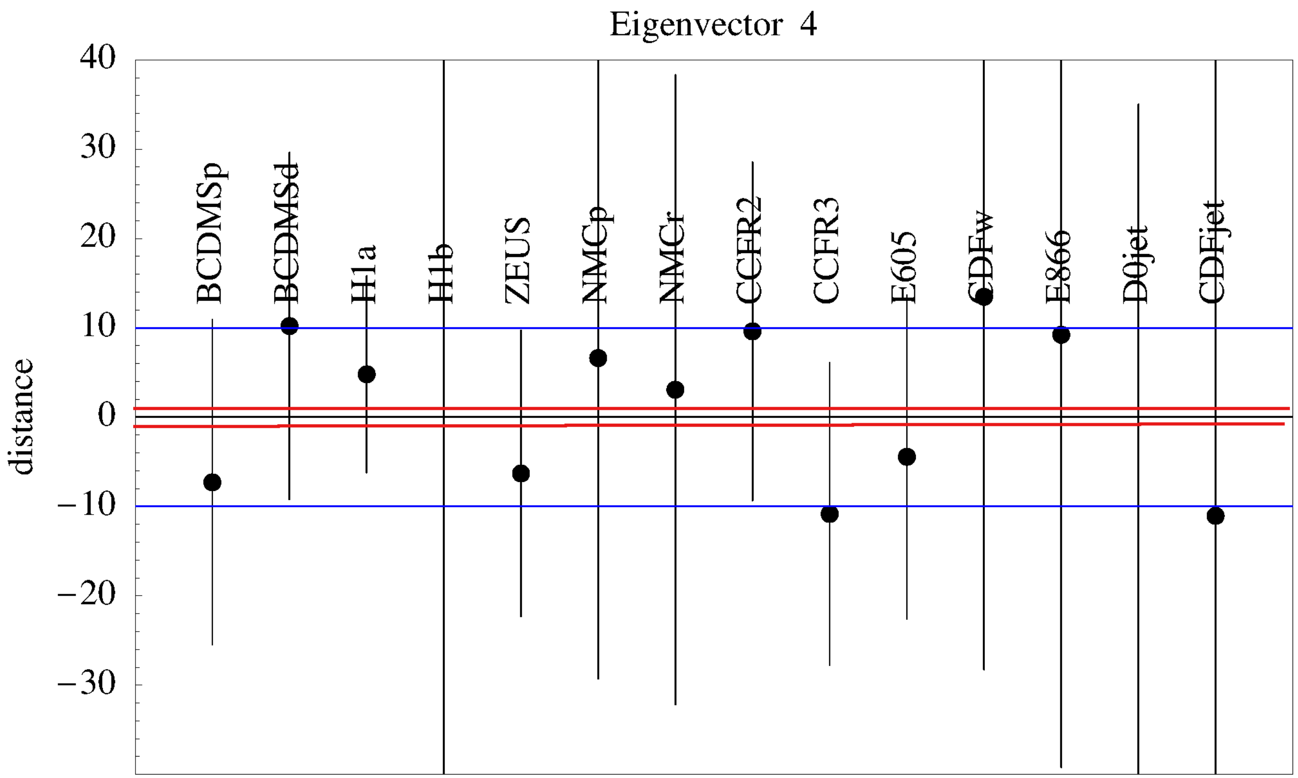


PDF UNCERTAINTIES

FIRST PDFs WITH UNCERTAINTIES (2002)
"TOLERANCE"

one sigma & ten sigma intervals for typical
covariance matrix eigenvalue

vs best value and uncertainty from individual experiments



- SPREAD OF BEST-FIT FROM DIFFERENT DATA HUGE W.R. TO $\Delta\chi^2 = 1$ ERROR PROPAGATION FORMULA
- PDF UNCERTAINTIES RESCALED BY "TOLERANCE" $T \sim 10$

THE HERA-LHC WORKSHOPS

HERALHC: 2004-2008

PDF4LHC: 2008–

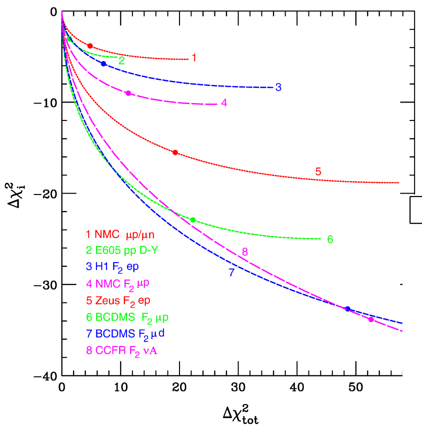
RELIABLE PDF UNCERTAINTY ESTIMATES?

DYNAMICAL TOLERANCE

"estimates of PDF uncertainties follow and ad-hoc recipe defined by the fitters" (C. Hays, 09)

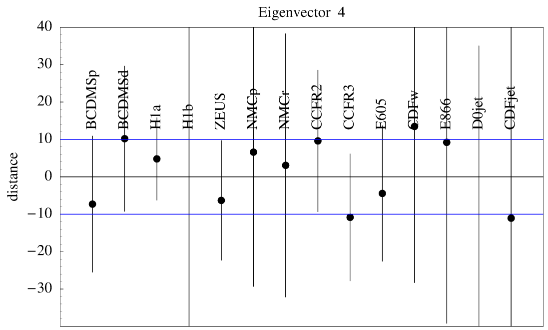
- TOLERANCE ⇒ ENVELOPE OF UNCERTAINTIES OF EXPERIMENTS
- DYNAMICAL ⇒ SEPARATELY DETERMINED FOR EACH HESSIAN EIGENVECTOR

MINIMUM χ_i^2 VS GLOBAL χ^2

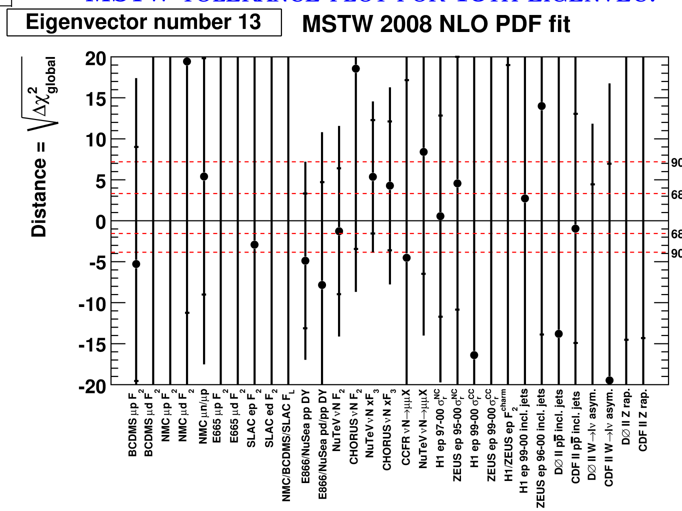


Collins, Pumplin 2001

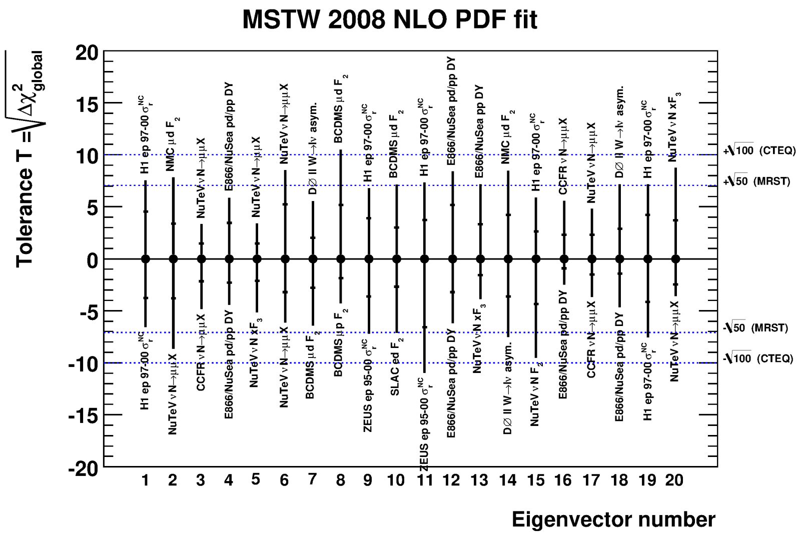
CTEQ TOLERANCE PLOT FOR 4TH EIGENVEC.



MSTW TOLERANCE PLOT FOR 13TH EIGENVEC.



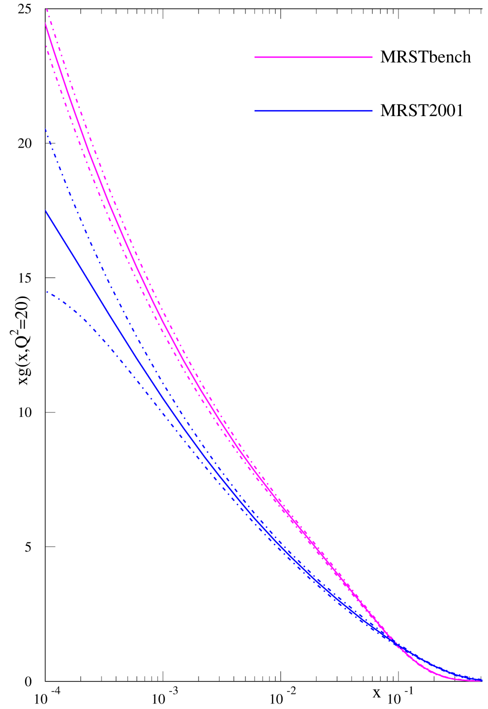
GLOBAL MSTW TOLERANCE



THE HERA-LHC BENCHMARK PROBLEM (2005)

- RESTRICTED AND VERY CONSISTENT DATASET USED
- RESULTS COMPARED TO THEN-BEST RESULT FROM FULL DATASET

BENCHMARK VS DEFAULT GLUON



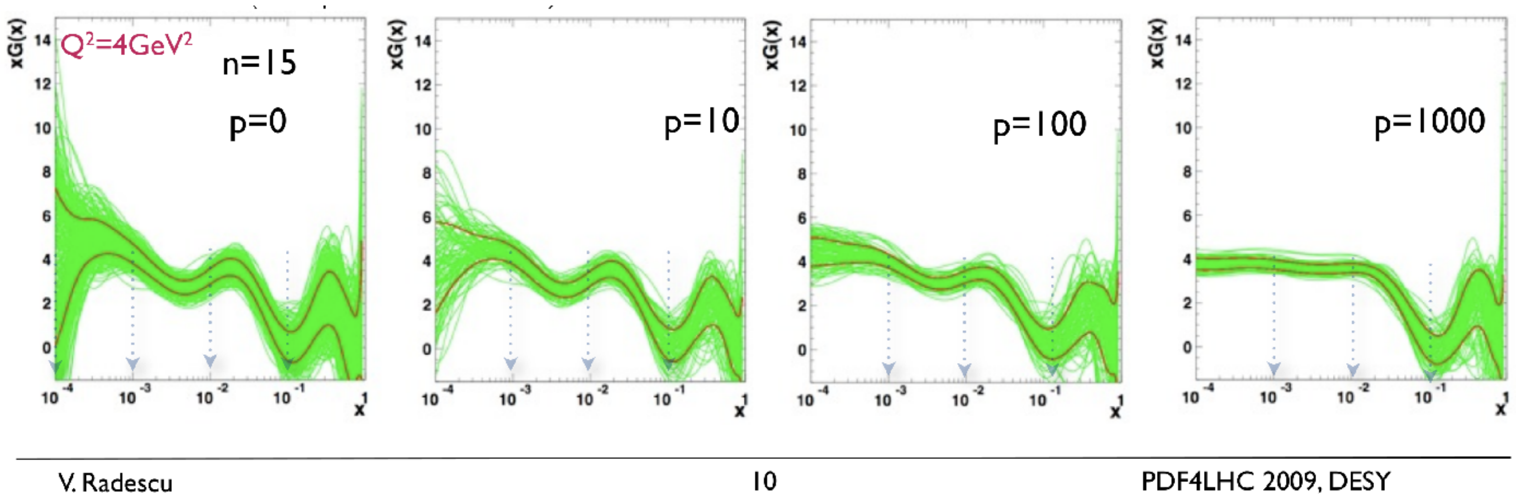
"...the partons extracted using a very limited data set are completely incompatible, even allowing for the uncertainties, with those obtained from a global fit with an identical treatment of errors...The comparison illustrates the problems in determining the true uncertainty on parton distributions." (R.Thorne, HERALHC, 2005)

ATTEMPTS AT A SOLUTION

(Glazov, Radescu, 2009)

CHEBYSHEV AND LENGTH PENALTY

- OLD IDEA (PARISI, SOURLAS, 1978):
  EXPAND PDFS OVER BASIS OF ORTHOGONAL POLYNOMIALS
- LENGTH PENALTY STABILIZATION:
  CONTRIBUTION TO χ^2 PROPORTIONAL TO THE ARCLENGTH WITH WEIGHT p
- RESULTS STRONGLY DEPENDENT ON ARBITRARY CHOICE OF p

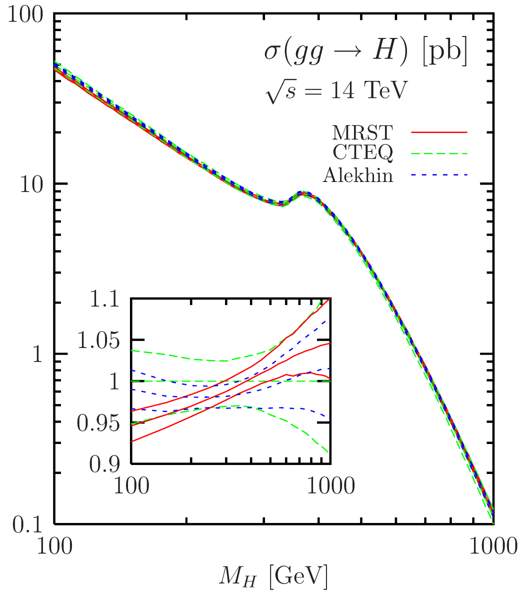


V. Radescu 10 PDF4LHC 2009, DESY

PDFs at the dawn of the LHC

can one calculate
the Higgs cross section??

before HERALHC
old disagreements



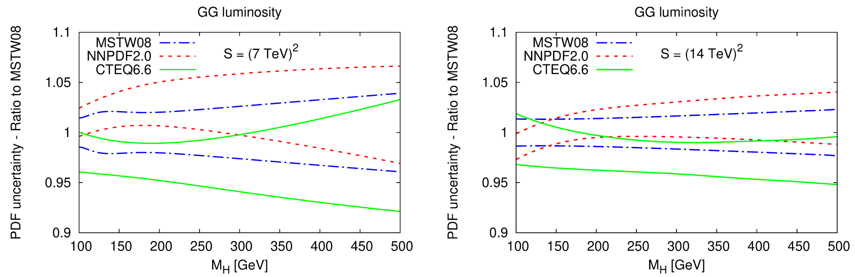
(Djouadi, Ferrag, 2004)

CTEQ, MRST (global); Alekhin (DIS)

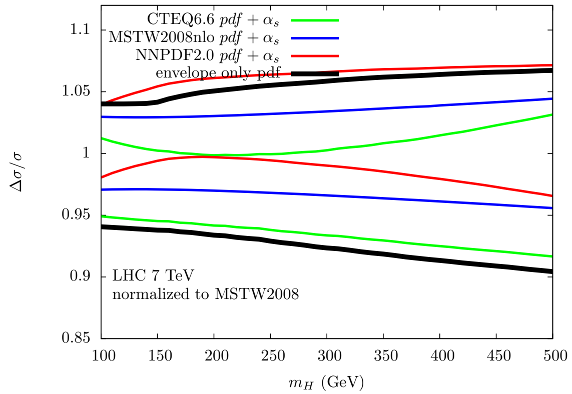
- widely different uncertainties
- poor agreement within uncertainties
- tolerance? $\Delta\chi^2 = 100$ (CTEQ); 50 (MRST);
  1 (Alekhin)

at LHC start
progress

gluon luminosities



gg→H cross section



(Demartin et al., 2010)

- three global (DIS+hadronic) PDF sets available
- reasonable agreement of central values & uncertainties

THE MODEL-DEPENDENT APPROACH: PROGRESS

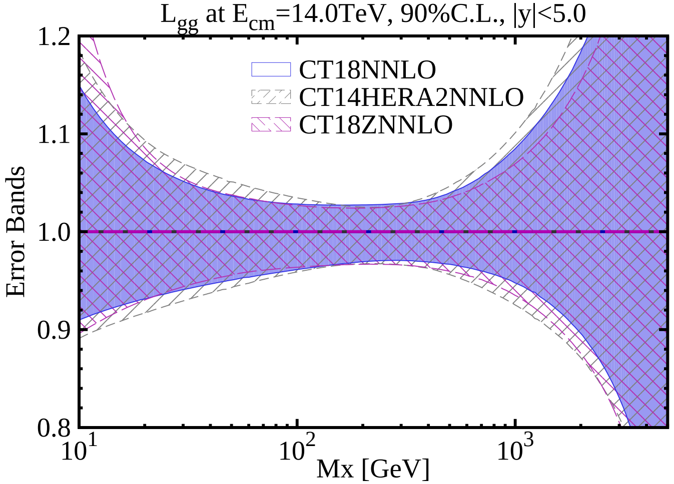
- INCREASINGLY COMPLEX PARAMETRIZATION
- UNDERLYING PHYSICALLY MOTIVATED ANSATZ (SINCE 1973!) $f_i(x, Q_0^2) = x^\alpha (1-x)^\beta g_i(x)$; $g_i(x)$ POLYNOMIAL IN x OR $\sqrt{x}$
- Example: MMHT 2015:
  - basis functions g; $u_v = u - \bar{u}$; $d_v = d - \bar{d}$; $S = 2(\bar{u} + \bar{d}) + s + \bar{s}$; $s_+ = s + \bar{s}$; $\Delta = \bar{d} - \bar{u}$; $s_- = s - \bar{s}$.
  - for all but Δ s_-, $g \Rightarrow x f_i(x, Q_0^2) = A x^\alpha (1-x)^\beta \left(1 + \sum_{i=1}^4 a_i T_i(y(x))\right)$; T_i Chebyshev polynomials, $y = 1 - 2\sqrt{x} \leftrightarrow$ must map $x = [0, 1]$ into $y = [-1, 1]$; $T_i(-1) = T_i(1) = 1$
  - gluon $xg(x, Q_0^2) = A x^\alpha (1-x)^\beta \left(1 + \sum_{i=1}^2 a_i T_i(y(x))\right) + A' x T \alpha' (1-x)^{\beta'}$
  - sea asymmetry $x\Delta(x, Q_0^2) = A x^\alpha (1-x)^\beta (1 + \gamma x + \epsilon x^2)$
  - strangeness asymmetry $x\Delta(x, Q_0^2) = A x^\alpha (1-x)^\beta (1 - x/x_0)$
  - 41 parameters, 4 fixed by sum rules
  - 12 parms fixed at best fit, remaining 25 used for (hessian) covariance matrix

THE MODEL-DEPENDENT APPROACH:
PROBLEMS
ADDING NEW DATA
PARTON PARAMETRIZATIONS

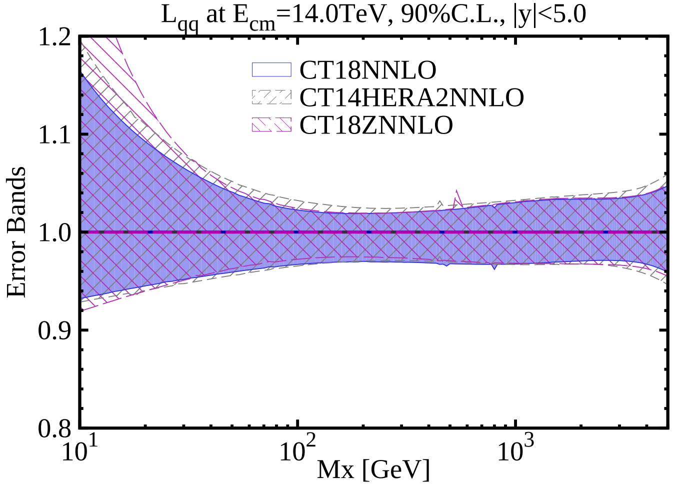
- CTEQ5 2002: $xg(x,Q_0^2) = A_0 x^{A_1}(1-x)^{A_2}(1+A_3 x^{A_4})$
- MRST-HERALHC 2005: $xg(x,Q_0^2) = A_g x^{\delta_g}(1-x)^{\eta_g}(1+\epsilon_g x^{0.5}+\gamma_g x) + A_{g'} x^{\delta_{g'}}(1-x)^{\eta_{g'}}$
- CT18: $g(x,Q=Q_0) = x^{a_1-1}(1-x)^{a2}\left[a_3(1-y)^3 + a_4 3y(1-y)^2 + a_5 3y^2(1-y) + y^3\right]$; $y=\sqrt{x}$; $a_5 = (3+2a_1)/3$.

CT18 VS. CT14: PARTON LUMINOSITY UNCERTAINTIES

GLUON-GLUON



QUARK-QUARK



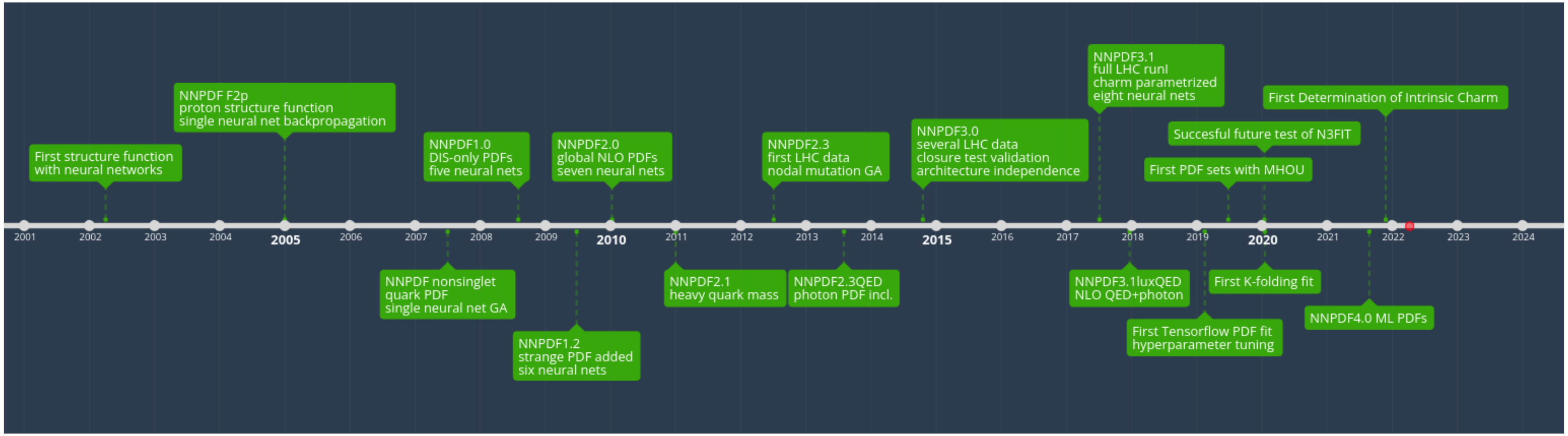
MORE DATA $\Rightarrow$ BIGGER UNCERTAINTIES (?!)

BIAS?

PDFs and AI

PROTON STRUCTURE AS AN AI PROBLEM:
NNPDF

UNBIASED MODELING WITH UNCERTAINTIES IN A SPACE OF FUNCTIONS

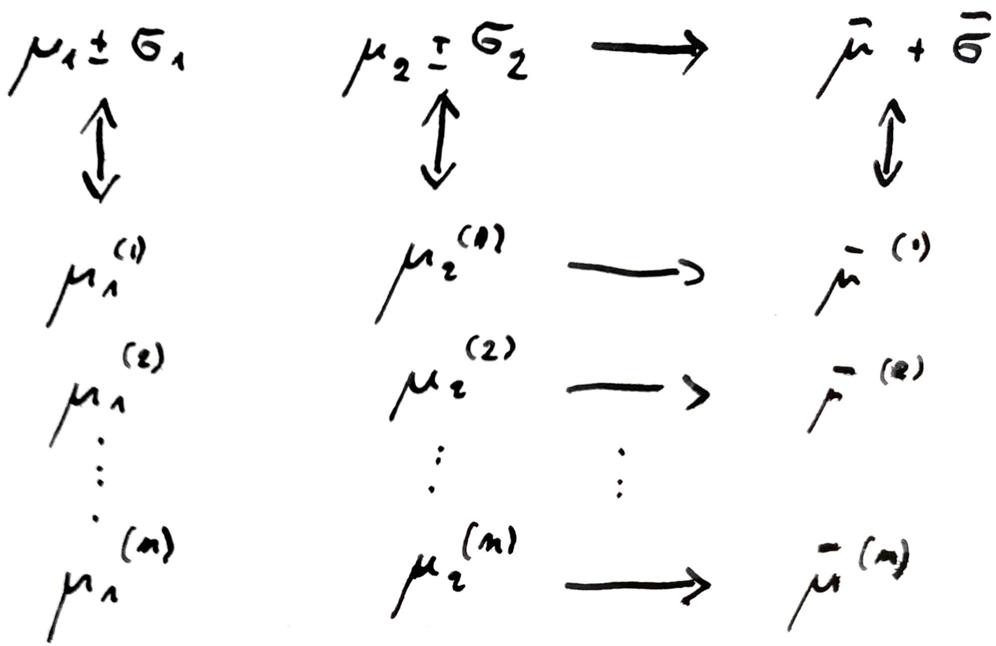


AI SOLUTION

- NEURAL NETWORK REGRESSION
- MONTE CARLO UNCERTAINTIES

MONTE CARLO COMBINATION

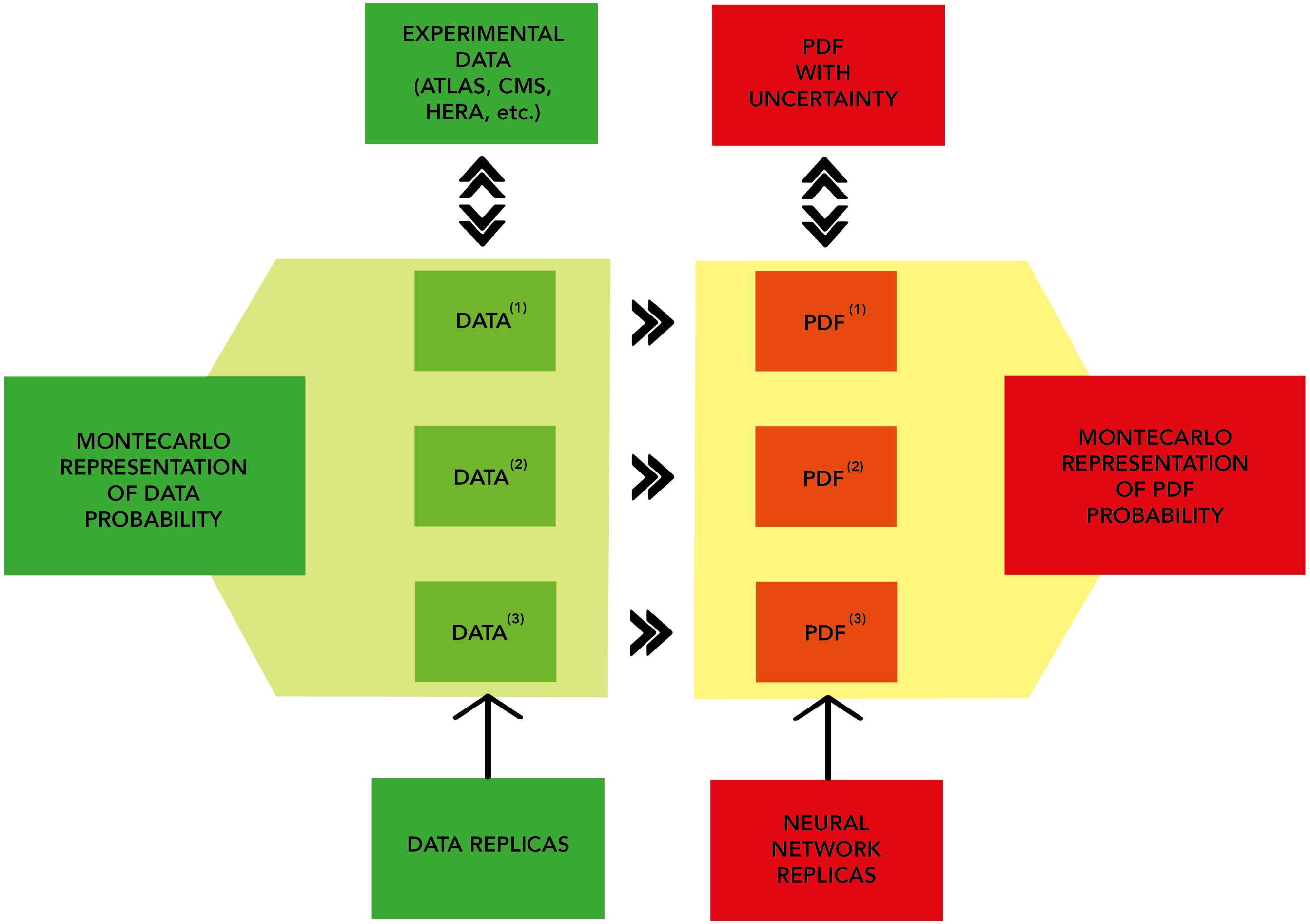
- TWO DATA WITH UNCERTAINTY $z_i = \mu_i \pm \sigma_i$
- SAMPLE OF DATA REPLICAS $\mu_i^{(k)} \to \mu_i = \langle \mu_i^{(k)} \rangle$; $\sigma_i^2 = \langle \left(\mu_i^{(k)} - \mu_i \right)^2 \rangle$.
- MAP COMBINATION $\mu_1^{(k)}, \mu^{(k)})_2 \to \bar{\mu}^{(k)}$
- $\mu^{(k)}$ REPLICA SAMPLE $\Rightarrow$ REPRESENTATION OF Max A Posteriori PROBABILITY $\bar{\mu} \pm \bar{\sigma}$
  $\bar{\mu} = \langle \bar{\mu}^{(k)} \rangle$; $\bar{\sigma}^2 = \langle \left(\bar{\mu}^{(k)} - \bar{\mu} \right)^2 \rangle$.



THE FUNCTIONAL MONTE CARLO

REPLICA SAMPLE OF FUNCTIONS $\Leftrightarrow$ PROBABILITY DENSITY IN FUNCTION SPACE

AVOIDS LIKELIHOOD IN FUNCTION SPACE

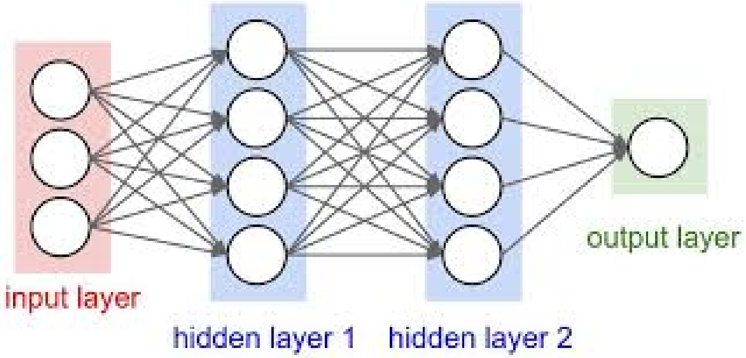


FINAL PDF SET: $f_i^{(a)}(x, \mu)$;

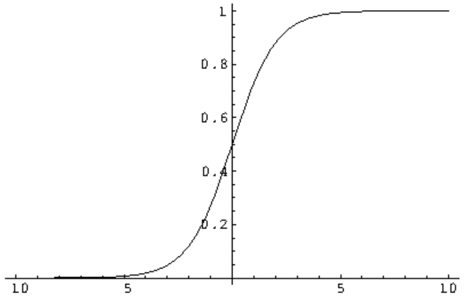
i =up, antiup, down, antidown, strange, antistrange, charm, gluon; $j = 1, 2, \ldots N_{\rm rep}$

FEED-FORWARD NEURAL NETWORKS

ARCHITECTURE



ACTIVATION FUNCTION



PARAMETERS

- WEIGHTS ω_{ij}
- THRESHOLDS θ_i

$$F^{(i)}_{\text{out}}(\vec{x}_{\text{in}}) = F\left(\sum_j \omega_{ij} x^j_{\text{in}} - \theta_i\right)$$

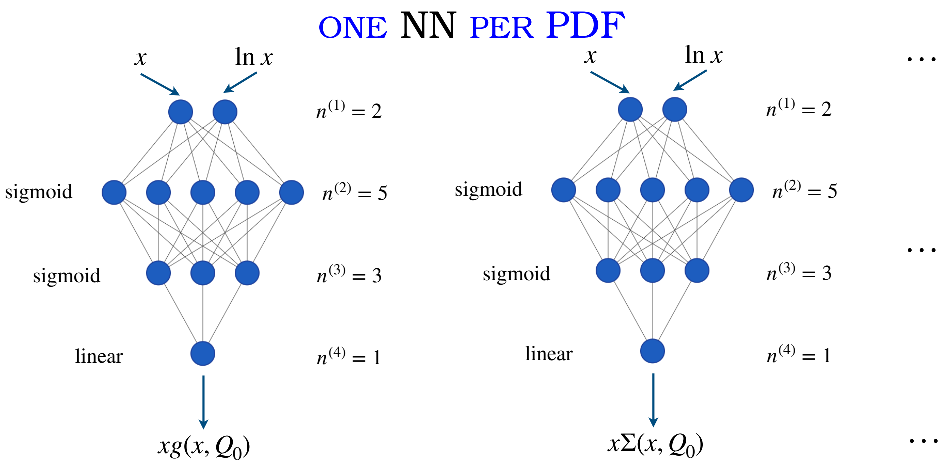
SIMPLEST EXAMPLE

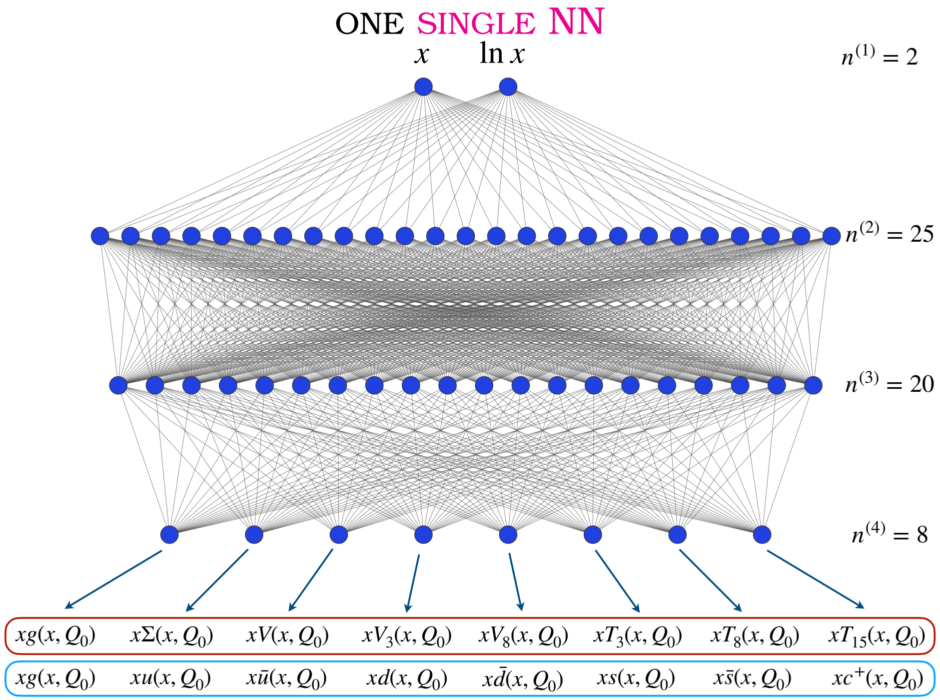
1-2-1

$$f(x) = \frac{1}{1+e^{\theta^{(3)}_1 - \frac{\omega^{(2)}_{11}}{1+e^{\theta^{(2)}_1 - x\omega^{(1)}_{11}}} - \frac{\omega^{(2)}_{12}}{1+e^{\theta^{(2)}_2 - x\omega^{(1)}_{21}}}}}$$

NEURAL NETWORKS
ARCHITECTURE

- HOW MANY INPUTS?
- HOW MANY INDEPENDENT NNs?



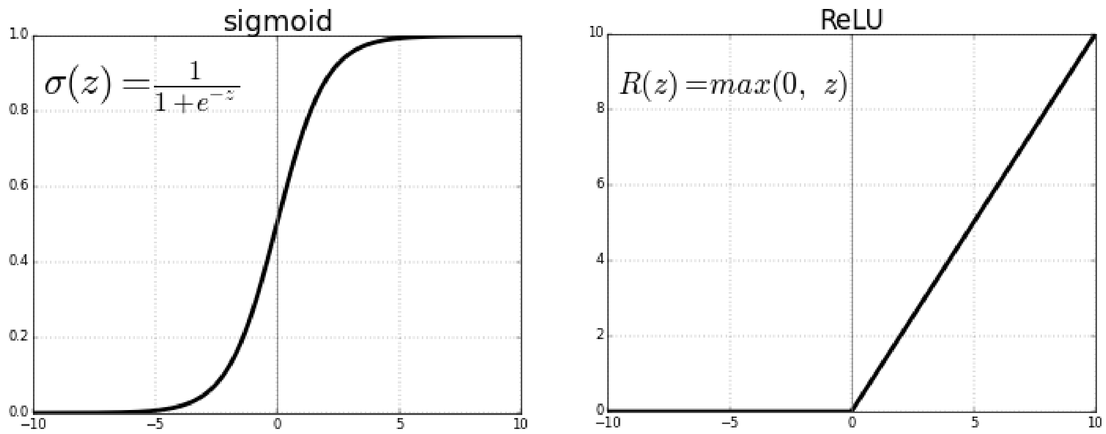


NEURAL NETWORKS

ACTIVATION FUNCTION

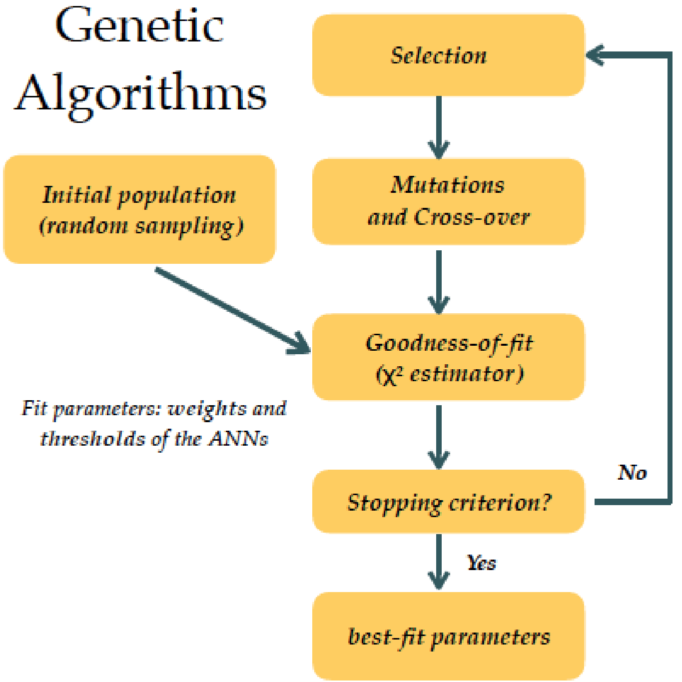
- LINEAR ACTIVATION $\Rightarrow$ MULTILINEAR REGRESSION
- + NONLINEAR PROFILE $\Rightarrow$ UNIVERSAL INTERPOL.
  - sigmoid $F(x) = \frac{1}{1+e^{-x}}$
  - arctan $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x$
  - RELU $F(x) \begin{cases} 0; & x < 0 \\ x; & x > 0 \end{cases}$

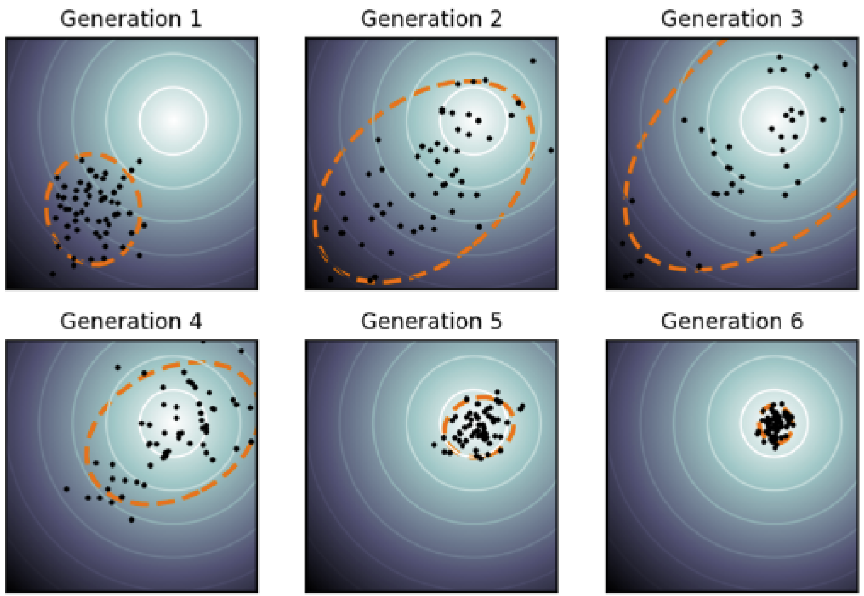
$$F_{\mathrm{out}}^{(i)}(\vec{x}_{\mathrm{in}}) = F\left(\sum_j \omega_{ij} x_{\mathrm{in}}^j - \theta_i\right)$$



NEURAL NETWORK TRAINING
GENETIC ALGORITHMS

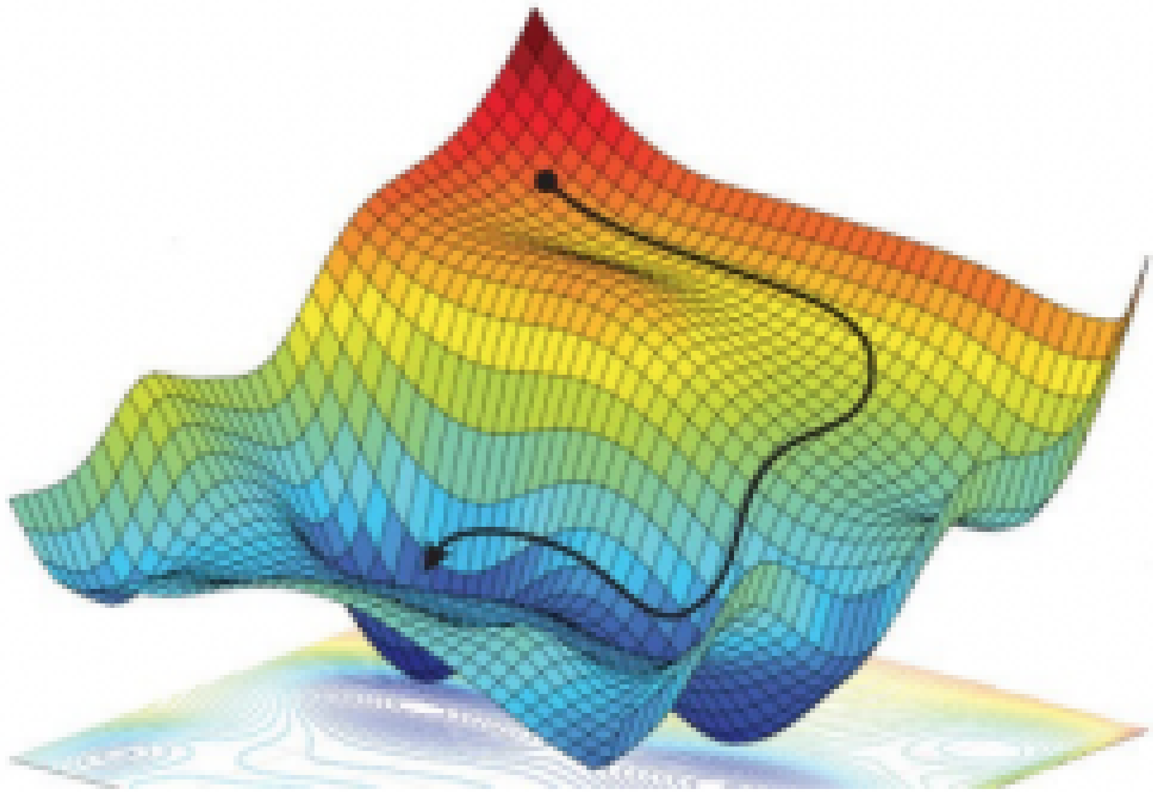
- BASIC IDEA: RANDOM MUTATION OF THE NN PARAMETER
- SELECTION OF THE FITTEST





NEURAL NETWORK TRAINING
GRADIENT DESCENT

- BASIC IDEA: COMPUTE GRADIENT OF LOSS W.R. TO PARAMETERS
- SELECT DIRECTION OF DESCENT



NEURAL NETWORK TRAINING

MINIMIZATION ALGORITHMS: DESIDERATA

- FAST CONVERGENCE
- DO NOT STOP ON LOCAL MINIMA
- EXPLORE SPACE OF MINIMA (DEGENERATE CASE)

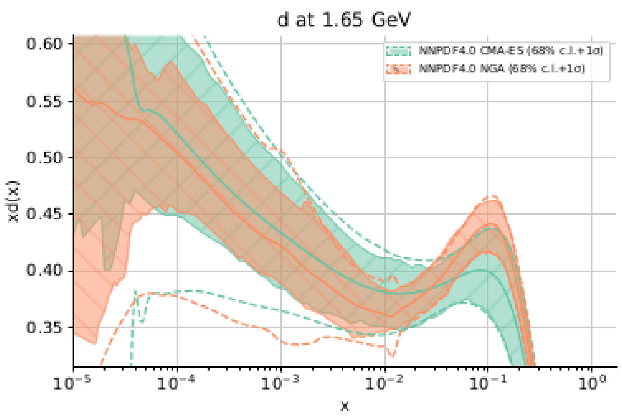
GENETIC ALGORITHMS

- DIFFERENT EPOCHS; VARIABLE MUTATION RATE
- REWEIGHTING DIFFERENT DATA CONTRIBUTIONS TO LOSS
- NODAL MUTATION
- COVARIANCE MATRIX ADAPTATION (CMA)

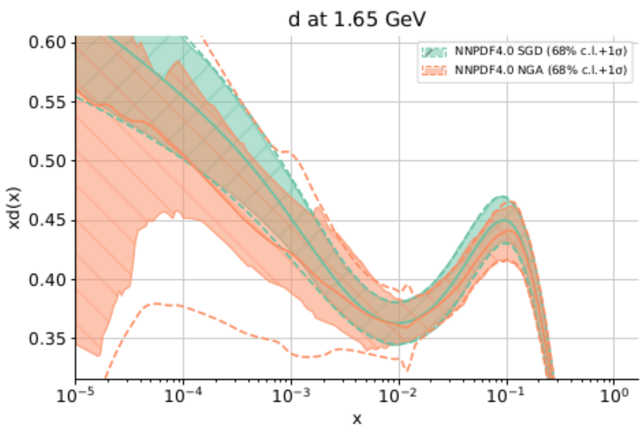
GRADIENT DESCENT

- GLOROT NORMAL/UNIFORM INITALIZATION
- ADAPTIVE GRADIENT / ADAPTIVE MOMENT
- STOCHASTIC GD
- BATCH GD

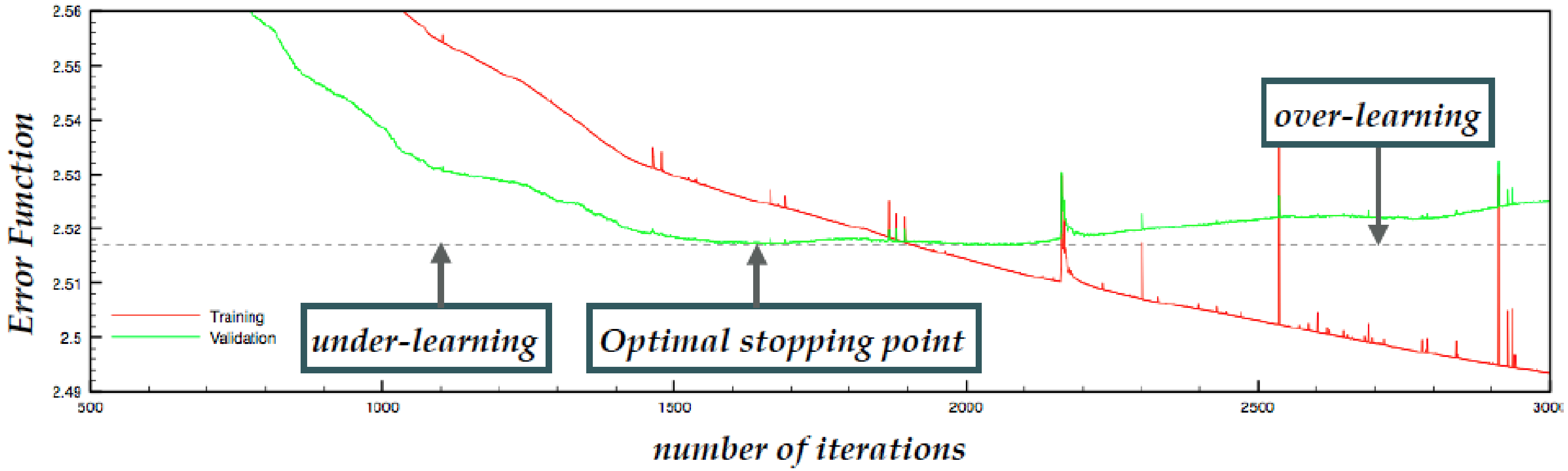
NAIVE GA VS. CMA



GA (NAIVE) VS GD (ADADELTA)



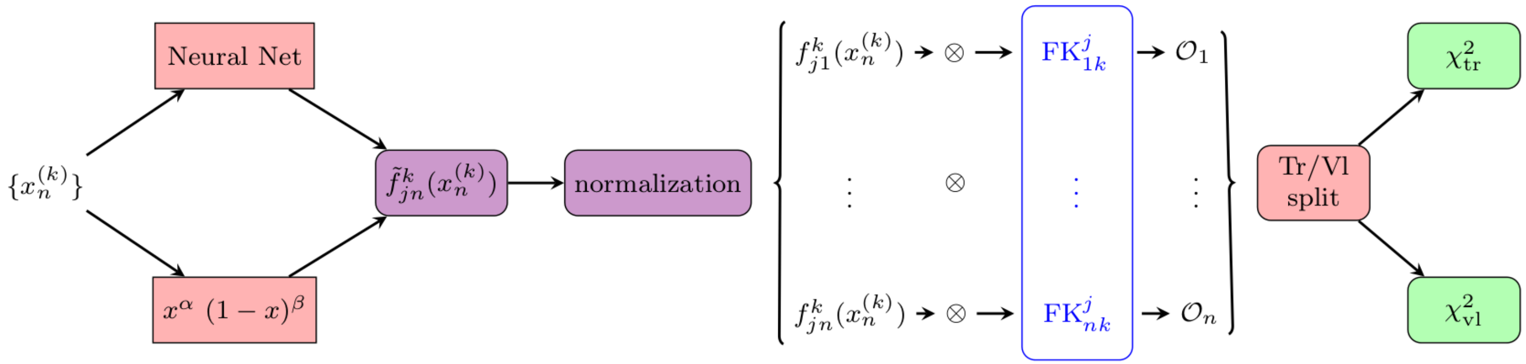
NEURAL NETWORK TRAINING
OVERLEARNING AND CROSS-VALIDATION



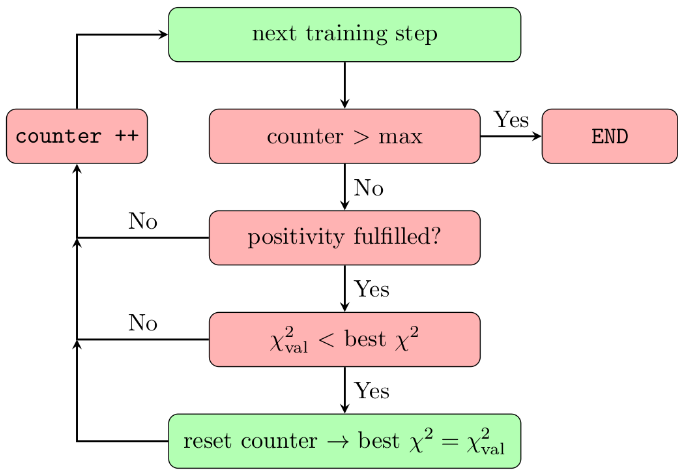
- NEURAL NET TRAINING $\Rightarrow$ LOSS MINIMIZATION (χ^2)
- RANDOM TRAINING-VALIDATION SPLIT, TRAINING LOSS MINIMIZED
- TRAINING STOPS AT MIMIMUM OF VALIDATION LOSS

PDFS AND MACHINE LEARNING

NEURAL NETS FOR PDFs
THE ALGORITHM



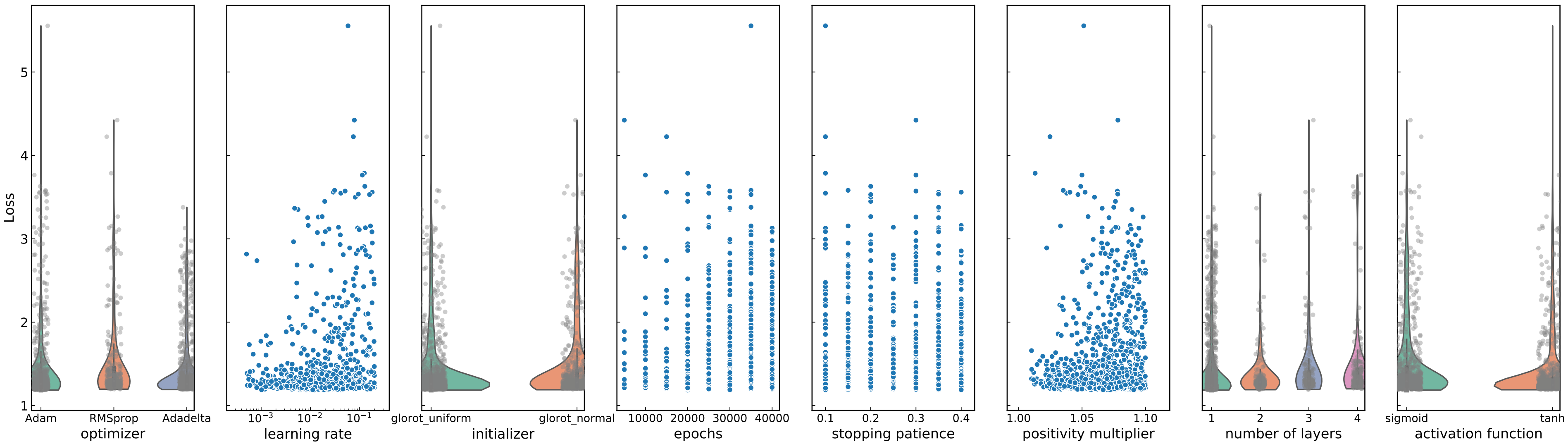
STOPPING



THE HYPERPARAMETERS

MODEL	MINIMIZATION
Number of layers	Optimizer
Size of each layer	Initializer
Activation functions	Learning rate
Initial positivity	Clipnorm
Initial integrability	Maximum number of epochs
	Stopping Patience

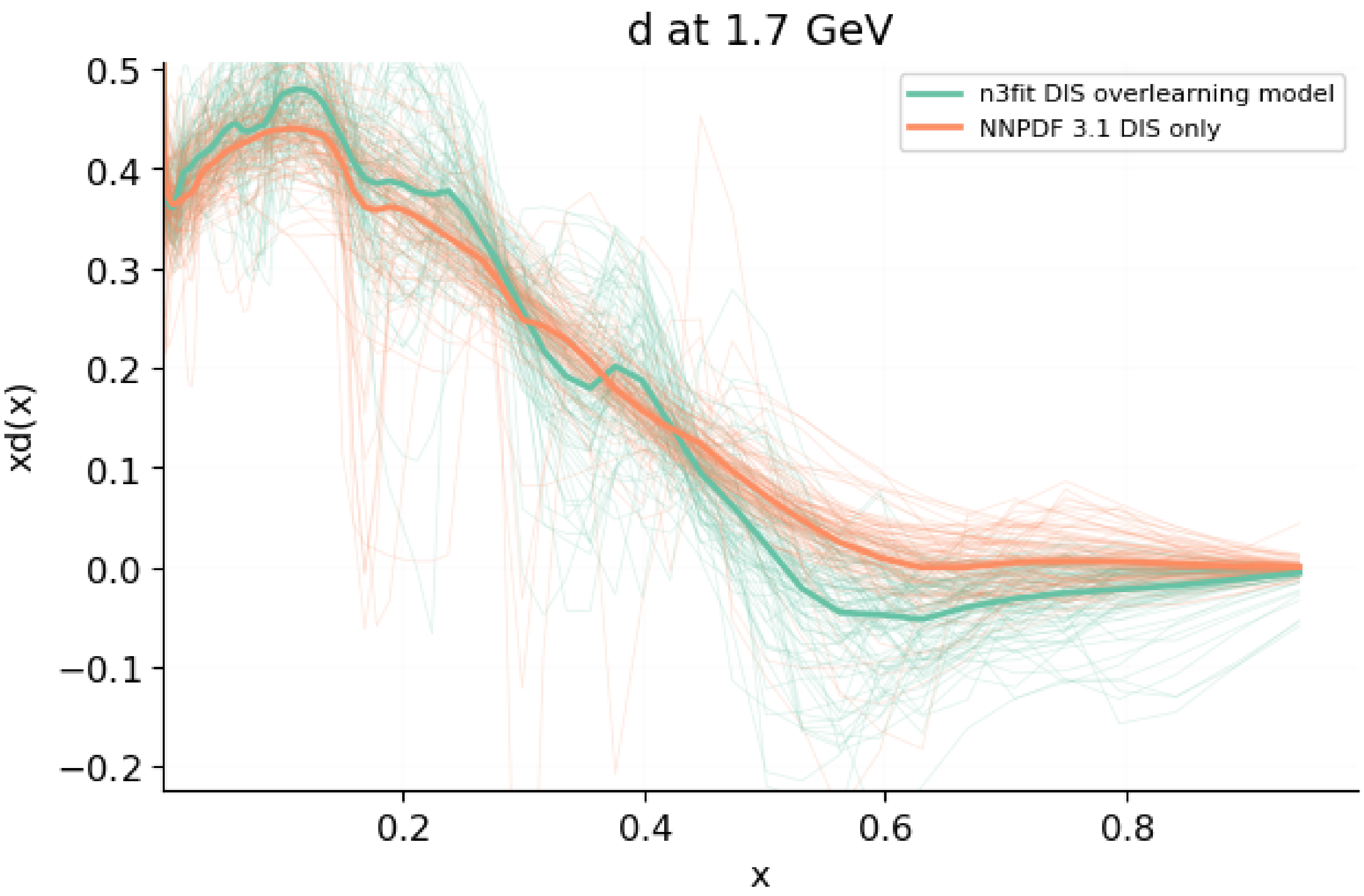
HYPERPARAMETER OPTIMIZATION



- BAYESIAN SCAN OF PARAMETER SPACE
- OPTIMIZE LOSS: VALIDATION χ^2

RESULTS: OVERFITTING!

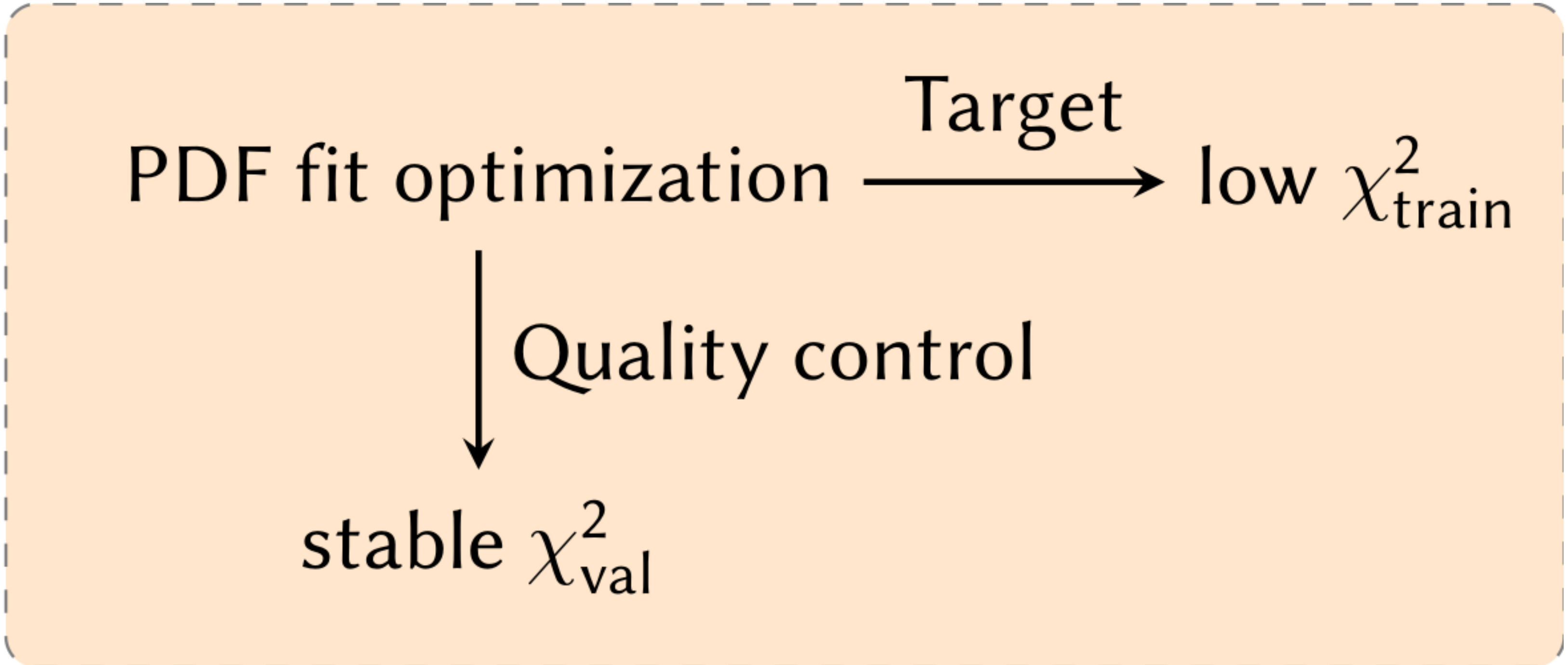
DOWN QUARK: HYPEROPTIMIZED VS. HAND-PICKED



- HAND-PICKED: WIGGLES: FINITE SIZE $\Rightarrow$ WILL GO AWAY AS $N_{\rm rep}$ GROWS
- HYPEROPT: WIGGLY PDFS $\Leftrightarrow$ OVERFITTING $\Rightarrow$ WILL **NOT** GO AWAY ($\chi^2_{\rm train} \ll \chi^2_{\rm valid}$ EVEN THOUGH VALIDATION LOSS MNIMIZED)

WHAT HAPPENED?

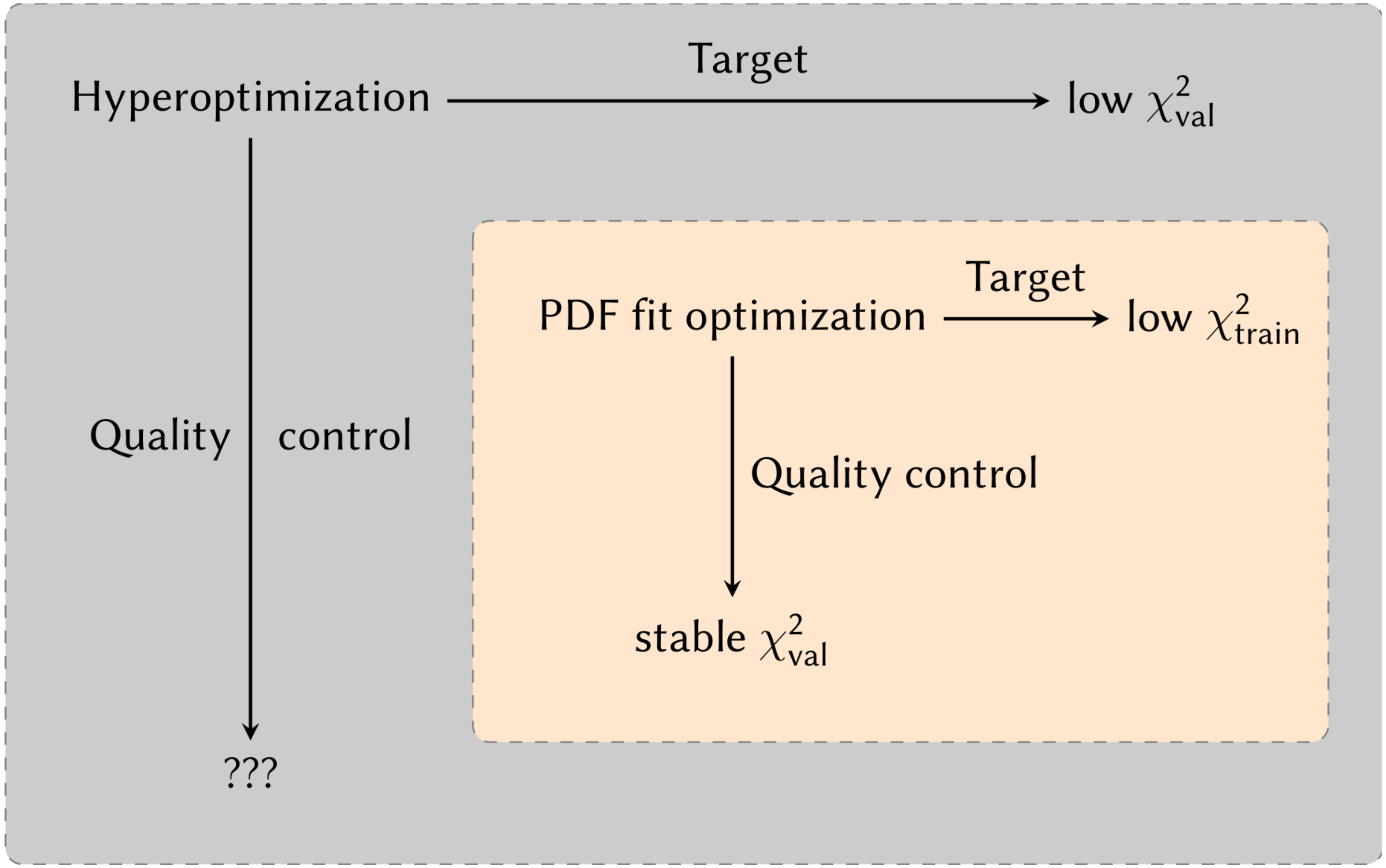
OPTIMIZATION



CROSS-VALIDATION SELECTS THE OPTIMAL MINIMUM

WHAT HAPPENED?

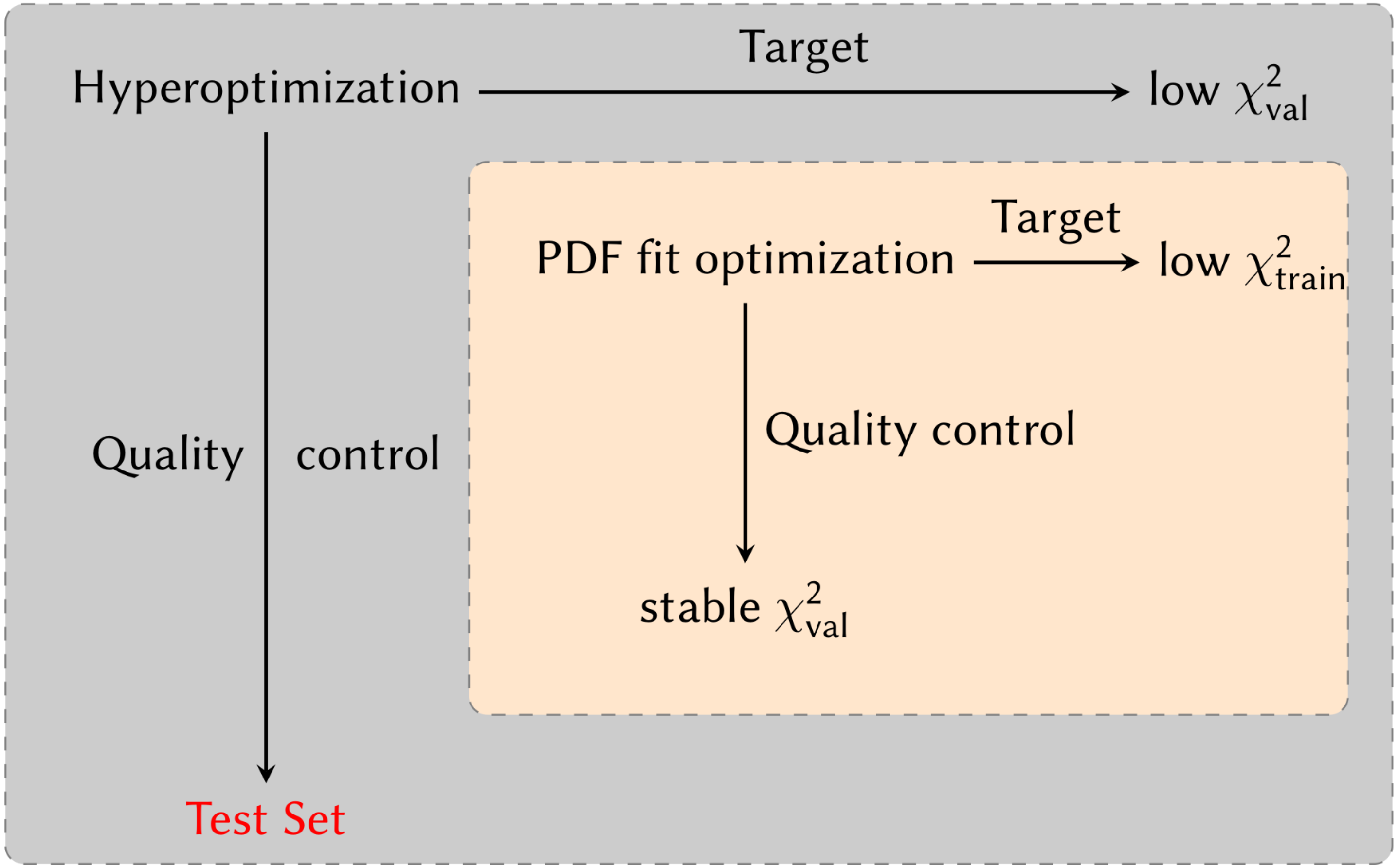
HYPEROPTIMIZATION



WE ARE MISSING A CONTROL CRITERION

THE SOLUTION

THE TEST SET



COMPARE TO A A TEST SET $\Rightarrow$ NEW DATA PREVIOUSLY NOT USED AT ALL
TESTS GENERALIZATION POWER

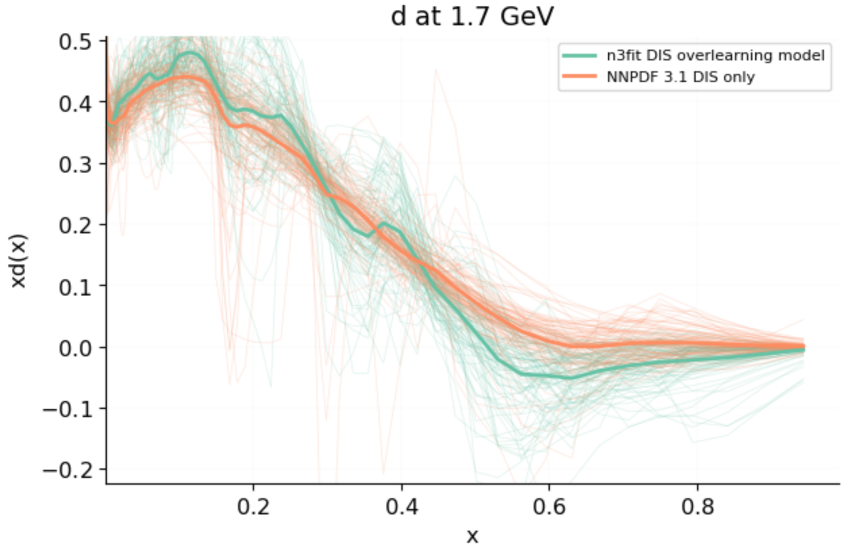
TEST SET RESULS

- COMPLETELY UNCORRELATED TEST SET
- OPTIMIZE ON WEIGHTED AVERAGE OF VALIDATION AND TEST
  $\Rightarrow$ NO OVERLEARNING

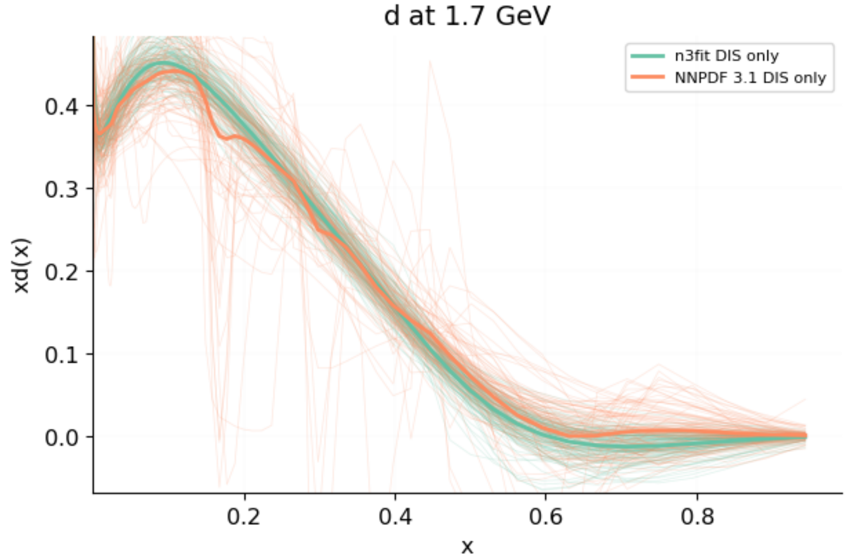
HYPEROPTIMIZED PDFs

DOWN QUARK

OVERFIT VS HAND-PICKED



TEST-SET VS HAND-PICKED



BUT WHO PICKS THE TEST SET?

K-FOLDS

THE BASIC IDEA:

- DIVIDE THE DATA INTO k REPRESENTATIVE SUBSETS
  EACH CONTAINING PROCESS TYPES, KINEMATIC RANGE OF FULL SET
- TRAIN $k-1$ SETS AND USE k-TH SET AS TEST
  $\Rightarrow k$ VALUES OF $\chi^2_{\text{test, i}}$

Fold 1		
CHORUS σ^{ν}_{CC}	HERA I+II inc NC e^+p 920 GeV	BCDMS p
LHCb Z 940 pb	ATLAS W, Z 7 TeV 2010	CMS Z p_T 8 TeV (p^{ll}_T, y_{ll})
DY E605 $\sigma^p_{\rm DY}$	CMS Drell-Yan 2D 7 TeV 2011	CMS 3D dijets 8 TeV
ATLAS single-$\bar{t}$ y (normalised)	ATLAS single top R_t 7 TeV	CMS $t\bar{t}$ rapidity $y_{t\bar{t}}$
CMS single top R_t 8 TeV		

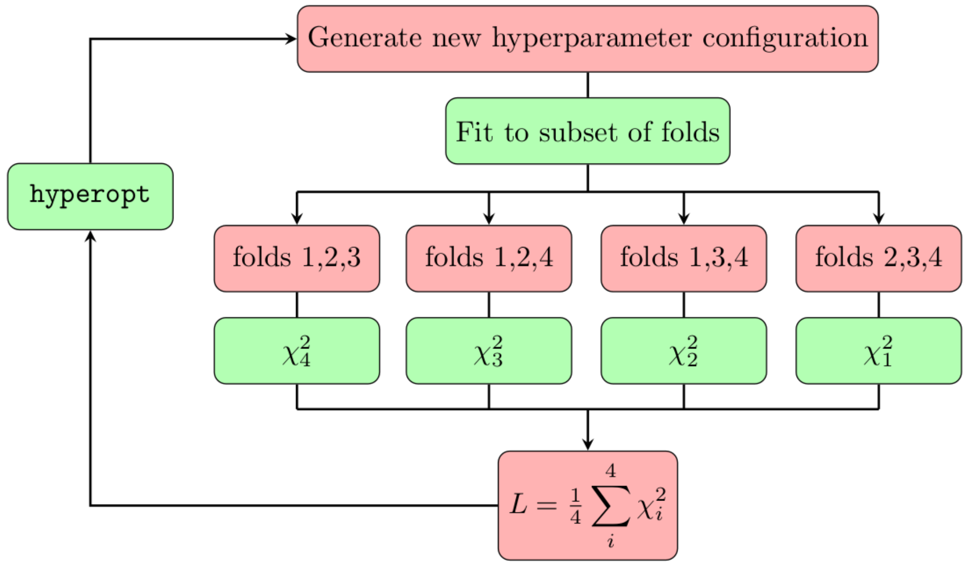
Fold 2		
HERA I+II inc CC e^-p	HERA I+II inc NC e^+p 460 GeV	HERA comb. $\sigma^{\rm red}_{b\bar{b}}$
NMC p	NuTeV $\sigma^{\bar{\nu}}_c$	LHCb $Z \to ee$ 2 fb
CMS W asymmetry 840 pb	ATLAS Z p_T 8 TeV (p^{ll}_T, M_{ll})	D0 $W \to \mu\nu$ asymmetry
DY E886 $\sigma^p_{\rm DY}$	ATLAS direct photon 13 TeV	ATLAS dijets 7 TeV, R=0.6
ATLAS single antitop y (normalised)	CMS $\sigma^{\rm tot}_{tt}$	CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV

Fold 3		
HERA I+II inc CC e^+p	HERA I+II inc NC e^+p 575 GeV	NMC d/p
NuTeV σ^{ν}_c	LHCb $W, Z \to \mu$ 7 TeV	LHCb $Z \to ee$
ATLAS W, Z 7 TeV 2011 Central selection	ATLAS W^++jet 8 TeV	ATLAS HM DY 7 TeV
CMS W asymmetry 4.7 fb	DYE 866 $\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$	CDF Z rapidity (new)
ATLAS $\sigma^{\rm tot}_{tt}$	ATLAS single top y_t (normalised)	CMS $\sigma^{\rm tot}_{tt}$ 5 TeV
CMS $t\bar{t}$ double diff. $(m_{t\bar{t}}, y_t)$		

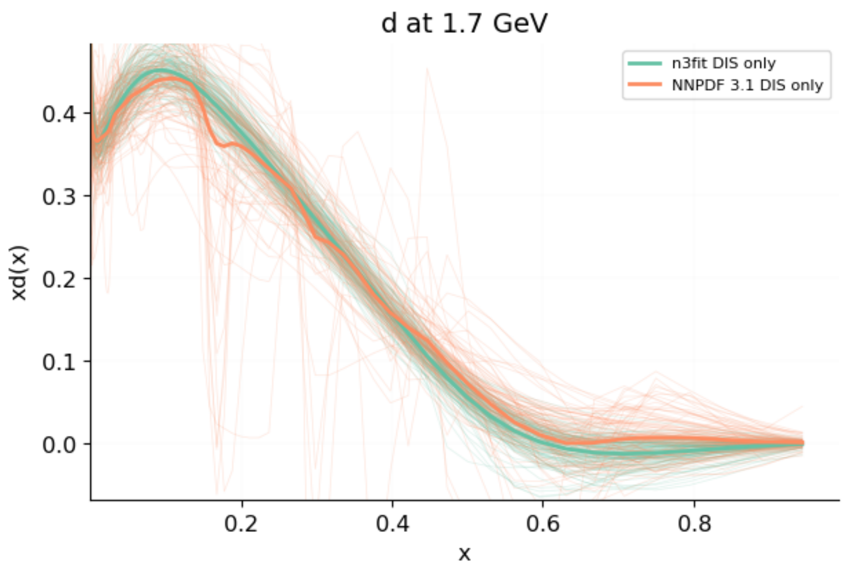
Fold 4		
CHORUS $\sigma^{\bar{\nu}}_{CC}$	HERA I+II inc NC e^+p 820 GeV	LHCb $W, Z \to \mu$ 8 TeV
LHCb $Z \to \mu\mu$	ATLAS W, Z 7 TeV 2011 Fwd	ATLAS W^-+jet 8 TeV
ATLAS low-mass DY 2011	ATLAS Z p_T 8 TeV (p^{ll}_T, y_{ll})	CMS W rapidity 8 TeV
D0 Z rapidity	CMS dijets 7 TeV	ATLAS single top y_t (normalised)
ATLAS single top R_t 13 TeV	CMS single top R_t 13 TeV	

K-FOLD VALIDATION

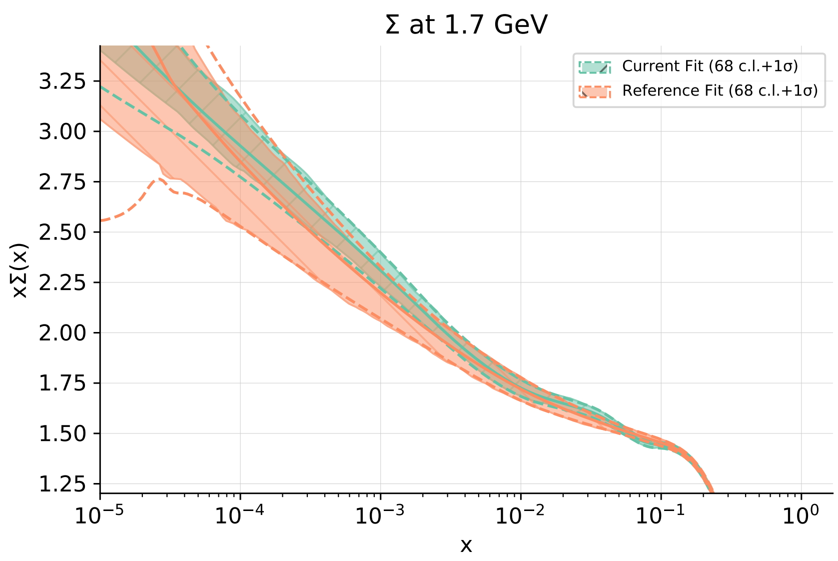
LOSS: AVERAGE χ^2 OF NON-FITTED FOLDS



TEST-SET VS HAND-PICKED



K-FOLD VS. TEST-SET



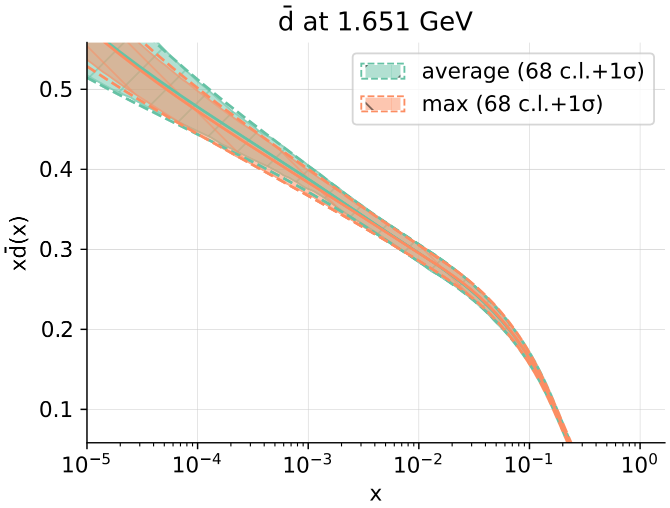
K-FOLD VALIDATION: RESULTS AND STABILITY

HYPEROPTIMIZED PARAMETERS

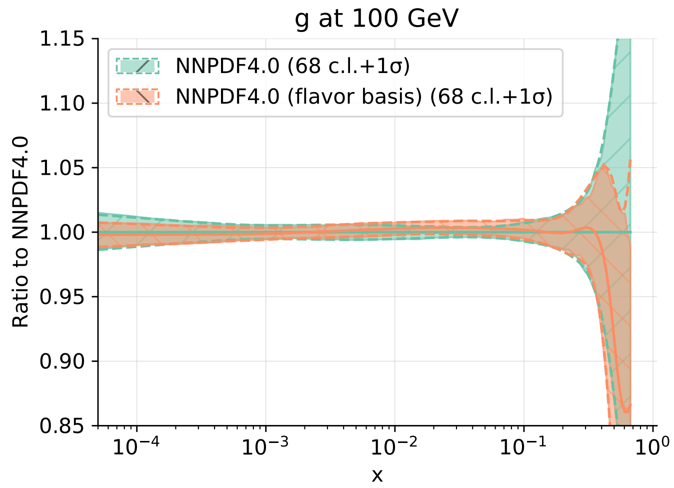
Parameter	NNPDF4.0	L as in Eq. (3.21)	Flavour basis Eq. (3.2)
Architecture	25-20-8	70-50-8	7-26-27-8
Activation function	hyperbolic tangent	hyperbolic tangent	sigmoid
Initializer	`glorot_normal`	`glorot_uniform`	`glorot_normal`
Optimizer	`Nadam`	`Adadelta`	`Nadam`
Clipnorm	6.0×10^{-6}	5.2×10^{-2}	2.3×10^{-5}
Learning rate	2.6×10^{-3}	2.5×10^{-1}	2.6×10^{-3}
Maximum # epochs	17×10^{3}	45×10^{3}	45×10^{3}
Stopping patience	10% of max epochs	12% of max epochs	16% of max epochs
Initial positivity $\Lambda^{(\mathrm{pos})}$	185	106	2
Initial integrability $\Lambda^{(\mathrm{int})}$	10	10	10

- DIFFERENT CHOICES OF LOSS: $L = \frac{1}{n_{\mathrm{fold}}} \sum_{k=1}^{n_{\mathrm{fold}}} \chi_k^2$ vs. $L = \max\left(\chi_1^2, \chi_2^2, \chi_3^2, \ldots, \chi_{n_{\mathrm{fold}}}^2\right)$
- PDF FLAVOR VS. EVOLUTION BASIS

CHANGE OF LOSS (ANTIDOWN)



CHANGE OF BASIS (GLUON)



VALIDATING AND UNDERSTANDING

VALIDATION CLOSURE TESTS

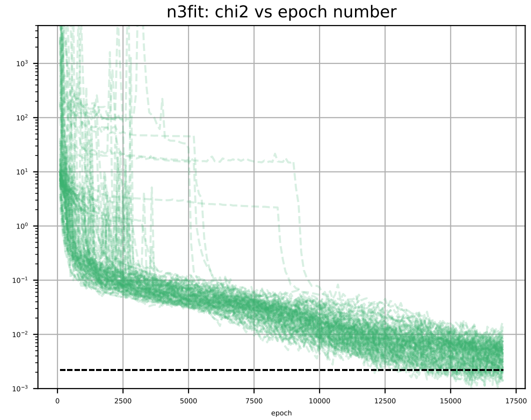
- ASSUME UNDERLYING "TRUTH" PDF (SAY A RANDOM PDF REPLICA)
- GENERATE DATA WITH STATISTICAL AND SYSTEMATIC SHIFTS
- DETERMINE PDFS & COMPARED TO "TRUTH"

THE NATURE OF UNCERTAINTIES

- LEVEL 0:
  - DATAPOINT EQUAL TO THE "TRUTH VALUE"; ZERO UNCERTAINTY
  - MUST FIND $\chi^2 = 0$ ("TRUTH")
  - INTERPOLATION/EXTRAPOLATION UNCERTAINTY
- LEVEL 1:
  - PSEUDO- DATAPOINTS $\Rightarrow$ FLUCTUATIONS ABOUT "TRUTH" $\Rightarrow$ "RUN OF THE UNIVERSE"
  - FIT DATA OVER AND OVER AGAIN
  - $\chi^2 \approx 1$
  - FUNCTIONAL UNCERTAINTY
- LEVEL 2:
  - DATA AS IN LEVEL 1
  - DATA REPLICAS OF THESE "DATA"
  - FIT PDF REPLICAS TO DATA REPLICAS
  - $\chi^2 \approx 2$ REPLICA TO REPLICA; $\chi^2 \approx 1$ AVERAGE TO TRUTH
  - DATA UNCERTAINTY

UNCERTAINTIES: TYPE AND SIZE
CLOSURE TEST RESULTS (NNPDF4.0)

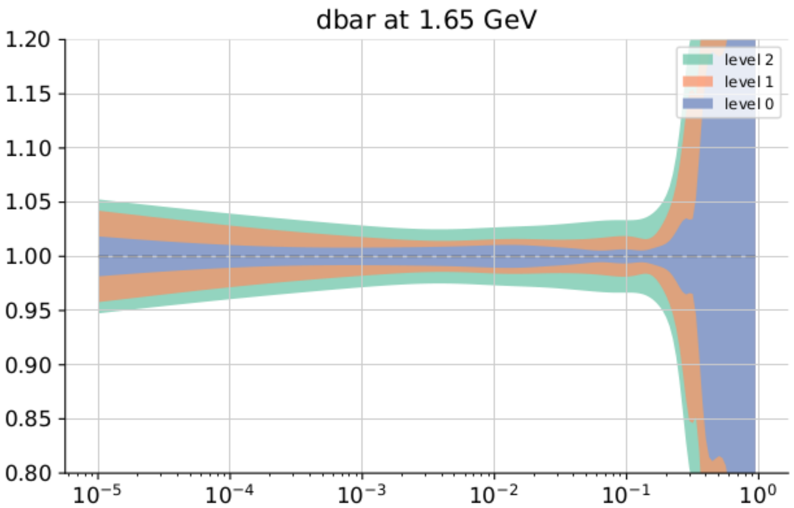
LEVEL 0 χ^2 VS TRAINING



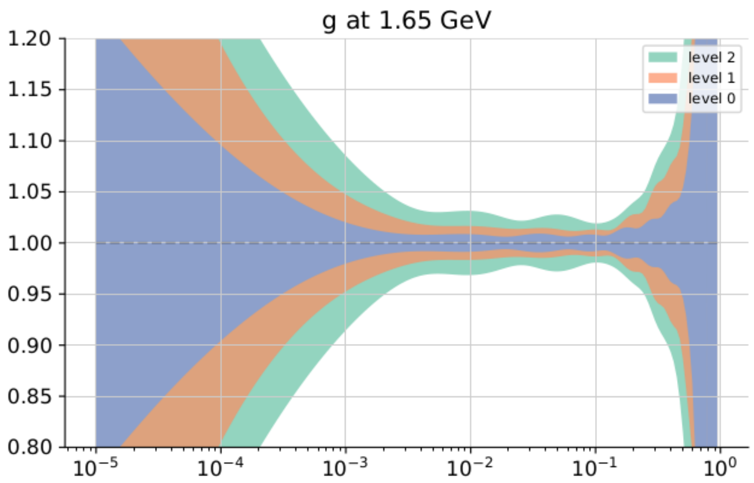
- LEVEL 0 (TRUTH DATA) $\Rightarrow \chi^2 \approx 0$, YET UNCERTAINTY NONZERO
  $\Rightarrow$ NEURAL NETS $\Leftrightarrow$ MANY FUNCTIONAL FORMS
- LEVEL 1 (RUNS OF UNIVERSE) $\Rightarrow$ REPLICAS ALL FITTED TO SAME DATA,
  YET UNCERTAINTY NONZERO
  $\Rightarrow$ DITTO
- LEVEL 0, 1 AND 2 UNCERTAINTIES COMPARABLE IN SIZE

LEVEL 0/1/2 UNCERTAINTIES

ANTIDOWN



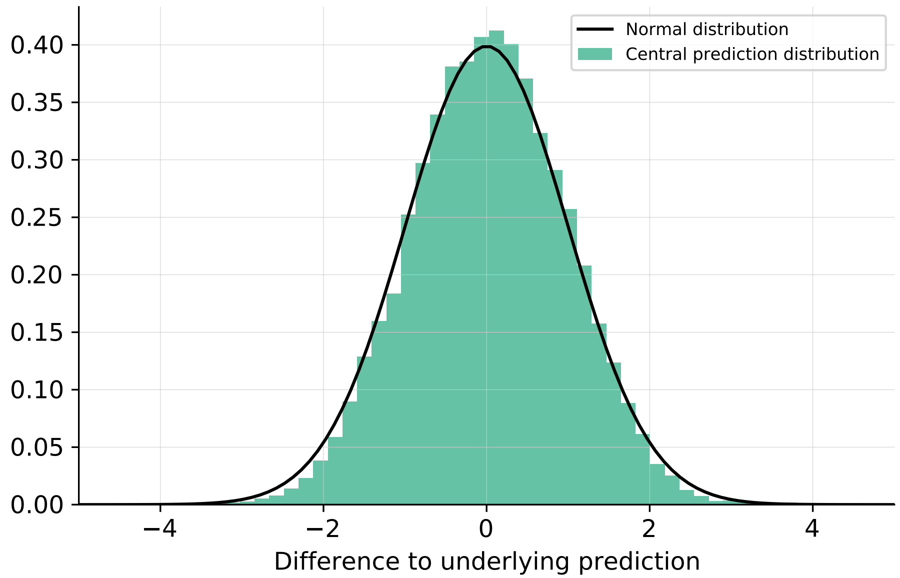
GLUON



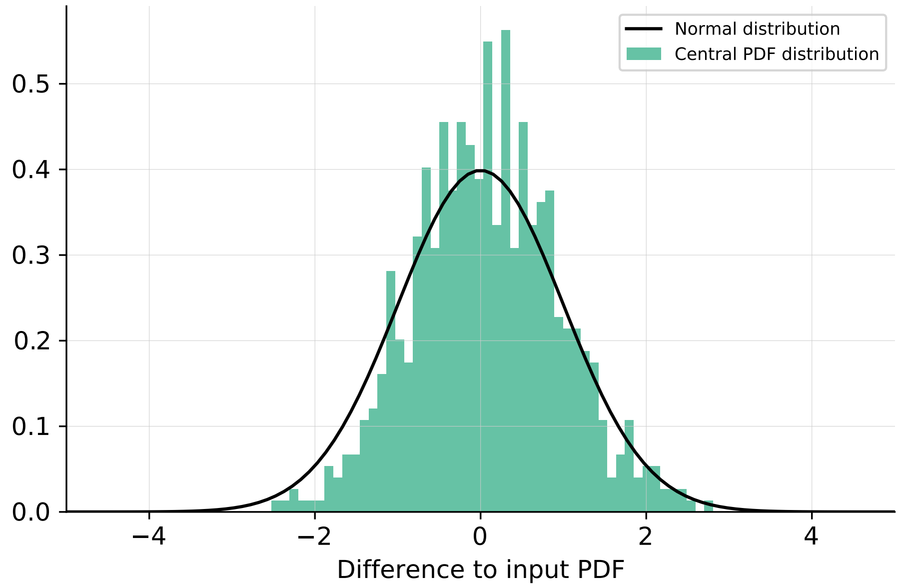
TESTING: THE INDICATORS

DISTRIBUTION OF DEVIATIONS FROM TRUTH

DATA SPACE (OUT OF SAMPLE)



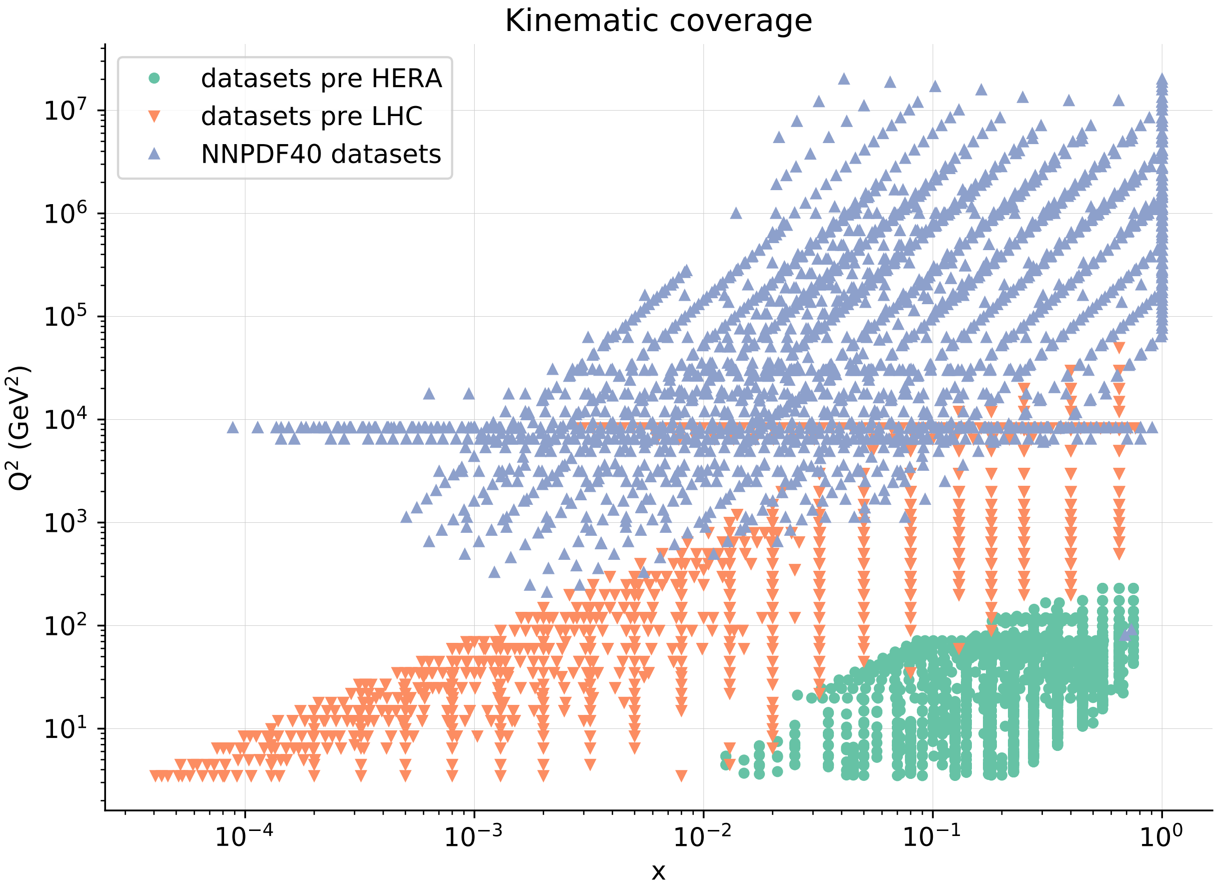
PDF SPACE



- PDF-SPACE MORE NOISY THAN DATA SPACE

FUTURE TESTS

IDEA: USE (REAL) HIERARCHICAL DATASETS



- DEFINE "PRE-HERA", " PRE-LHC" AND "CURRENT" DATASETS
  EACH LATER DATASET IS EXTRAPOLATION OF PREVIOUS
- DETERMINE PDFs & COMPARE TO "FUTURE" DATA
- COMPUTE χ^2 TO FUTURE DATA:
  - WITHOUT PDF UNCERTAINTIES $\Rightarrow$ IF $\gg 1$, MISSING INFORMATION
  - WITH PDF UNCERTAINTY $\Rightarrow$ IF ~ 1, TEST PASSED
    MISSING INFO REPRODUCED BY UNCERTAINTY

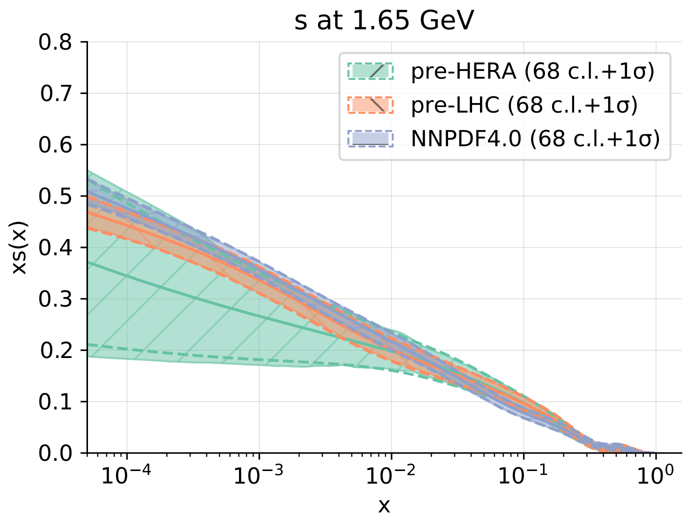
ASSESSING EXTRAPOLATION UNCERTAINTIES

FUTURE TEST RESULTS (NNPDF4.0)

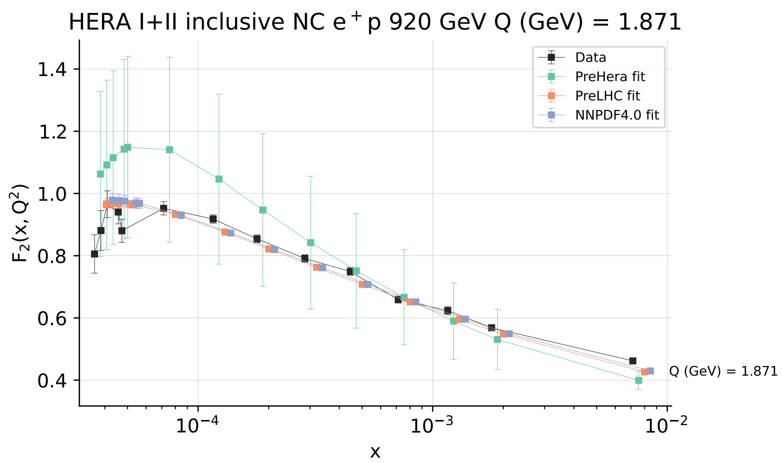
χ^2: FITTED VS EXTRAPOLATED: WITHOUT/WITH PDF UNC.

PROCESS	PRE-HERA	PRE-LHC	NNPDF4.0
FT DIS (NC)	1.05	1.18	1.23
FT DIS (CC)	0.80	0.85	0.87
FT DY	0.92	1.27	1.59
HERA	27.20/1.23	1.22	1.20
COLL. DY (TEV.)	5.52/1.02	0.99	1.11
COLL. DY (LHC)	18.91/1.31	2.63/1.58	1.53
TOP QUARK	20.01/1.06	1.30/0.87	1.01
JETS	2.69/0.98	2.12/1.10	1.26
TOTAL OUT OF SAMPLE	19.48/1.16	2.10/1.15	–

strange PDF



HERA F_2^p

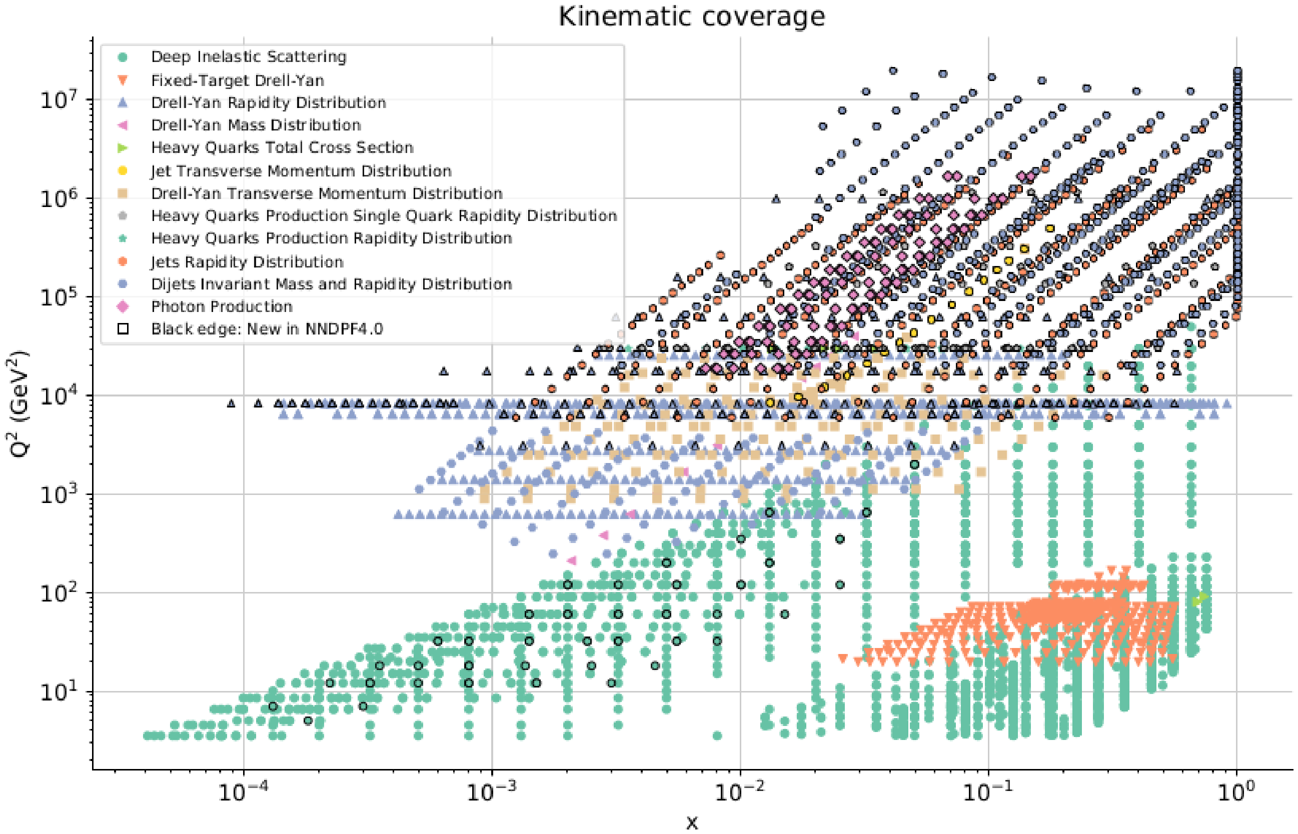


ML PREDICTS THE RISE OF F_2 AT HERA

CONTEMPORARY PDF DETERMINATION

THE DATA

Experimental data in NNPDF4.0



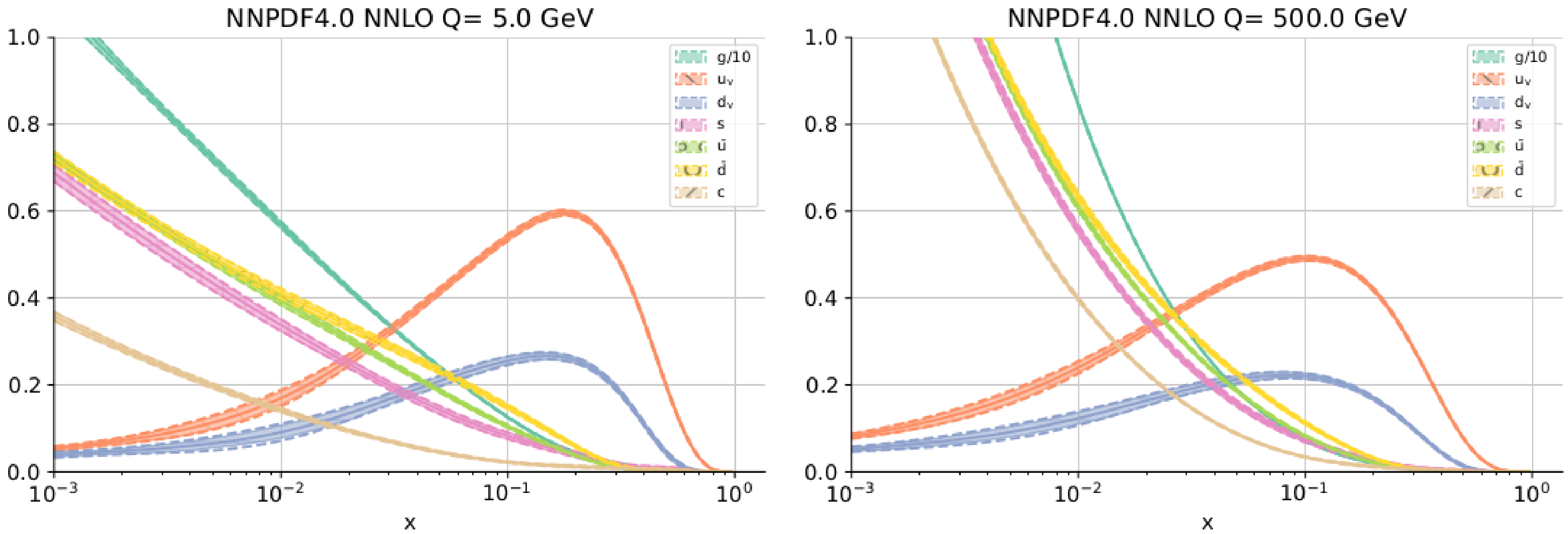
More than 4000 datapoints!

New processes:

- direct photon
- single top
- dijets
- W+jet
- DIS jet

CONTEMPORARY PDF DETERMINATION

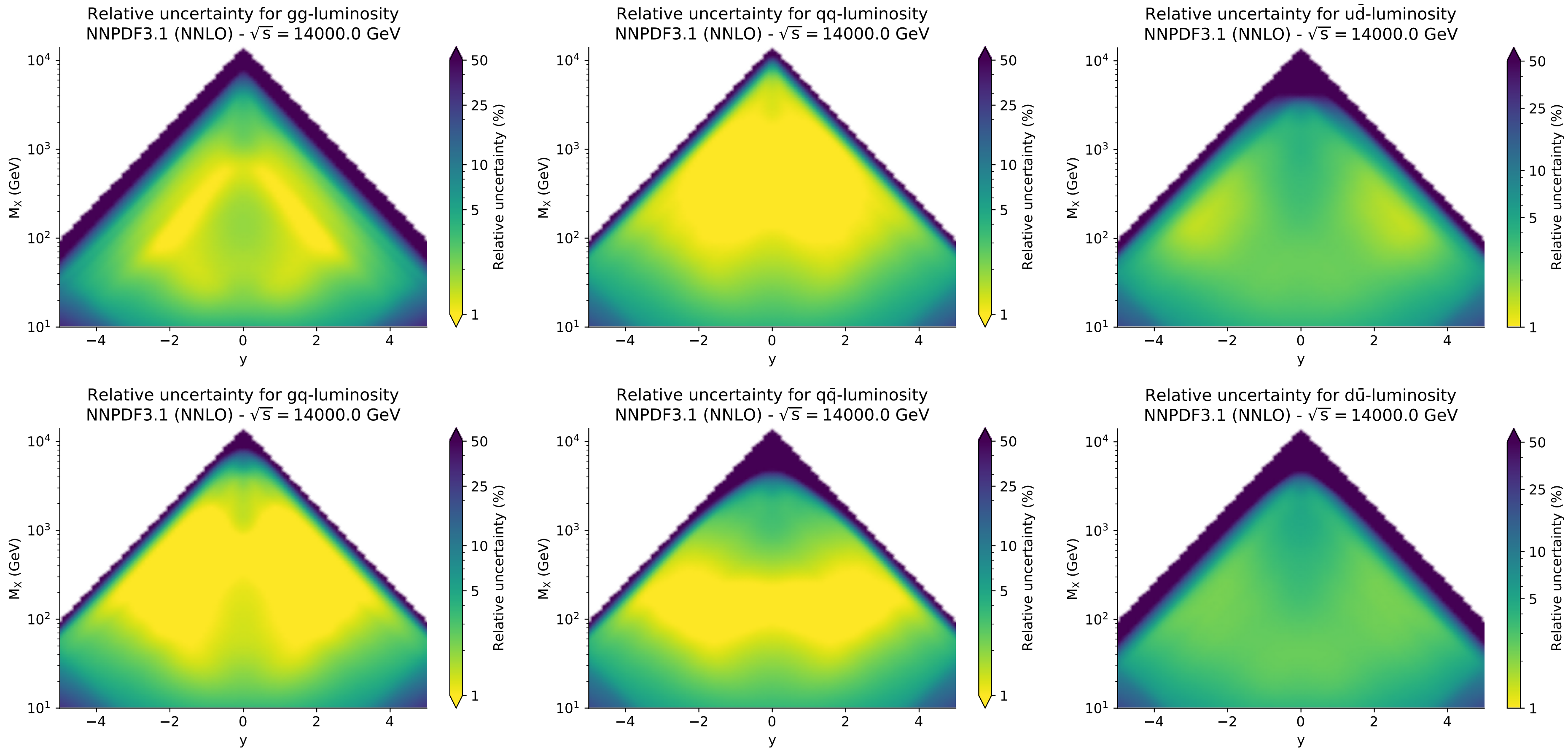
THE PDFs



CONTEMPORARY PDF DETERMINATION

THE UNCERTAINTIES (2016)

GLUON SINGLET FLAVORS

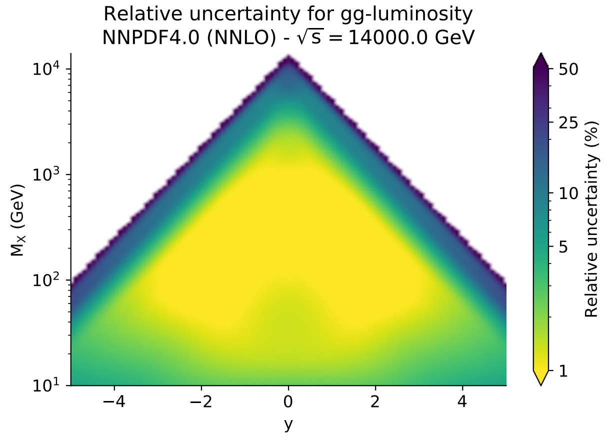


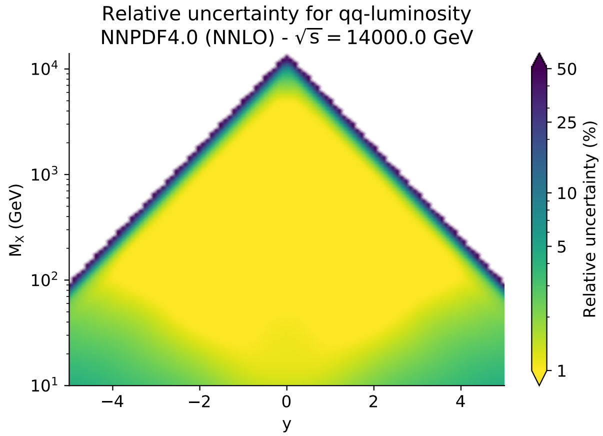
- TYPICAL UNCERTAINTIES IN DATA REGION: SINGLET $\sim 3\%$, NONSINGLET $\sim 5\%$
- DATA REGION: $10^2 \lesssim M_X \lesssim 10^3$ TeV, $-2 \lesssim y \lesssim 2$

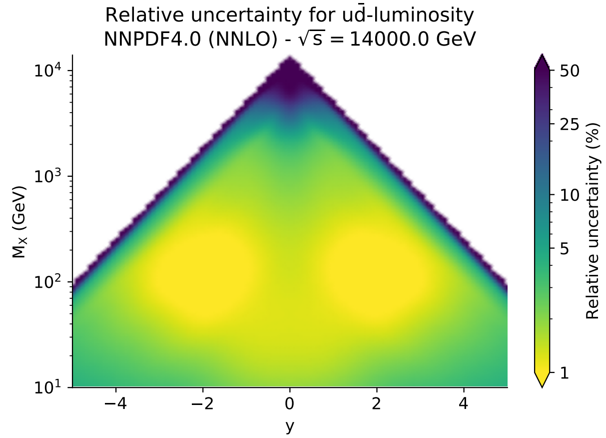
CONTEMPORARY PDF DETERMINATION

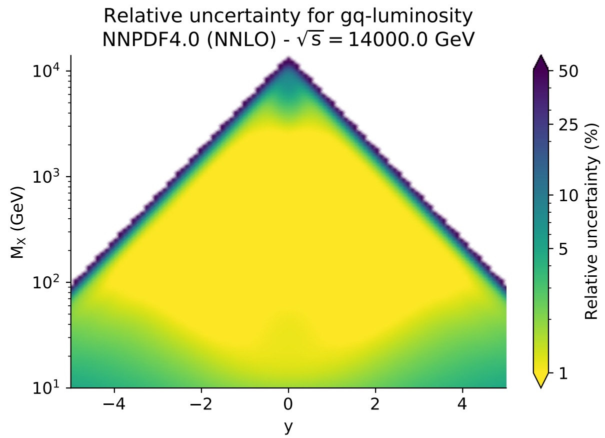
THE UNCERTAINTIES (2022)

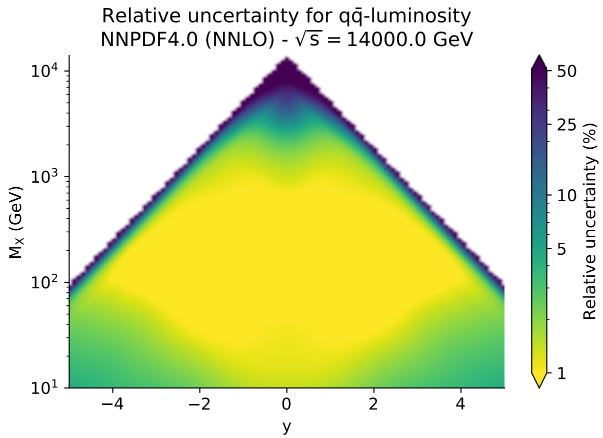
GLUON | SINGLET | FLAVORS

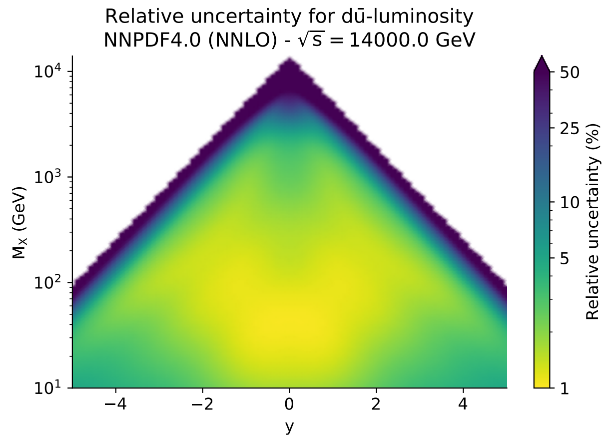










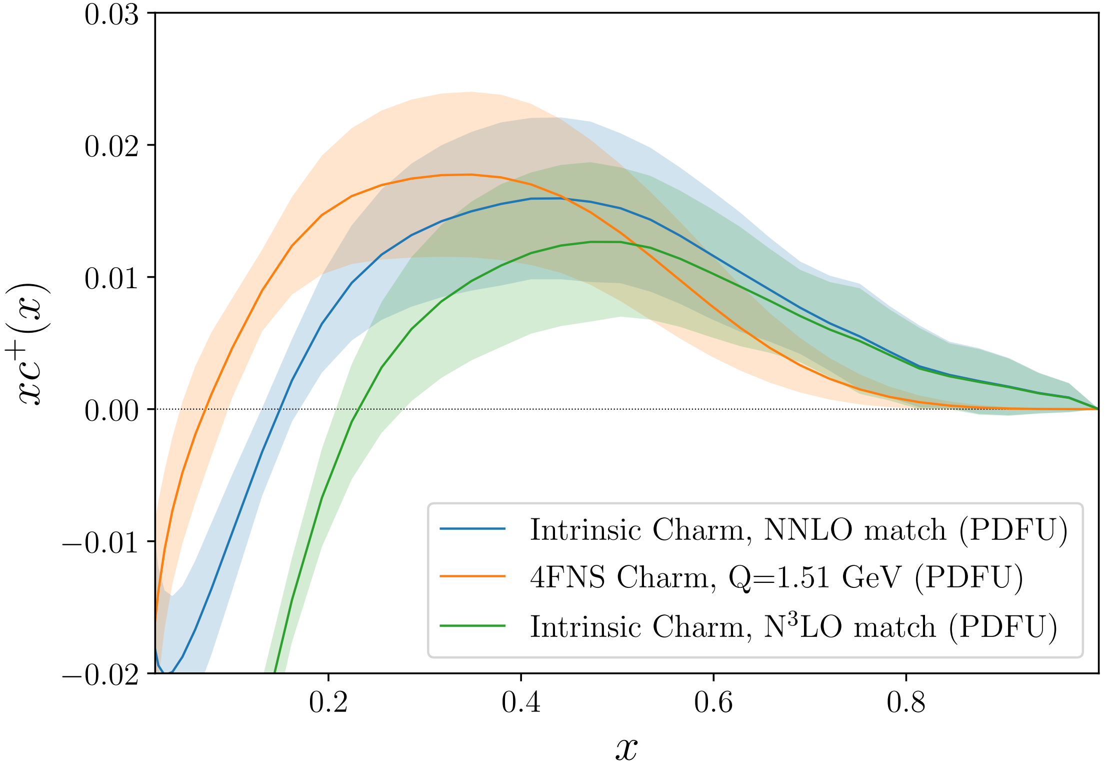


- TYPICAL UNCERTAINTIES IN DATA REGION: SINGLET $\sim 1\%$, NONSINGLET $\sim 2-3\%$
- DATA REGION: $10 \lesssim M_X \lesssim 3 \cdot 10^3$ TeV, $-4 \lesssim y \lesssim 4$

CONTEMPORARY PDF DETERMINATION:
THE NEED FOR THEORETICAL ACCURACY
INTRINSIC CHARM

- PERTURBATIVE CHARM ($N_f = 4$) DETERMINED BY MATCHING CONDITIONS
- LARGE HIGHER ORDER CORRECTIONS ⇒ N^{3}LO AVAILABLE (Blümlein, Ablinger et al.)
- INTRINSIC CHARM ⇒ INVERT MATCHING CONDITIONS
  INVERSION ⇒ EKO CODE (Candido, Hekhorn, Magni, 2022)

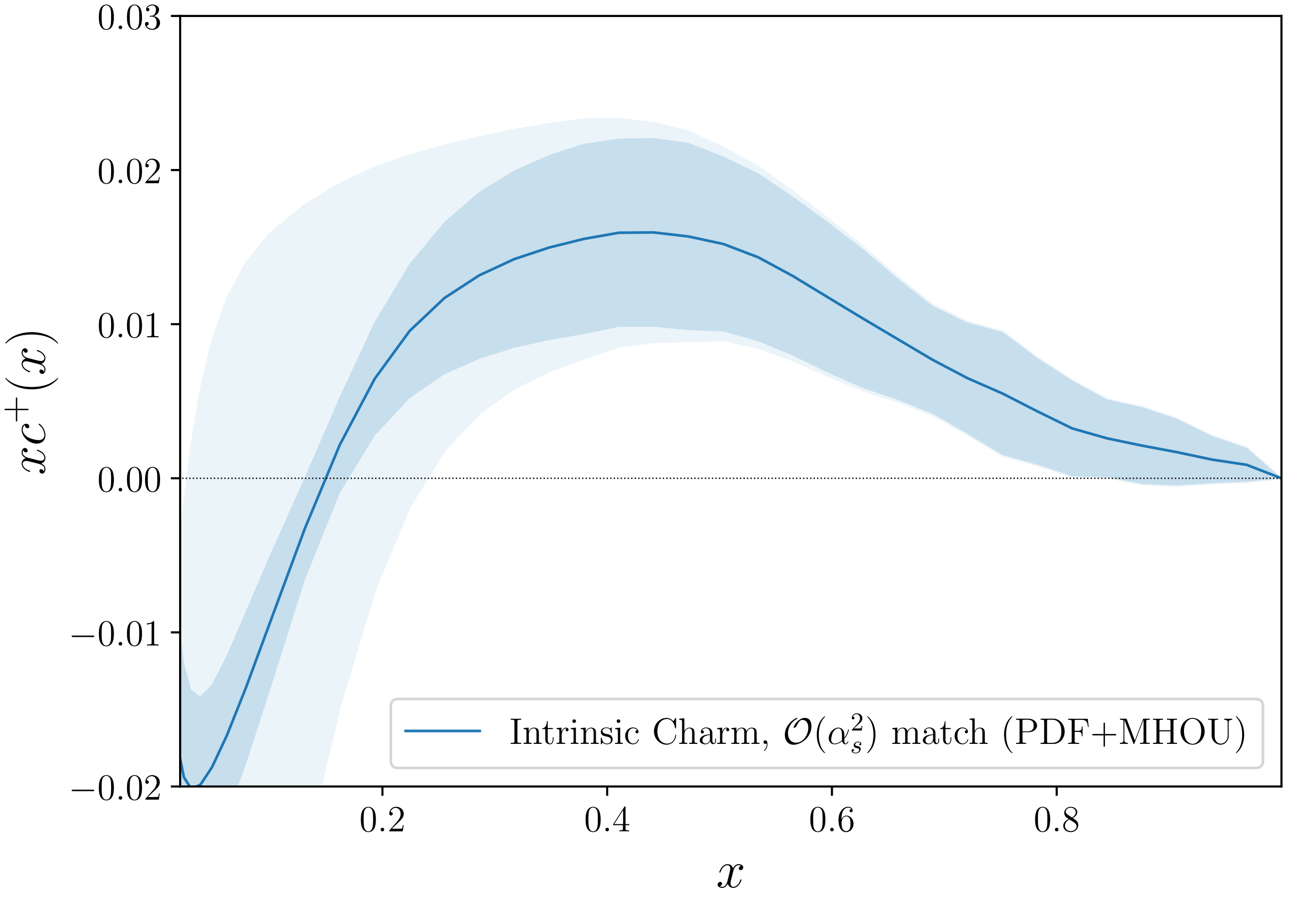
CHARM PDF: $N_f = 4$ VS $N_f = 3$ (NNLO & N^{3}LO CONVERSION)



CONTEMPORARY PDF DETERMINATION

INTRINSIC CHARM!

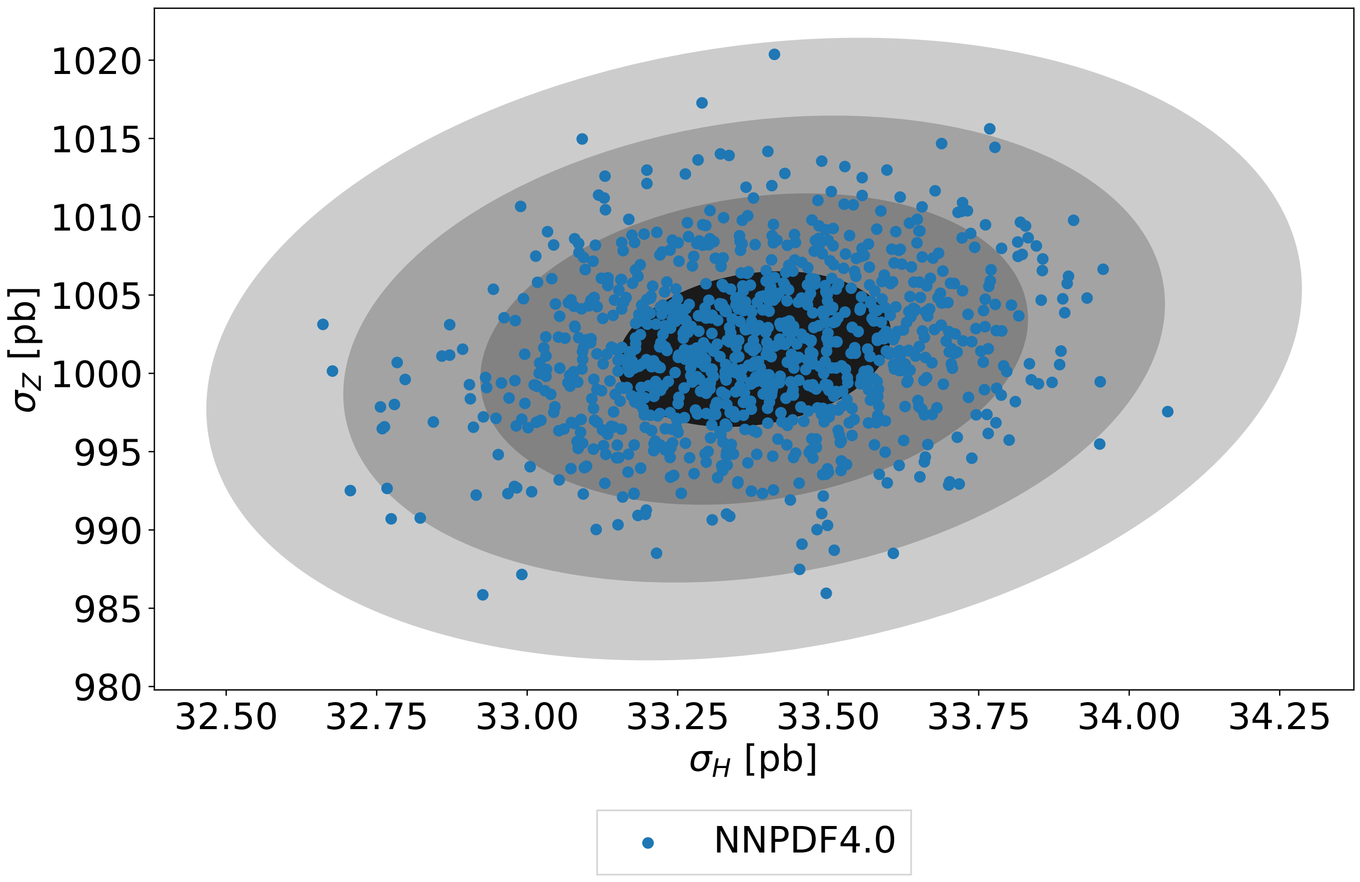
- MHOU ESTIMATED FROM N^3LO-NNLO DIFFERENCE
  - LARGE UNCERTAINTY AT SMALL x
  - NEGLIGIBLE UNCERTAINTY IN VALENCE REGION
- COMPATIBLE WITH ZERO AT SMALL x
- CLEAR EVIDENCE FOR INTRINSIC VALENCE PEAK



PDFs and XAI

CONTEMPORARY PDF DETERMINATION
DISTRIBUTION IN FUNCTION SPACE :

- PLOT RESULTS IN (σ_H, σ_Z) PREDICTION SPACE
- DISTRIBUTION OF REPLICAS $\Rightarrow$ IMPORTANCE SAMPLING OF UNDERLYING PROBABILITY

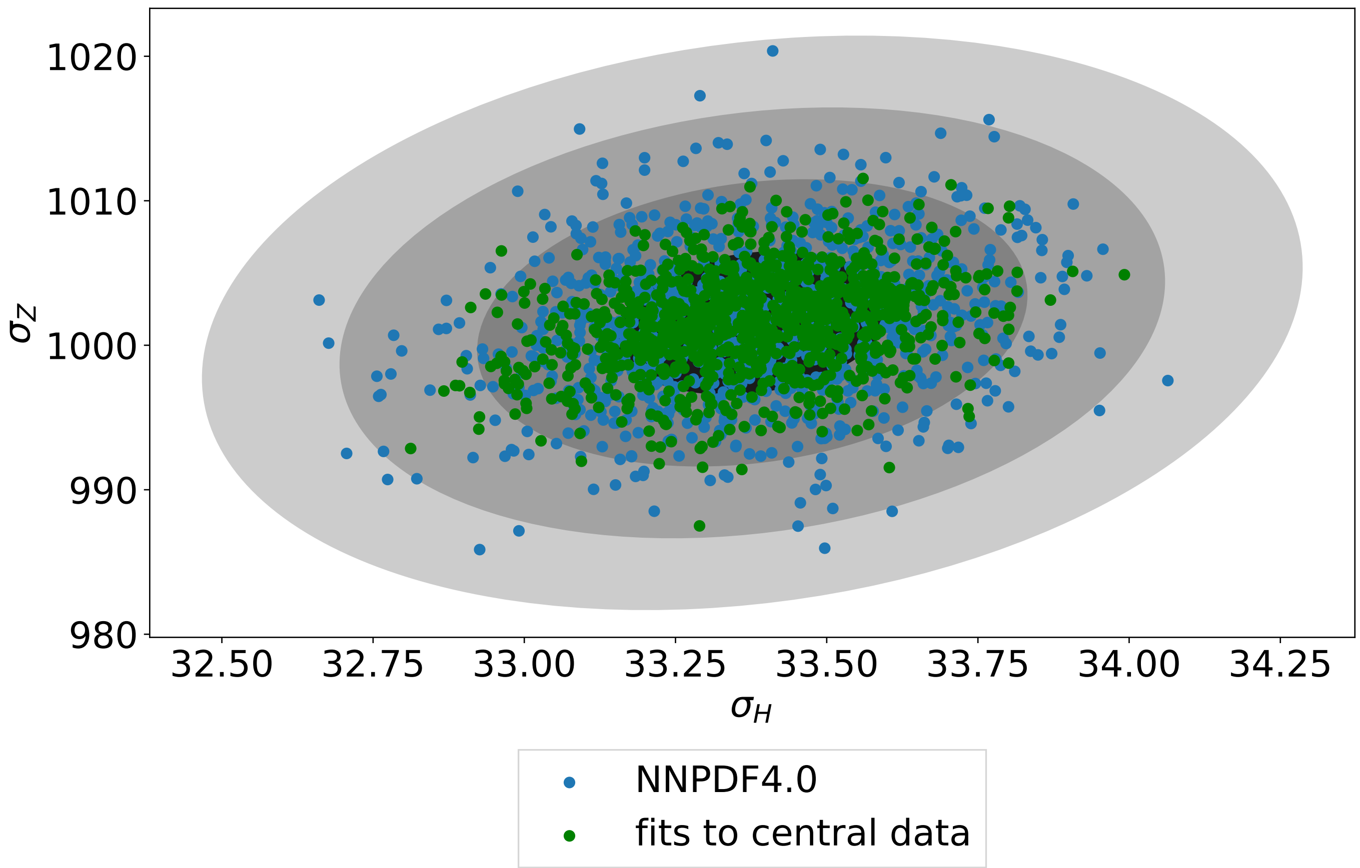


DISTRIBUTION OF REPLICAS DRIVEN BY

- DATA UNCERTAINTIES $\Rightarrow$ DATA REPLICA FLUCTUATION
- INTERPOLATION, EXTRAPOLATION AND FUNCTIONAL UNCERTAINTIES $\Rightarrow$ BEST FIT DEGENERACY

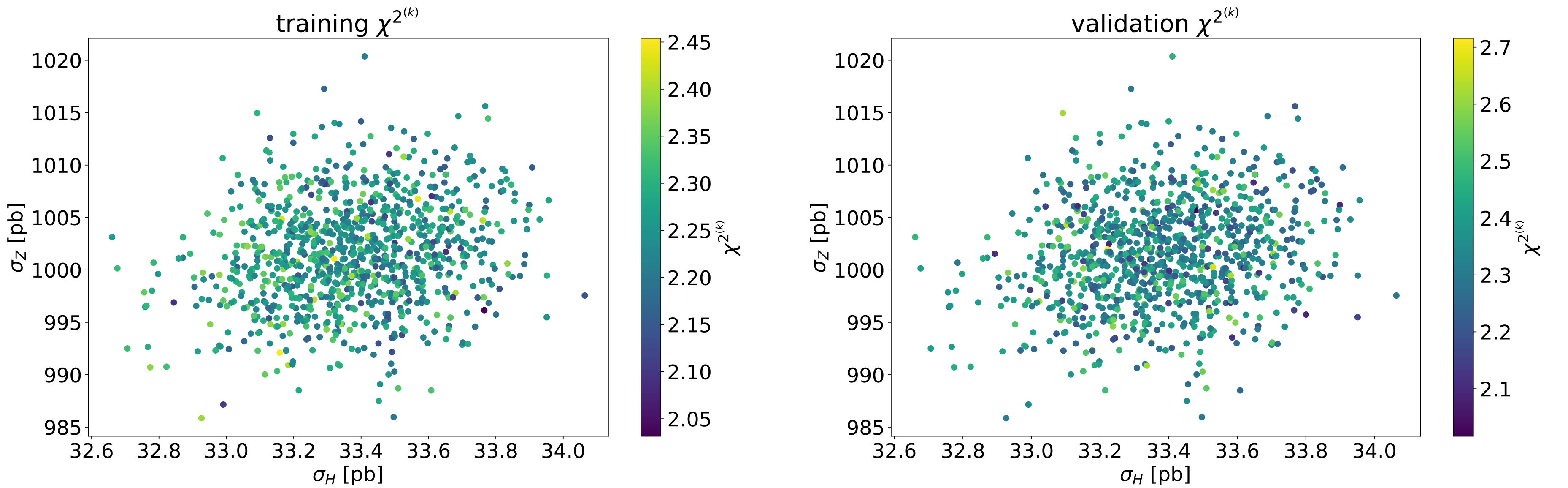
LEVEL-1 vs FULL UNCERTAINTIES

- REPLICA FLUCTUATION ⇒ DATA UNCERTAINTIES
- NO REPLICA FLUCTUATION ⇒ MODEL UNCERTAINTY



THE REPLICA DISTRIBUTION

LOSS QUALITY

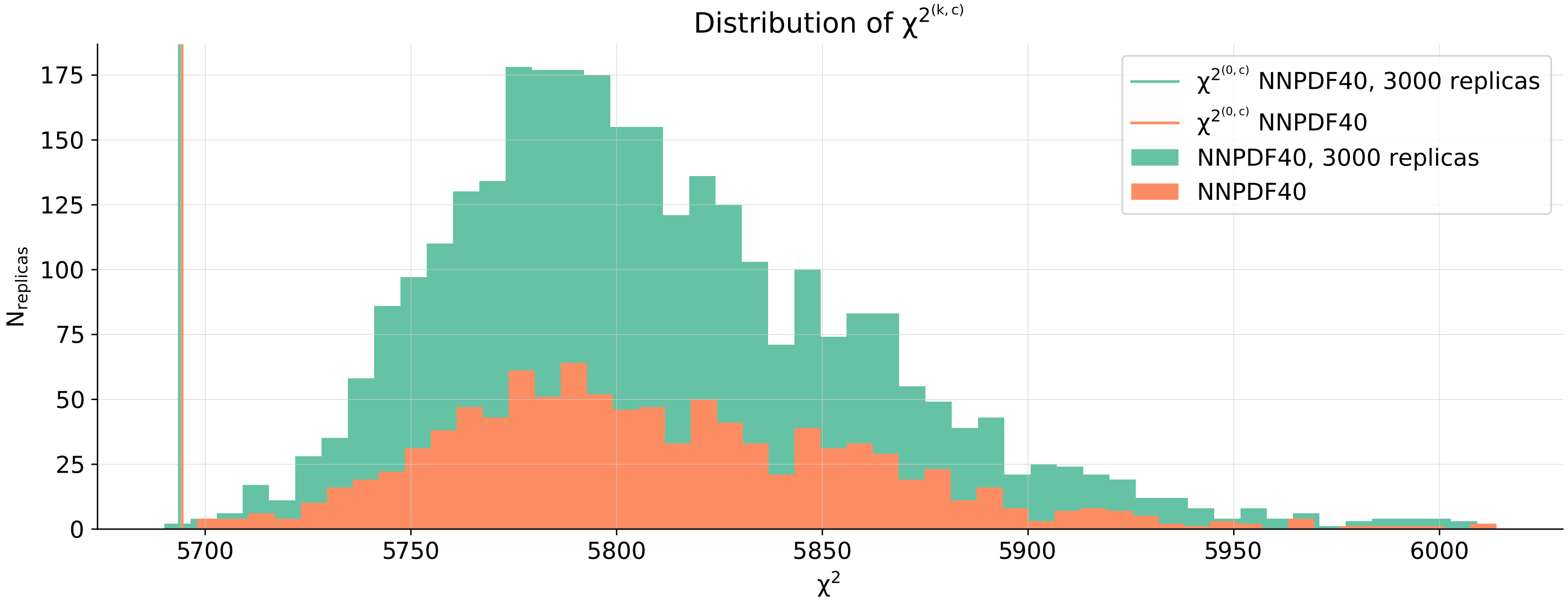


- COMPARE TRAINING AND VALIDATION LOSS FOR EACH REPLICA
- NO CORRELATION BETWEEN FIT QUALITY AND POSITION IN THE (σ_H, σ_Z) PLANE
- UNIFORM QUALITY

CONTEMPORARY PDF DETERMINATION
DISTRIBUTION IN FEATURE SPACE
LOSS TO CENTRAL DATA

- EACH PDF REPLICA FITTED TO A DATA REPLICA
- LOSS COMPUTED TO CENTRAL DATA STATISTICALLY DISTRIBUTED

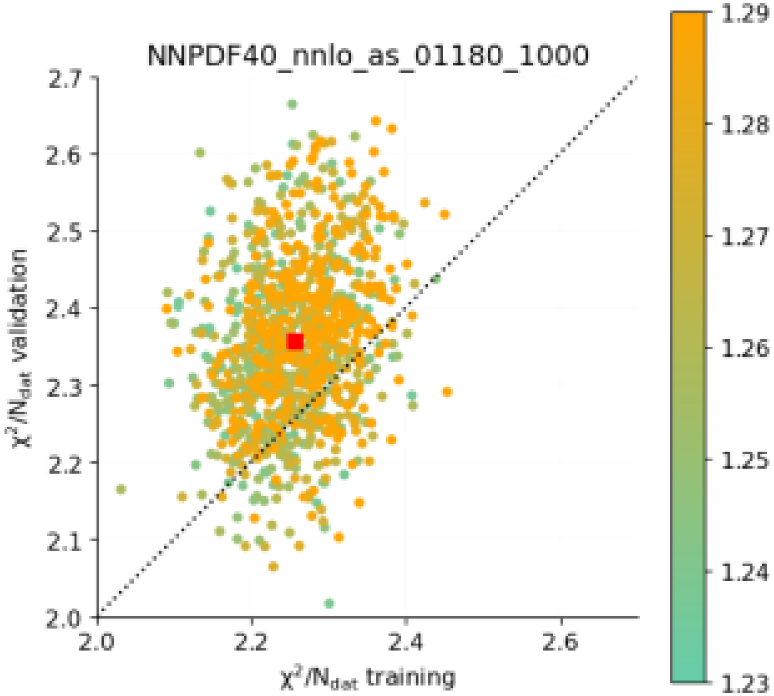
1000 REPLICAS VS. 3000 REPLICAS



- AVERAGE $\Rightarrow$ CENTRAL PREDICTION PDF $\Rightarrow$ LOW LOSS
- NOT NECESSARILY LOWEST

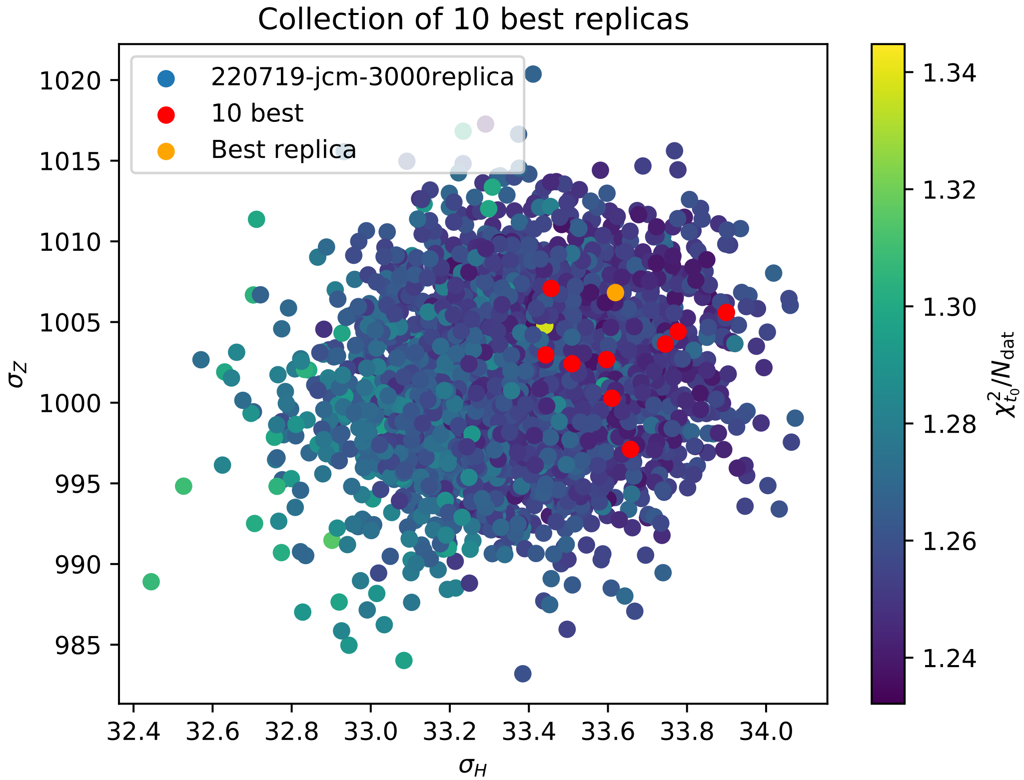
REPLICA LOSS DISTRIBUTION
TRAINING AND VALIDATION

- ARE FITS WITH HIGH LOSS TO CENTRAL DATA POOR (UNDERLEARNT)?



- NO CORRELATION BETWEEN LOSS TO CENTRAL DATA AND TRAINING, VALIDATION LOSS
- UNIFORM FIT QUALITY
- DISPERSION DUE
  - DATA REPLICA FLUCTUATION ⇒ DATA UNCERTAINTIES
  - MODEL UNCERTAINTIES
    ⇒ INTERPOLATION, EXTRAPOLATION AND FUNCTIONAL UNCERTAINTIES

FEATURE SPACE vs. FUNCTION SPACE: CORRELATION

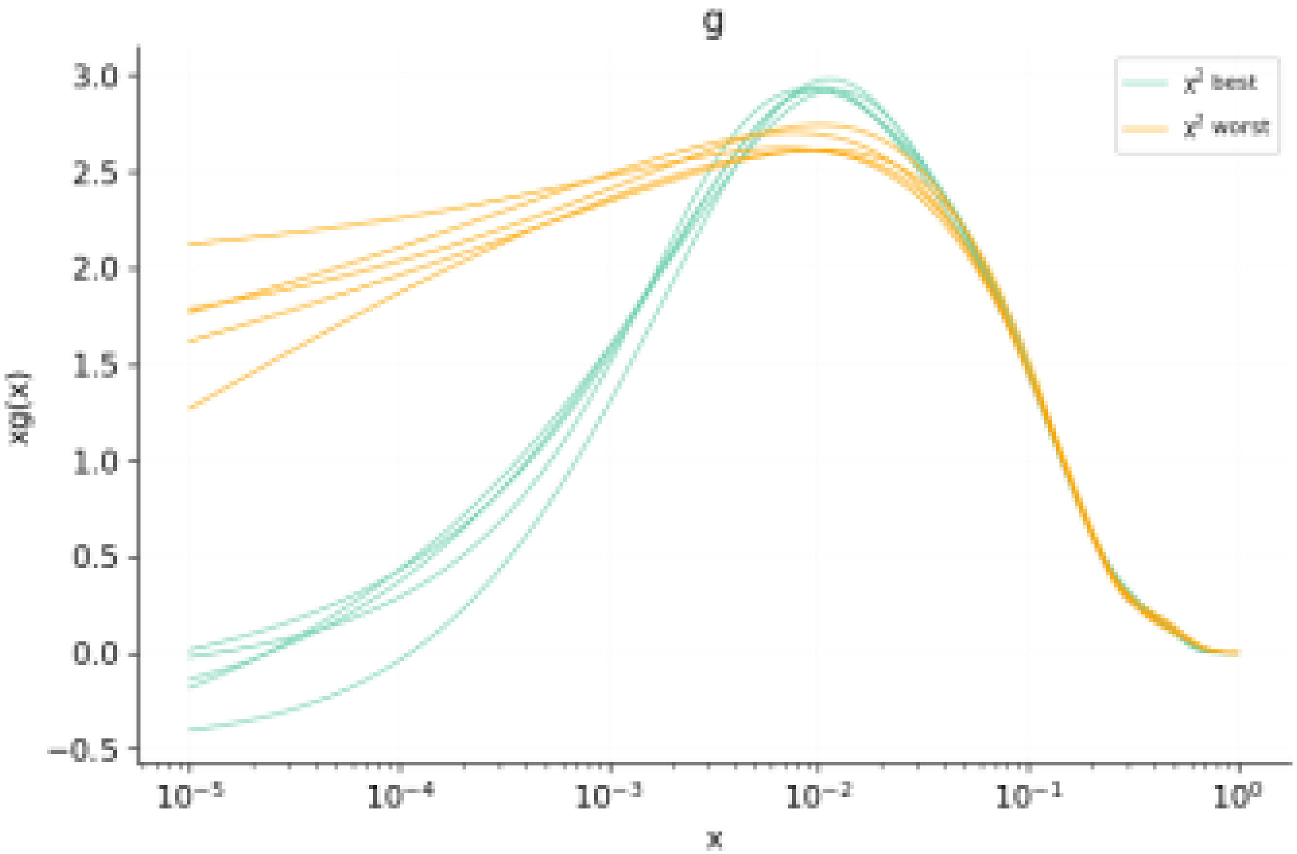


LOSS TO CENTRAL DATA

- CORRELATED TO POSITION IN (σ_H, σ_z) PLANE
- CORRELATED TO A FEATURE?

FEATURE SPACE VS. FUNCTION SPACE: REPLICAS WITH LOWEST & HIGHEST LOSS TO CENTRAL DATA

THE GLUON



- REPLICAS CLOSER TO CENTRAL DATA $\Rightarrow$ MORE STRUCTURE

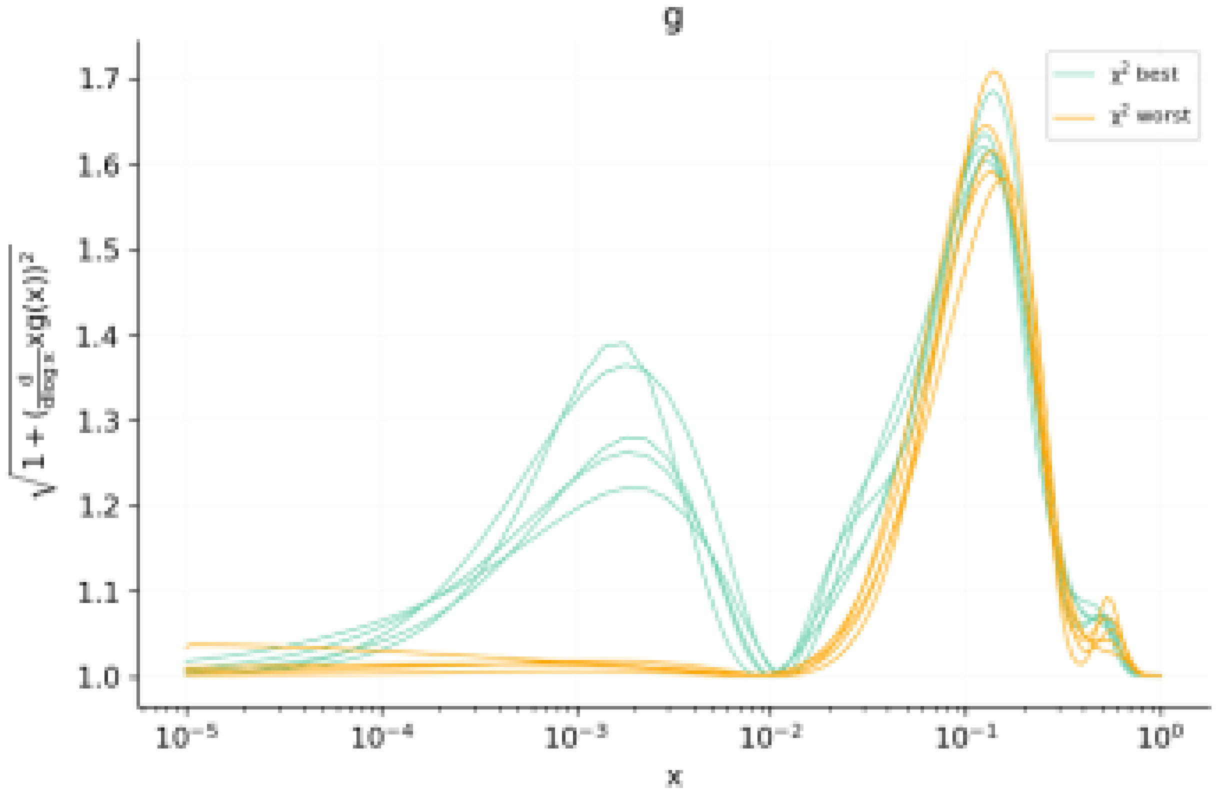
THE PDF KINETIC ENERGY

REPLICAS WITH LOWEST & HIGHEST LOSS TO CENTRAL DATA

$$\mathrm{KE} = \sqrt{1 + \left(\frac{d}{d\ln x} x f(x, Q^2)\right)^2}$$

ARCLENGTH OF THE NN OUTPUT IN TERMS OF INPUT

THE GLUON

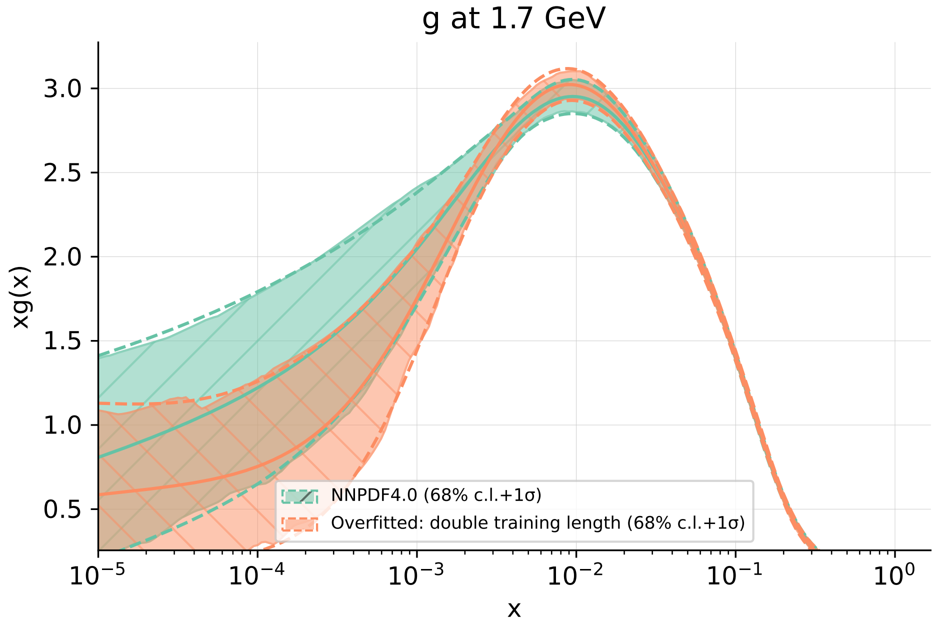


- REPLICAS CLOSER TO CENTRAL DATA $\Rightarrow$ MORE STRUCTURE
- HIGHER KINETIC ENERGY

OVERLEARNING FEATURES

- INDUCE OVERLEARING: DOUBLE TRAINING LENGTH

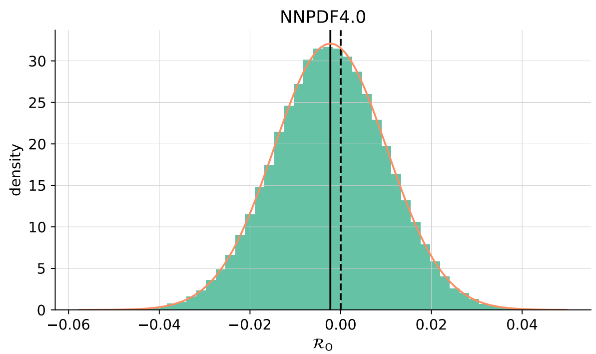
THE GLUON



- LOOK AT THE OUTPUT ⇒ MORE STRUCTURE IN GLUON

THE OVERFIT METRIC

DEFAULT NNPDF4.0



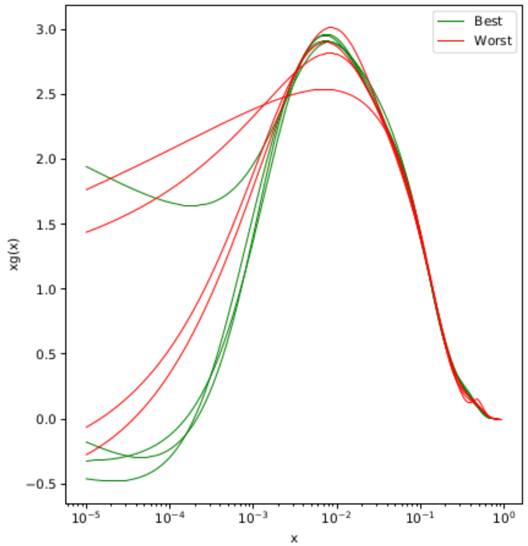
OVERFIT VARIANT

EXPLANATION
GENERALIZATION

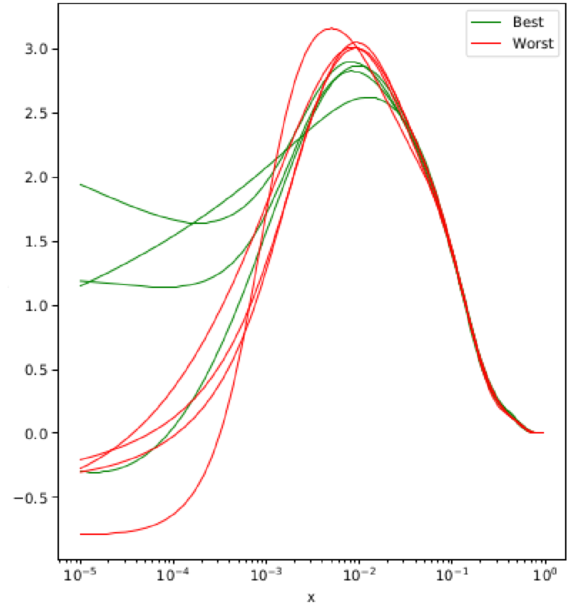
- OVERFITTING ⇒ POOR GENERALIZATION
- KEPT IN CHECK BY K-FOLDING (NOT CROSS-VALIDATION)
- LOOK AT BEST LOSS TO FITTED VS. EXCLUDED FOLDS

THE GLUON

FITTED FOLDS



EXCLUDED FOLD

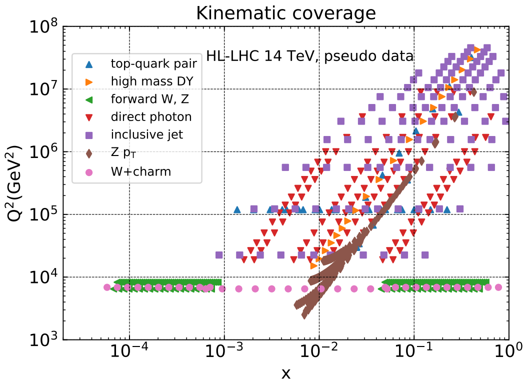


- BEST VS WORST REVERSED
- HIGH K.E. SOLUTIONS DO NOT GENERALIZE

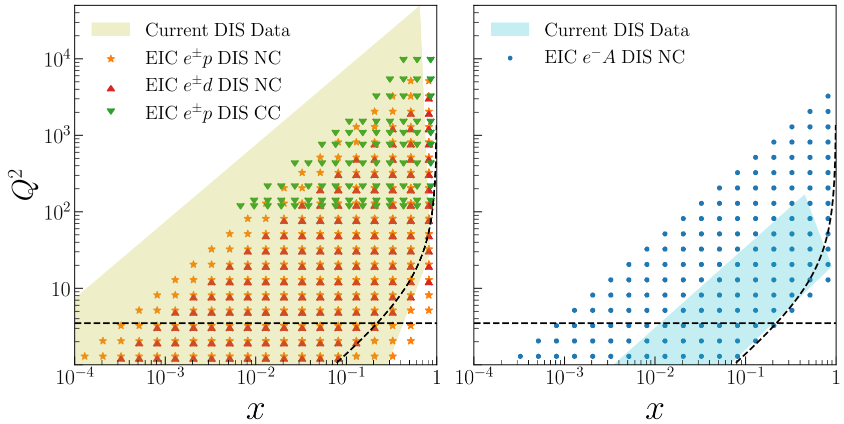
THE ERA OF THE EIC

THE IMPACT OF THE EIC: KINEMATIC COVERAGE

HL-LHC



EIC

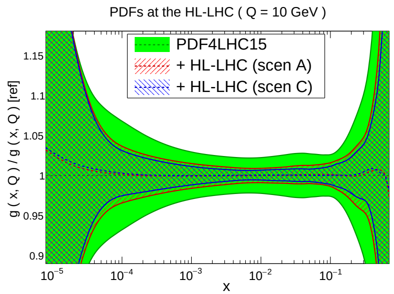


- COMPLEMENTARY KINEMATICS AND INFORMATION COMPARED TO HL-LHC
- LARGE x KINEMATICS, POLARIZATION, NUCLEAR

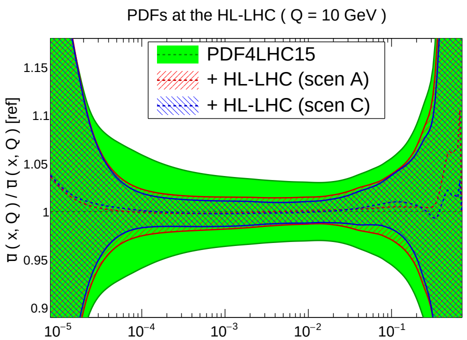
PDFS

HL-LHC

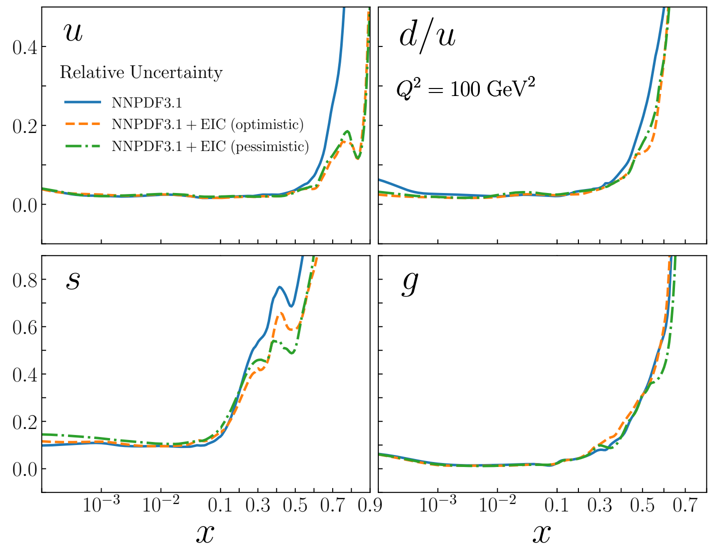
gluon



antiup



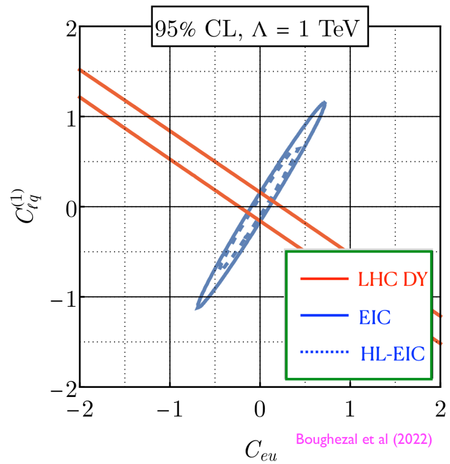
EIC



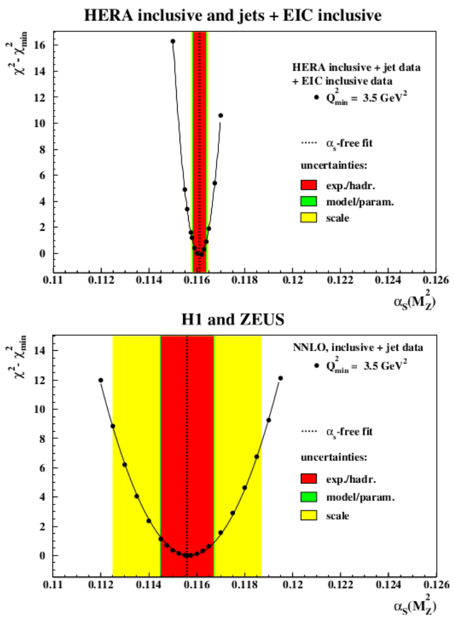
THE IMPACT OF THE EIC

- SMEFT FITS: RESOLVING DEGENERACIES (Boughezal et al., 2021-2023)
- α_s DETERMINATION AND PDFs (Cerci et al., 2023)
- LARGE x PDFs AND HIGH-MASS STATES (NNPDF, 2022)

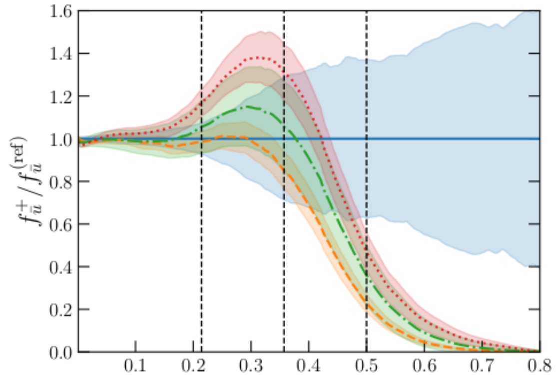
SMEFT: $\bar{e}\gamma^\mu e\bar{u}\gamma_\mu u$, $\bar{l}\gamma^\mu l\bar{q}\gamma_\mu q$ plane



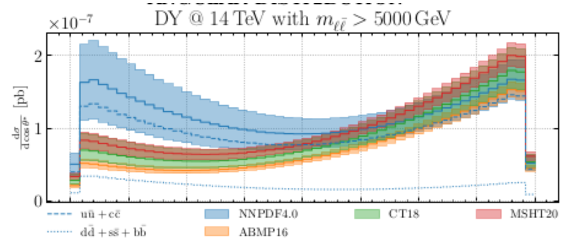
α_s: hera vs. eic



large x PDFS: antiup



DY forward-backward asym



EIC AND MACHINE LEARNING

Artificial Intelligence for the Electron Ion Collider (AI4EIC)

C. Allaire[61‖], R. Ammendola[24‖], E.-C. Aschenauer[3‖], M. Balandat[35‖], M. Battaglieri[38§], J. Bernauer[6,45§], M. Bondì[37‖], N. Branson[34,15‖], T. Britton[29‖], A. Butter[30‖], I. Chahrour[56], P. Chatagnon[29], E. Cisbani[39§], E. W. Cline[48], S. Dash[25§], C. Dean[33‖], W. Deconinck[55§], A. Deshpande[6,4], M. Diefenthaler[29§‖], R. Ent[29‖], C. Fanelli[64,29†††‡*], M. Finger[10], M. Finger, Jr.[10], E. Fol[5‖], S. Furletov[29‖], Y. Gao[3], J. Giroux[64,57‖††], N. C. Gunawardhana Waduge[59], R. Harish[9,12], O. Hassan[55,58], P. L. Hegde[9], R. J. Hernández-Pinto[17], A. Hiller Blin[27‖], T. Horn[49‡], J. Huang[3‖], D. Jayakodige[22,29], B. Joo[41‖], M. Junaid[57], P. Karande[32], B. Kriesten[8], R. Kunnawalkam Elayavalli[62‖], M. Lin[3], F. Liu[41‖], S. Liuti[59‖], G. Matousek[16], M. McEneaney[16‖], D. McSpadden[29††], T. Menzo[53‖], T. Miceli[18‖], V. Mikuni[31‖], R. Montgomery[54§], B. Nachman[31§‖], R. R. Nair[36], J. Niestroy[64], S. A. Ochoa Oregon[17], J. Oleniacz[63], J. D. Osborn[3§], C. Paudel[19], C. Pecar[16‖], C. Peng[1‖], G. N. Perdue[18§], W. Phelps[11,29‖], M. L. Purschke[3], K. Rajput[29‖††], Y. Ren[31§‖], D. F. Renteria-Estrada[17], D. Richford[2], B. J. Roy[40,23], D. Roy[47], N. Sato[29‖], T. Satogata[29,42‖], G. Sborlini[13,21], M. Schram[29§], D. Shih[46‖], J. Singh[44], R. Singh[4,7], A. Siodmok[28], P. Stone[64], J. Stevens[64§], L. Suarez[64], K. Suresh[57††], A.-N. Tawfik[20], F. Torales Acosta[31‖], N. Tran[18‖], R. Trotta[49], F. J. Twagirayezu[52], R. Tyson[54], S. Volkova[43‖], A. Vossen[29,16§], E. Walter[64††], D. Whiteson[51‖], M. Williams[33‖], S. Wu[55] and N. Zachariou[60] and P. Zurita[14,26§]

- PDFs: COLLINEAR AND GPDs
- MONTE CARLO EVENT GENERATORS
- DETECTOR SIMULATION
- CROSS-SECTION INFERENCE
- EVENT RECONSTRUCTION AND PARTICLE IDENTIFICATION
- HARDWARE ACCELERATION
- STREAMING READOUT DATA AQUISITION

CONCLUSION

NO EFFECT THAT REQUIRES MORE THAN 10% ACCURACY IN MEASUREMENT IS WORTH INVESTIGATING

Walther Nernst

~~NO EFFECT THAT REQUIRES MORE THAN 10% ACCURACY IN MEASUREMENT IS WORTH INVESTIGATING~~

Walther Nernst

ACCURACY OF OBSERVATION IS THE EQUIVALENT OF ACCURACY OF THINKING

Wallace Stevens