CAN WE TRUST MACHINE LEARNING?

PDFS AS A CASE STUDY

Stefano Forte

Università di Milano & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

Istituto Nazionale di Fisica Nucleare

precision QCD for the EIC

Stony Brook, September 19, 2023

THE PROBLEM
GENERALIZATION

WIKIPEDIA
The Free Encyclopedia

Search Wikipedia

Machine learning

Article Talk

From Wikipedia, the free encyclopedia

Generalization [edit]

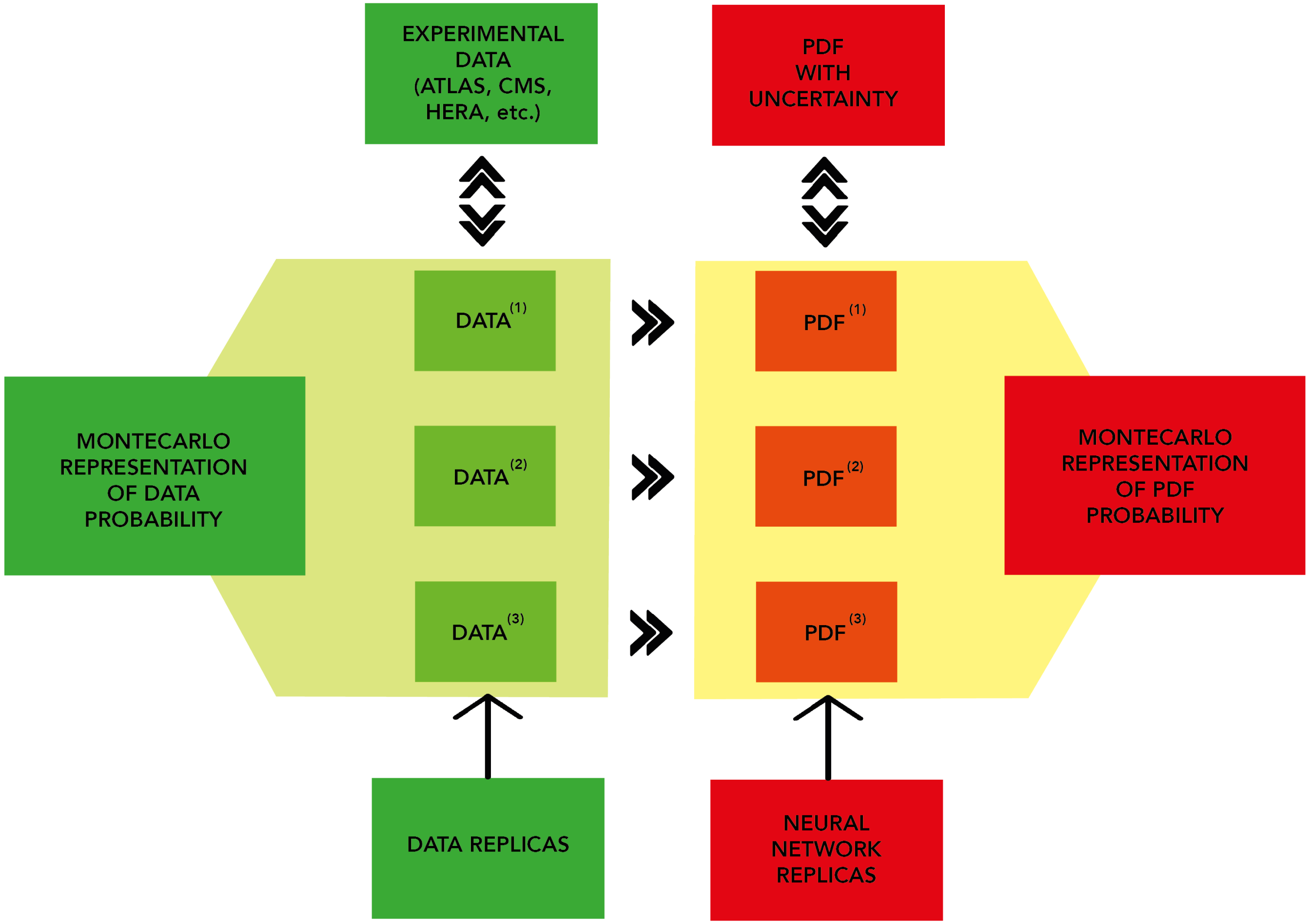
The difference between optimization and machine learning arises from the goal of generalization: while optimization algorithms can minimize the loss on a training set, machine learning is concerned with minimizing the loss on unseen samples. Characterizing the generalization of various learning algorithms is an active topic of current research, especially for deep learning algorithms.

- CAN WE TEST IT? $\Rightarrow$ VALIDATION
- CAN WE UNDERSTAND IT? $\Rightarrow$ EXPLANATION (XML)

PDF/NNPDF RECAP SEQUENCE

THE FUNCTIONAL MONTE CARLO

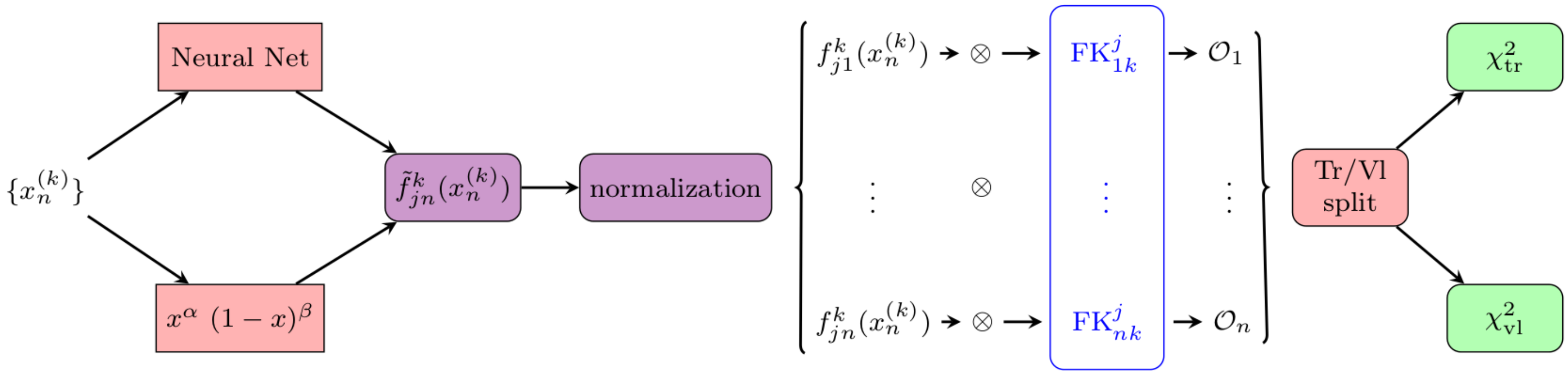
REPLICA SAMPLE OF FUNCTIONS $\Leftrightarrow$ PROBABILITY DENSITY IN FUNCTION SPACE
KNOWLEDGE OF LIKELIHHOD SHAPE (FUNCTIONAL FORM) NOT NECESSARY

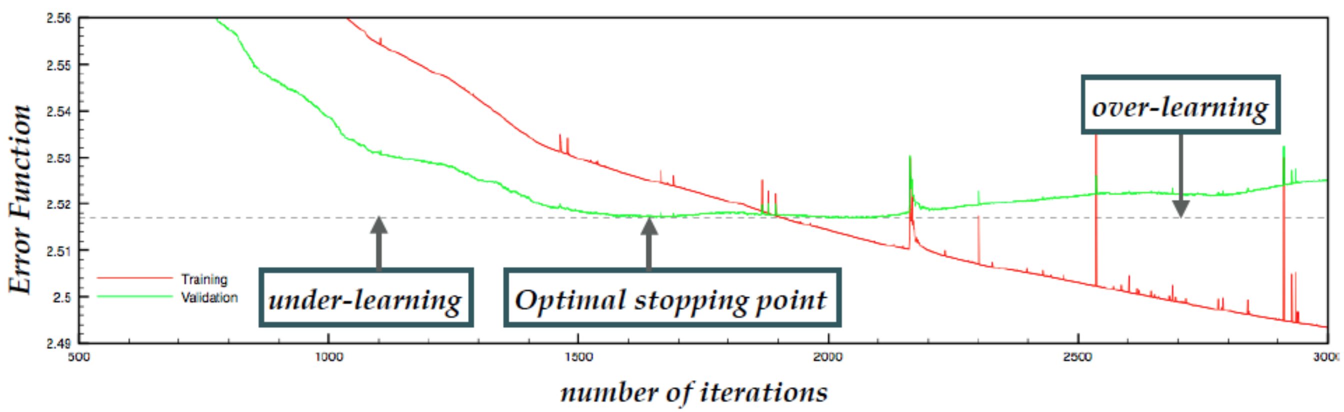


FINAL PDF SET: $f_i^{(a)}(x,\mu)$;
i =up, antiup, down, antidown, strange, antistrange, charm, gluon; $j = 1, 2, \ldots N_{\rm rep}$

MINIMIZATION AND CROSS-VALIDATION

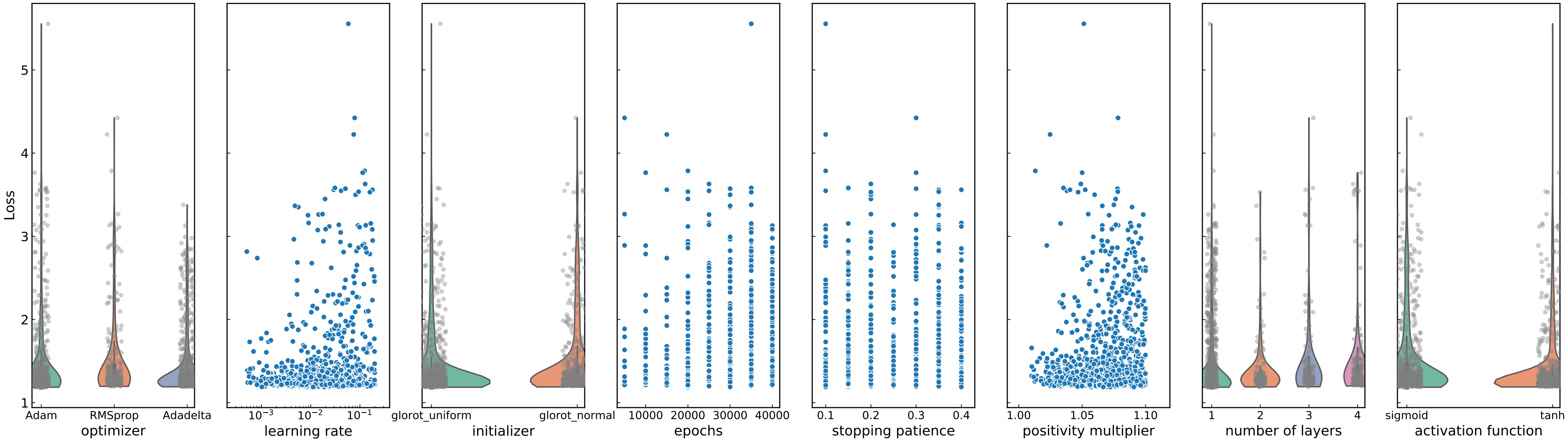
- NEURAL NET PARAMETERS DETERMINED BY χ^2 MINIMIZATION THROUGH GRADIENT DESCENT
- RANDOM TRAINING-VALIDATION SPLIT, χ^2 TO TRAINING DATA REPLICAS MINIMIZED
- TRAINING STOPS IF VALIDATION χ^2 GROWS FOR A WHILE (PATIENCE)
- LOWEST VALIDATION χ^2 $\Rightarrow$ OPTIMAL FIT





FITTING THE METHODOLOGY

HYPEROPTIMIZATION

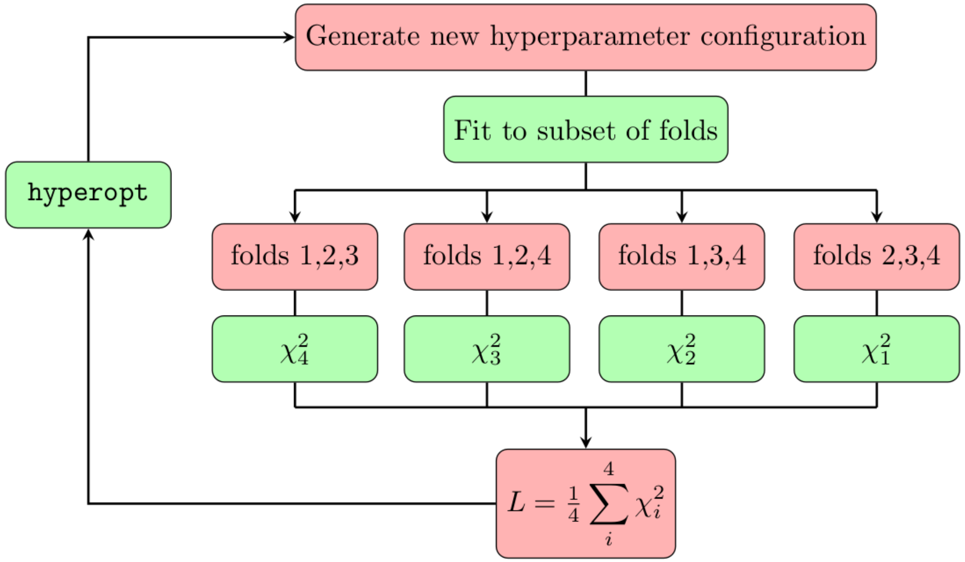


HYPEROPT PARAMETERS

Neural Network	Fit options
Number of layers (*)	Optimizer (*)
Size of each layer	Initial learning rate (*)
Dropout	Maximum number of epochs (*)
Activation functions (*)	Stopping Patience (*)
Initialization functions (*)	Positivity multiplier (*)

- SCAN PARAMETER SPACE
- OPTIMIZE FIGURE OF MERIT: K-FOLDING LOSS

K-FOLDING



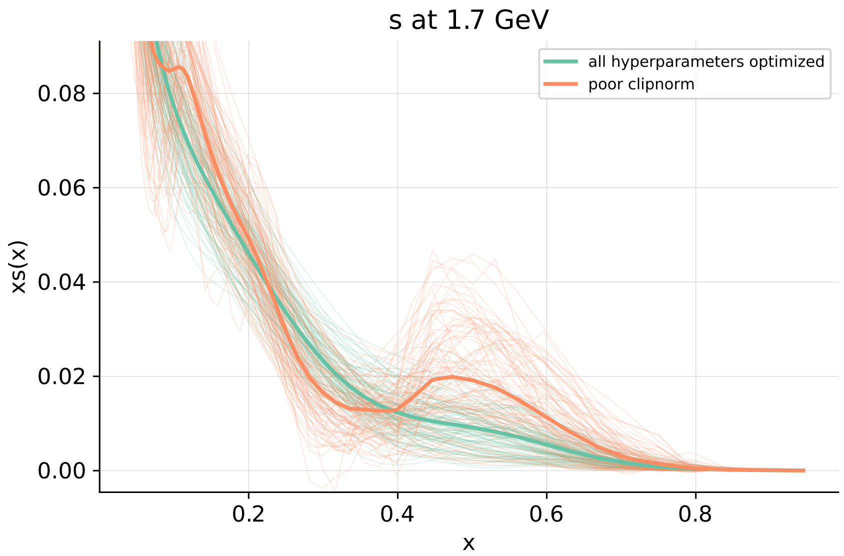
Fold 1		
CHORUS σ_{CC}^{ν}	HERA I+II inc NC e^+p 920 GeV	BCDMS p
LHCb Z 940 pb	ATLAS W, Z 7 TeV 2010	CMS Z p_T 8 TeV (p_T^{ll}, y_{ll})
DY E605 $\sigma_{\rm DY}^{p}$	CMS Drell-Yan 2D 7 TeV 2011	CMS 3D dijets 8 TeV
ATLAS single-$\bar{t}$ y (normalised)	ATLAS single top R_t 7 TeV	CMS $t\bar{t}$ rapidity $y_{t\bar{t}}$
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. $\sigma_{b\bar{b}}^{\rm red}$
NMC p	NuTeV $\sigma_c^{\bar{\nu}}$	LHCb $Z \to ee$ 2 fb
CMS W asymmetry 840 pb	ATLAS Z p_T 8 TeV (p_T^{ll}, M_{ll})	D0 $W \to \mu\nu$ asymmetry
DY E886 $\sigma_{\rm DY}^{p}$	ATLAS direct photon 13 TeV	ATLAS dijets 7 TeV, R=0.6
ATLAS single antitop y (normalised)	CMS $\sigma_{tt}^{\rm tot}$	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_c^{ν}	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 $\sigma_{\rm DY}^{d}/\sigma_{\rm DY}^{p}$	CDF Z rapidity (new)
ATLAS $\sigma_{tt}^{\rm tot}$	ATLAS single top y_t (normalised)	CMS $\sigma_{tt}^{\rm tot}$ 5 TeV
CMS $t\bar{t}$ double diff. $(m_{t\bar{t}}, y_t)$		

Fold 4		
CHORUS $\sigma_{CC}^{\bar{\nu}}$	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_T^{ll}, 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	

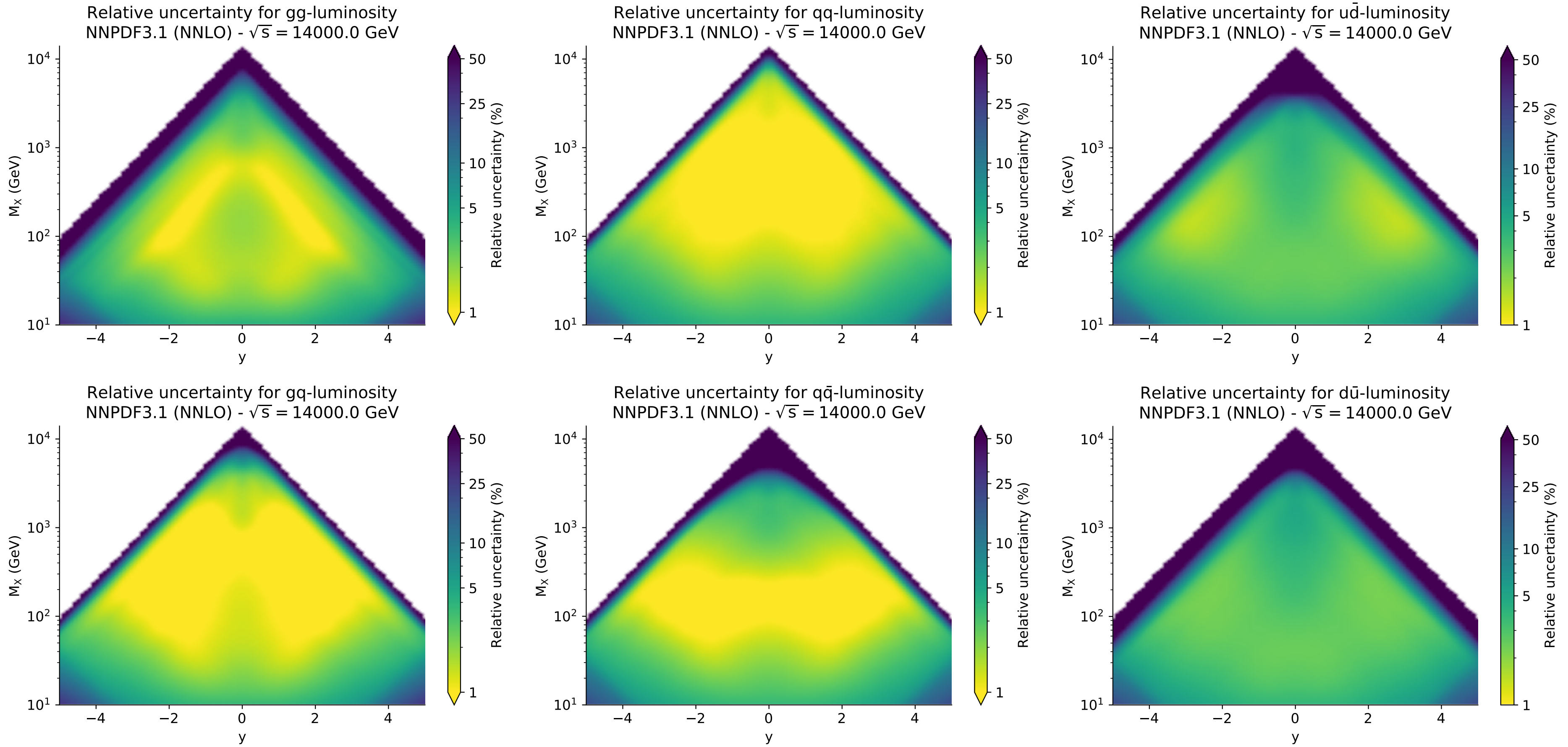
K-FOLDING VS NO K-FOLDING



- EACH FOLD REPRODUCES FEATURES OF FULL DATASET
- LOSS: AVERAGE χ^2 OF NON-FITTED FOLDS
- OVERFITTING REMOVED $\Rightarrow$ CORRECT GENERALIZATION

UNCERTAINTIES 2016

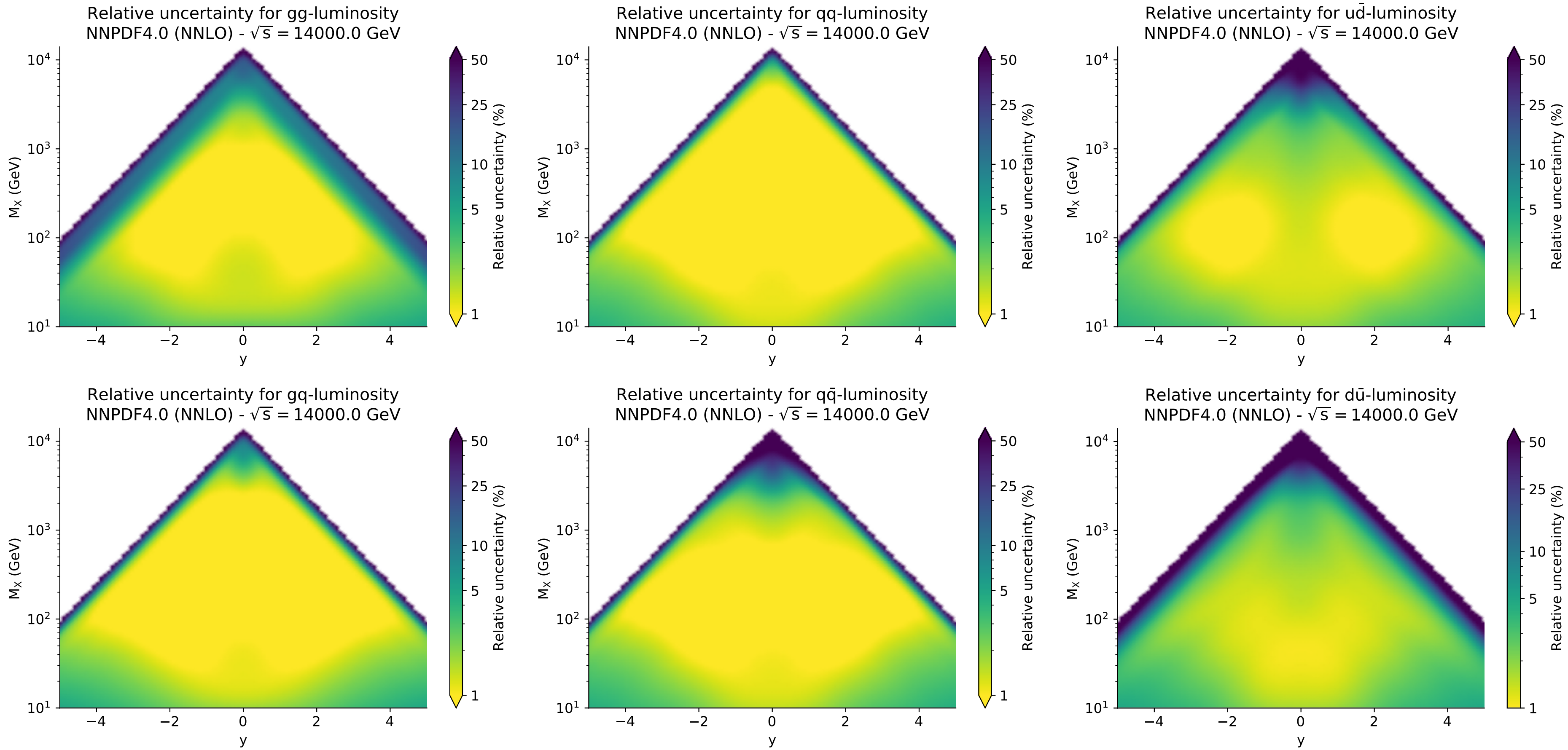
GLUON | SINGLET | FLAVORS



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

UNCERTAINTIES 2022

GLUON SINGLET FLAVORS



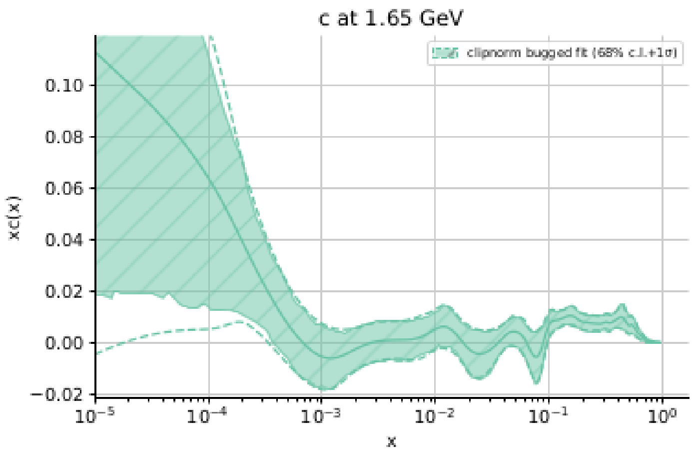
- TYPICAL UNCERTAINTIES IN DATA REGION: SINGLET $\sim 1\%$, NONSINGLET $\sim 2-3\%$
- DATA REGION: $10 \lesssim M_X \lesssim 3 \cdot 10^3$ TeV, $-4 \lesssim y \lesssim 4$

ONLINE VALIDATION: OVERFITTING METRIC

- RECOMPUTE VALIDATION $\chi^2_{\text{val}'}$ FOR ALL DATA REPLICAS
  - KEEPING SAME TRAINING-VALIDATION SPLIT
  - BUT DIFFERENT FLUCTUATED VALIDATION DATA
- COMPUTE AVERAGE OVER REPLICAS $\langle \chi^2_{\text{val}'} \rangle$ &
  DETERMINE DIFFERENCE TO STANDARD VALIDATION χ^2_{val}
  OVERFITNESS: $\mathcal{R}_O = \chi^2_{\text{val}} - \langle \chi^2_{\text{val}'} \rangle$
- NEGATIVE OVERFITNESS $\mathcal{R}_O \Rightarrow$ OVERFIT

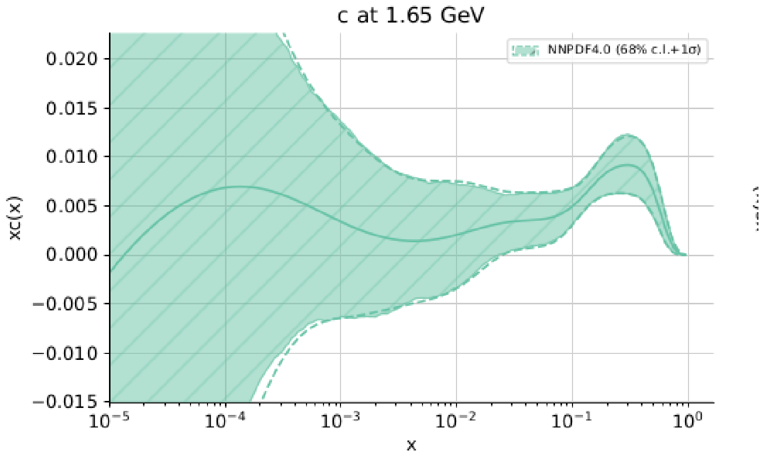
CHARM PDF

OVERFIT (NO CLIPNORM)



$$\mathcal{R}_O = -0.024 \pm 0.012$$

PROPER FIT (NNPDF4.0)



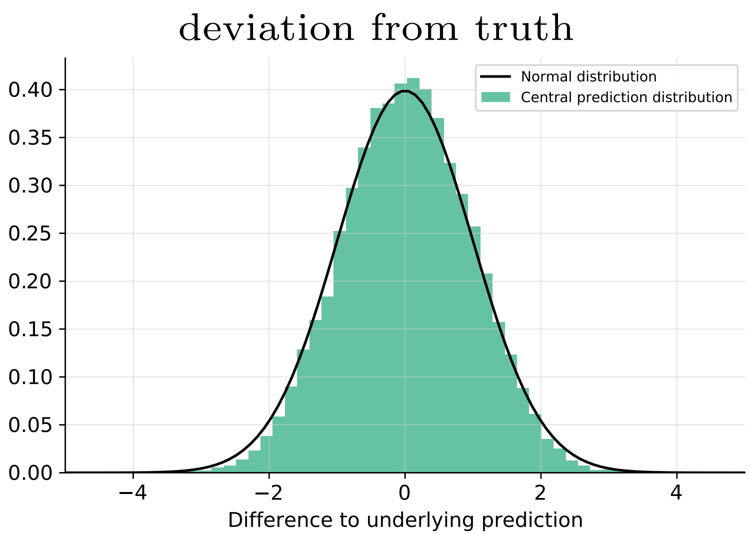
$$\mathcal{R}_O = -0.001 \pm 0.013$$

OFFLINE VALIDATION: CLOSURE TESTS

FAITHFUL UNCERTAINTIES IN DATA REGION?

- ASSUME "TRUE" UNDERLYING PDF $\Rightarrow$ E.G. SOME RANDOM PDF REPLICA
- GENERATE DATA DISTRIBUTED ACCORDING TO EXPERIMENTAL COVARIANCE MATRIX
- RUN WHOLE METHDOLOGY ON THESE DATA
- DO STATISTICS ON "RUNS OF THE UNIVERSE", POSSIBLE THANKS TO EFFICIENT METHDOLOGY: COMPARE TO TRUE VALUES OF OBSERVABLES (NOT FITTED)
  - BIAS/VARIANCE: MEAN SQUARE DEVIATION WR TO TRUTH VS UNCERTAINTY
  - IS TRUTH WITHIN ONE SIGMA 68% OF TIMES?

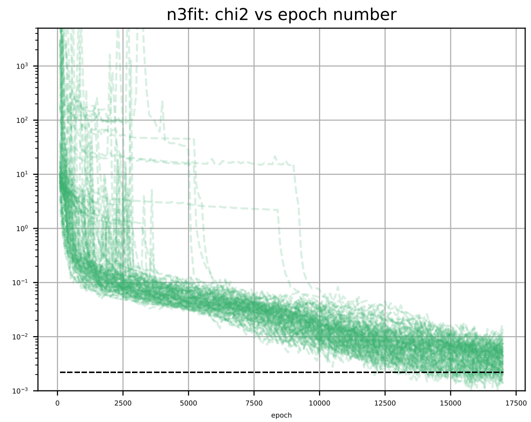
RESULTS



Dataset	$\sqrt{\text{bias/variance}}$	$\xi_{1\sigma}^{(\text{data})}$
DY	0.99 ± 0.08	0.69 ± 0.02
Top-pair	0.75 ± 0.06	0.75 ± 0.03
Jets	1.14 ± 0.05	0.63 ± 0.03
Dijets	0.99 ± 0.07	0.70 ± 0.03
Direct photon	0.71 ± 0.06	0.81 ± 0.03
Single top	0.87 ± 0.07	0.69 ± 0.04
Total	1.03 ± 0.05	0.68 ± 0.02

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

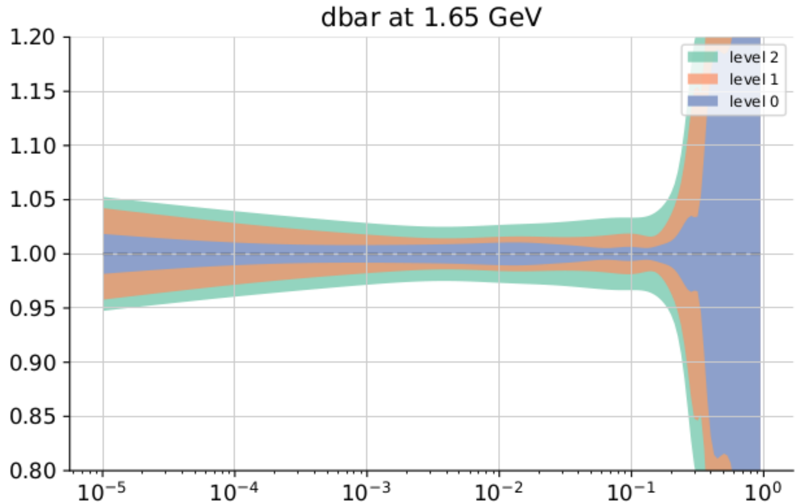
LEVEL 0 χ^2 VS TRAINING



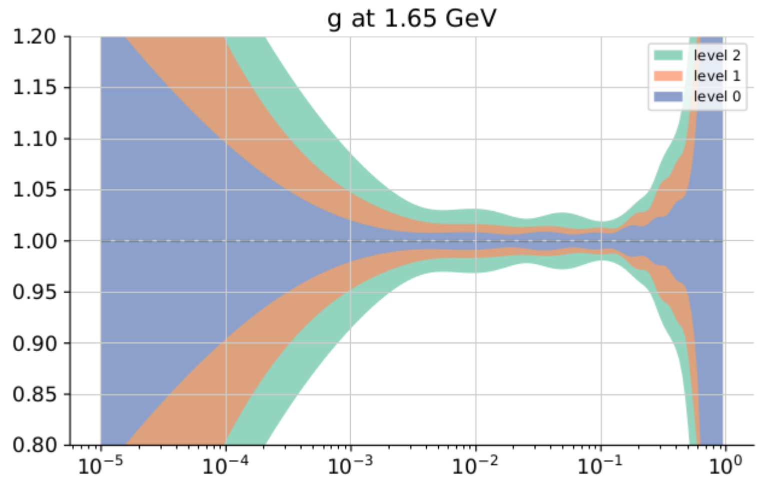
- LEVEL 0 (TRUTH DATA) $\Rightarrow$ $\chi^2 \approx 0$, YET UNCERTAINTY NONZERO
  $\Rightarrow$ NEURAL NETS $\Leftrightarrow$ MANY FUNCTIONAL FORMS
- LEVEL 1 (RUNS OF UNIVERSE) $\Rightarrow$ REPLICAS ALL FITTED TO SAME DATA,
  YET UNCERTAINTY NONZERO
  $\Rightarrow$ DITTO
- LEVEL 0, 1 AND 2 UNCERTAINTIES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

LEVEL 0/1/2 UNCERTAINTIES

ANTIDOWN



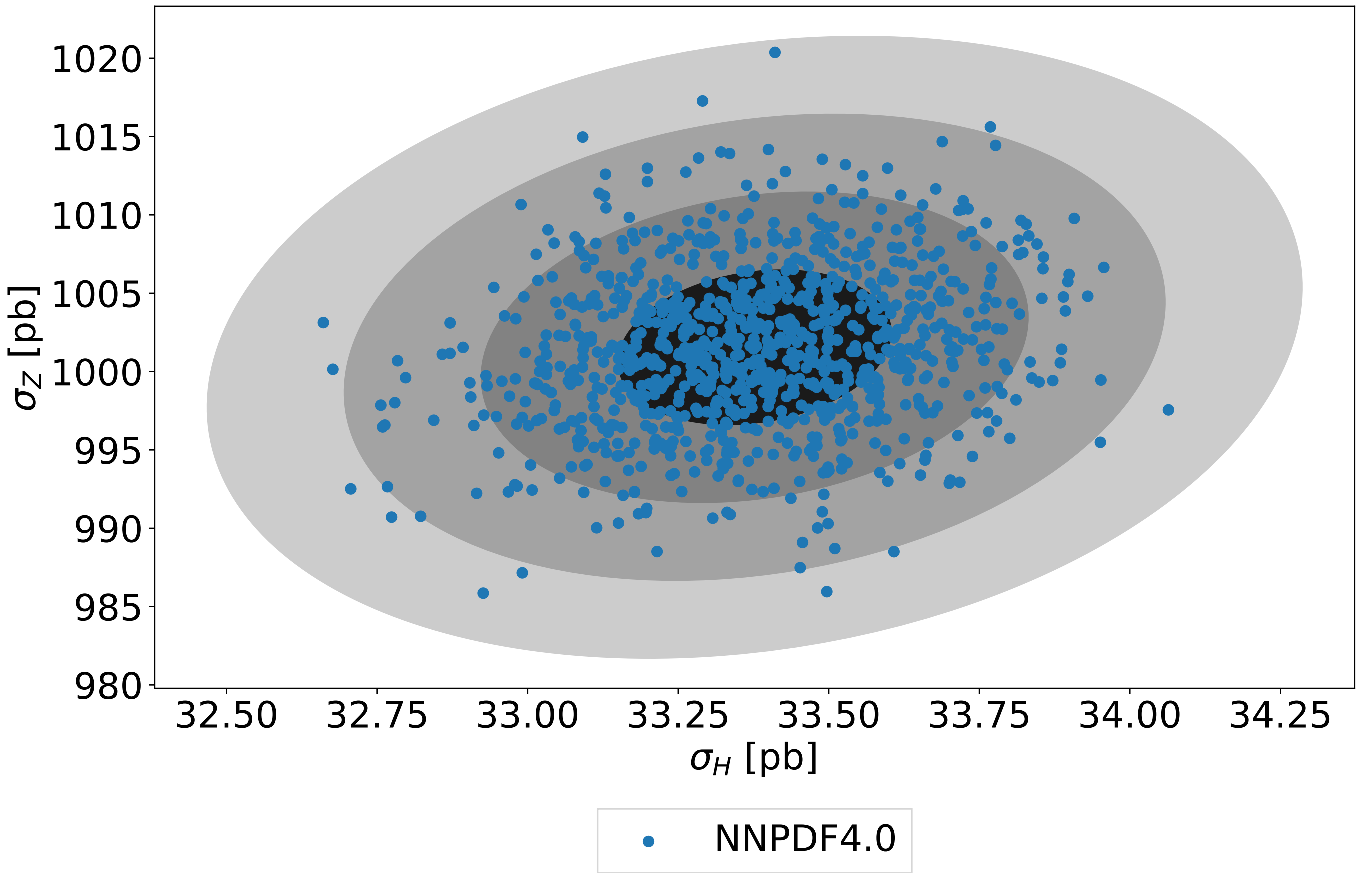
GLUON



THE MC DISTRIBUTION: UNCERTAINTIES

HOW DOES IT LOOK LIKE?

- PLOT RESULTS IN (σ_H, σ_Z) PREDICTION SPACE $\Rightarrow$ GAUSSIAN!
- DISTRIBUTION OF REPLICAS $\Rightarrow$ OPTIMAL IMPORTANCE SAMPLING



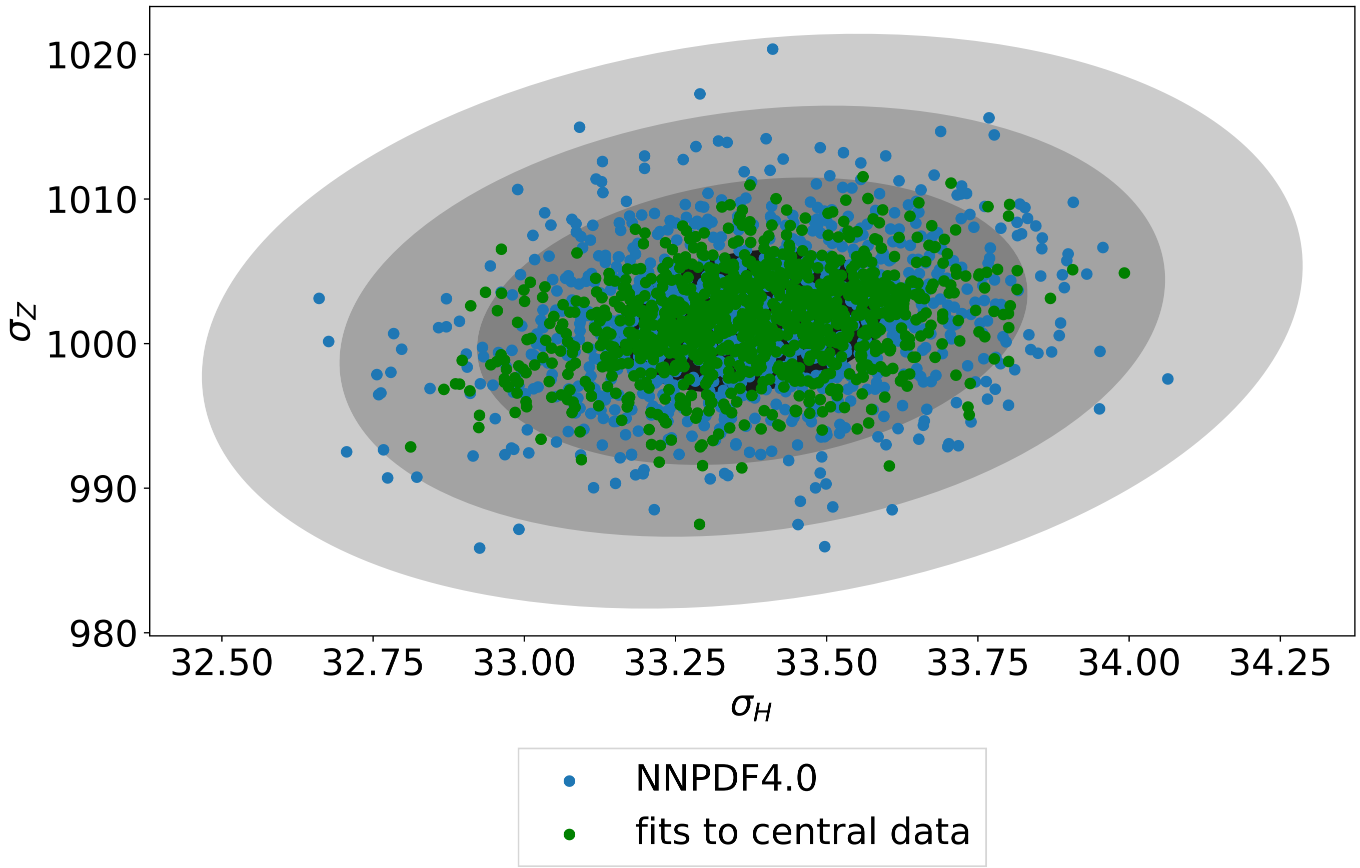
DISTRIBUTION OF REPLICAS DRIVEN BY

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

EXPLANATION

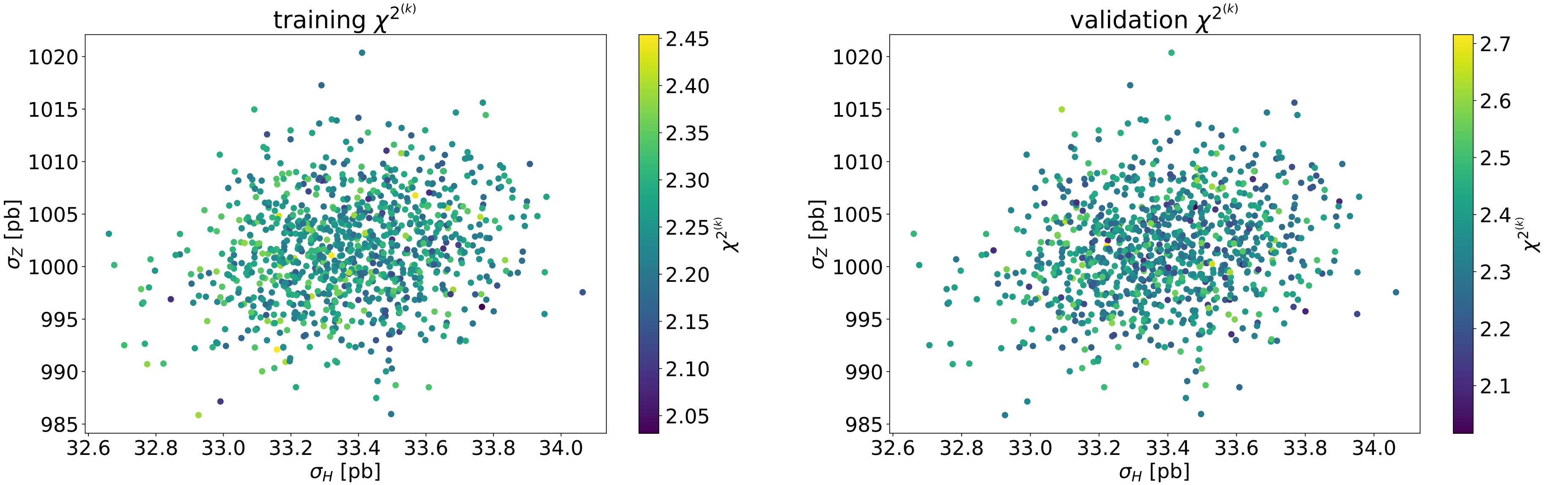
THE REPLICA DISTRIBUTION

- REPLICA FLUCTUATION $\Rightarrow$ DATA UNCERTAINTIES
- NO REPLICA FLUCTUATION $\Rightarrow$ FIT DEGENERACY



THE REPLICA DISTRIBUTION

ARE ALL FITS EQUALLY GOOD?



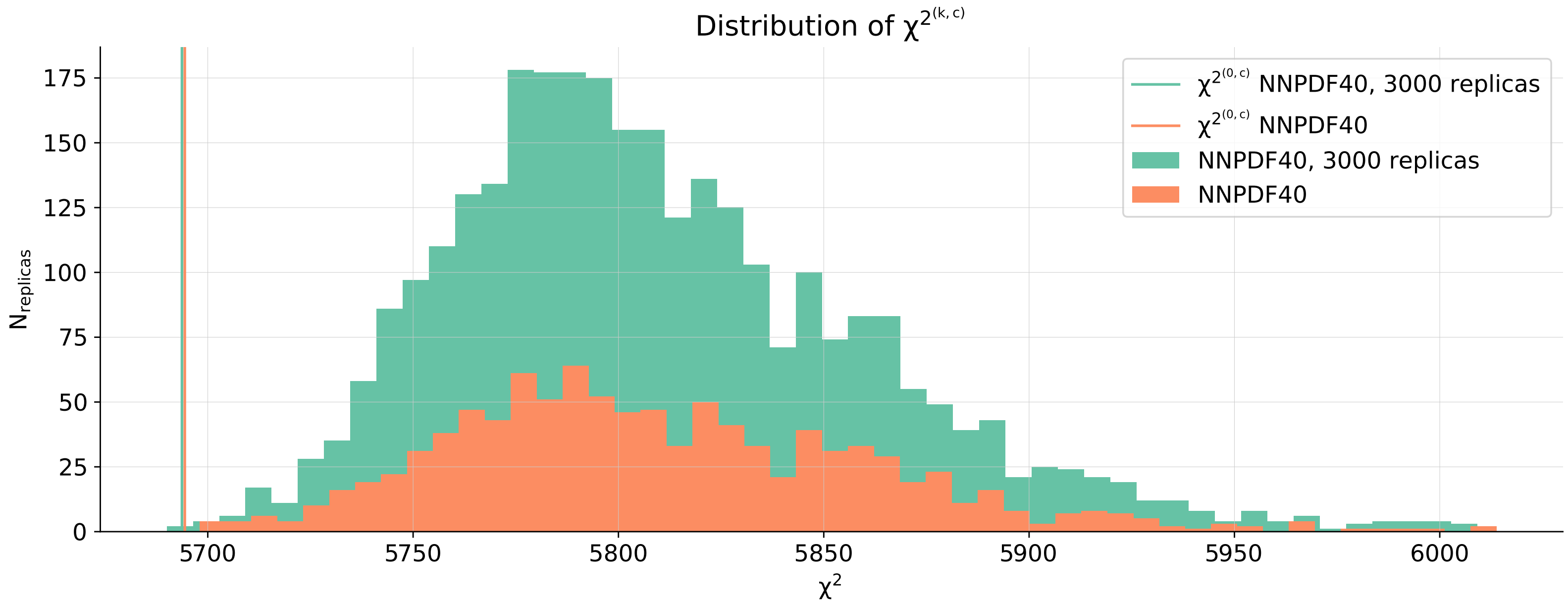
- COMPARE TRAINING AND VALIDATION χ^2 FOR EACH REPLICA
- NO CORRELATION BETWEEN FIT QUALITY AND POSITION IN THE (σ_H, σ_Z) PLANE
- UNIFORM FIT QUALITY

METRICS IN REPLICA SPACE

THE REPLICA DISTRIBUTION
COMPARISON TO CENTRAL DATA

- EACH PDF REPLICA FITTED TO A DATA REPLICA
- FIT QUALITY TO CENTRAL DATA STATISTICALLY DISTRIBUTED

1000 REPLICAS VS. 3000 REPLICAS

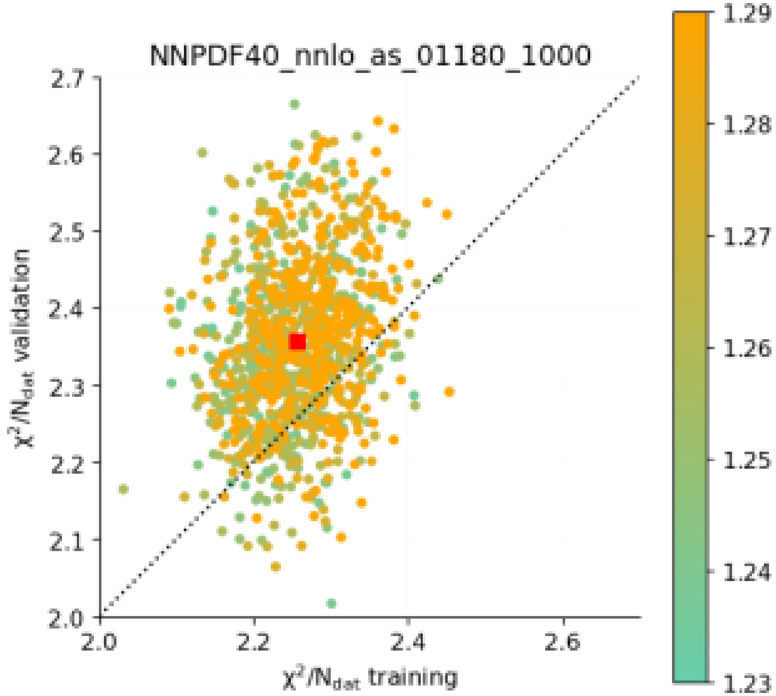


- AVERAGE BEST FIT PDF $\Rightarrow$ LOW χ^2
- NOT NECESSARILY LOWEST

THE REPLICA DISTRIBUTION

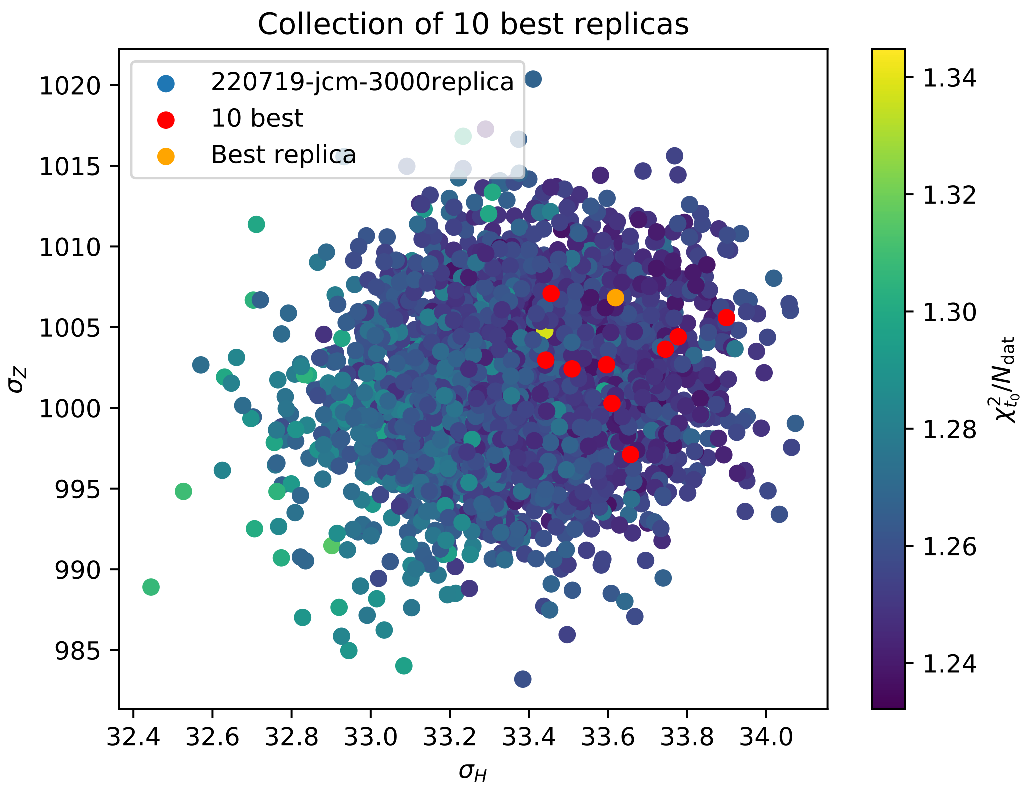
COMPARISON TO CENTRAL DATA

- ARE FITS WITH HIGH χ^2 TO CENTRAL DATA POOR (UNDERLEARNT)?



- NO CORRELATION BETWEEN χ^2 TO CENTRAL DATA AND TRAINING, VALIDATION χ^2
- UNIFORM FIT QUALITY
- DISPERSION DUE
  - DATA REPLICA FLUCTUATION $\Rightarrow$ DATA UNCERTAINTIES
  - BEST FIT DEGENERACY
    $\Rightarrow$ INTERPOLATION, EXTRAPOLATION AND FUNCTIONAL UNCERTAINTIES

REPLICA LOSS DISTRIBUTION
CORRELATION TO FEATURES



χ^2 TO CENTRAL DATA

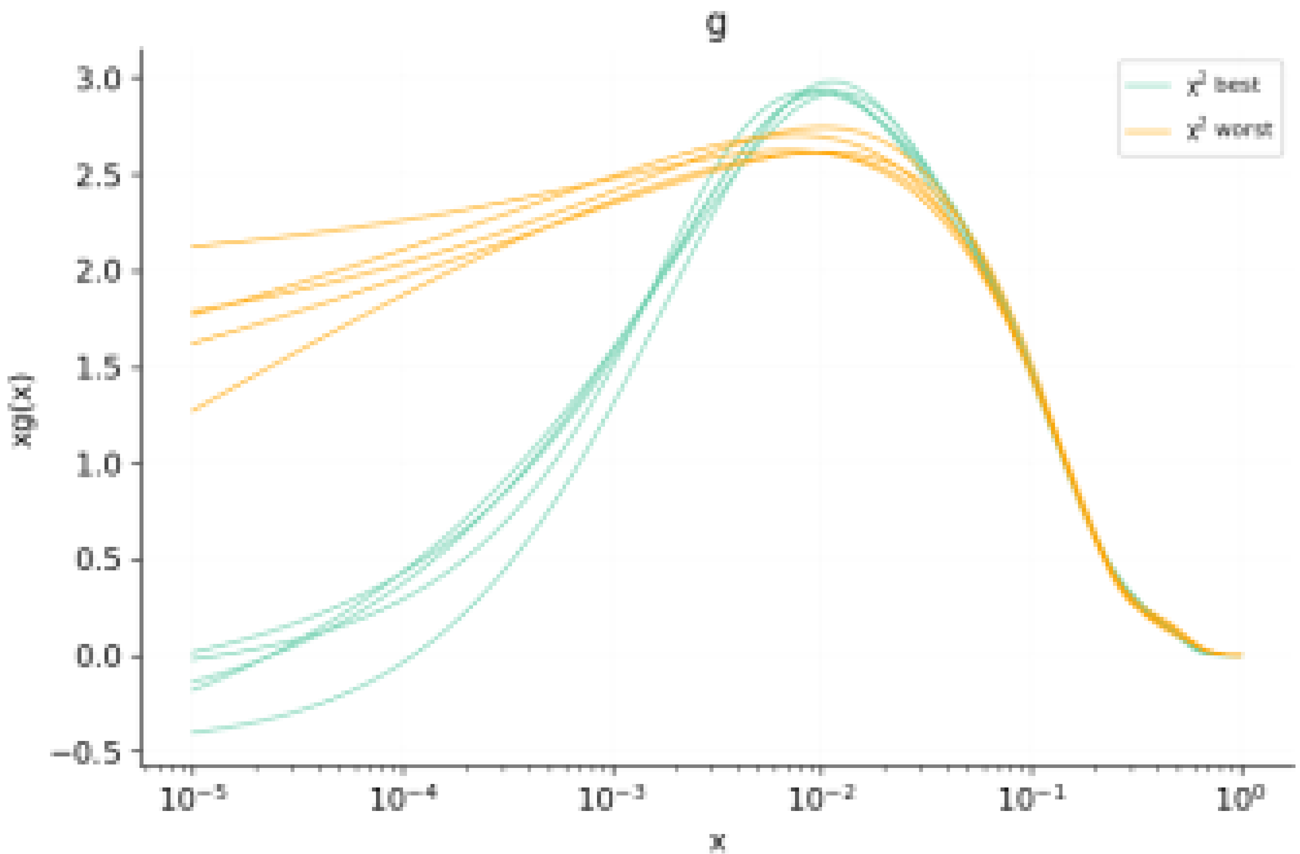
- CORRELATED TO POSITION IN (σ_H, σ_z) PLANE
- CORRELATED TO A FEATURE?

EXPLANATION

LOOKING FOR FEATURES

REPLICAS WITH LOWEST & HIGHEST χ^2 TO CENTRAL DATA

THE GLUON



- REPLICAS CLOSER TO CENTRAL DATA $\Rightarrow$ MORE STRUCTURE
- CORRELATED TO A FEATURE?

EXPLANATION

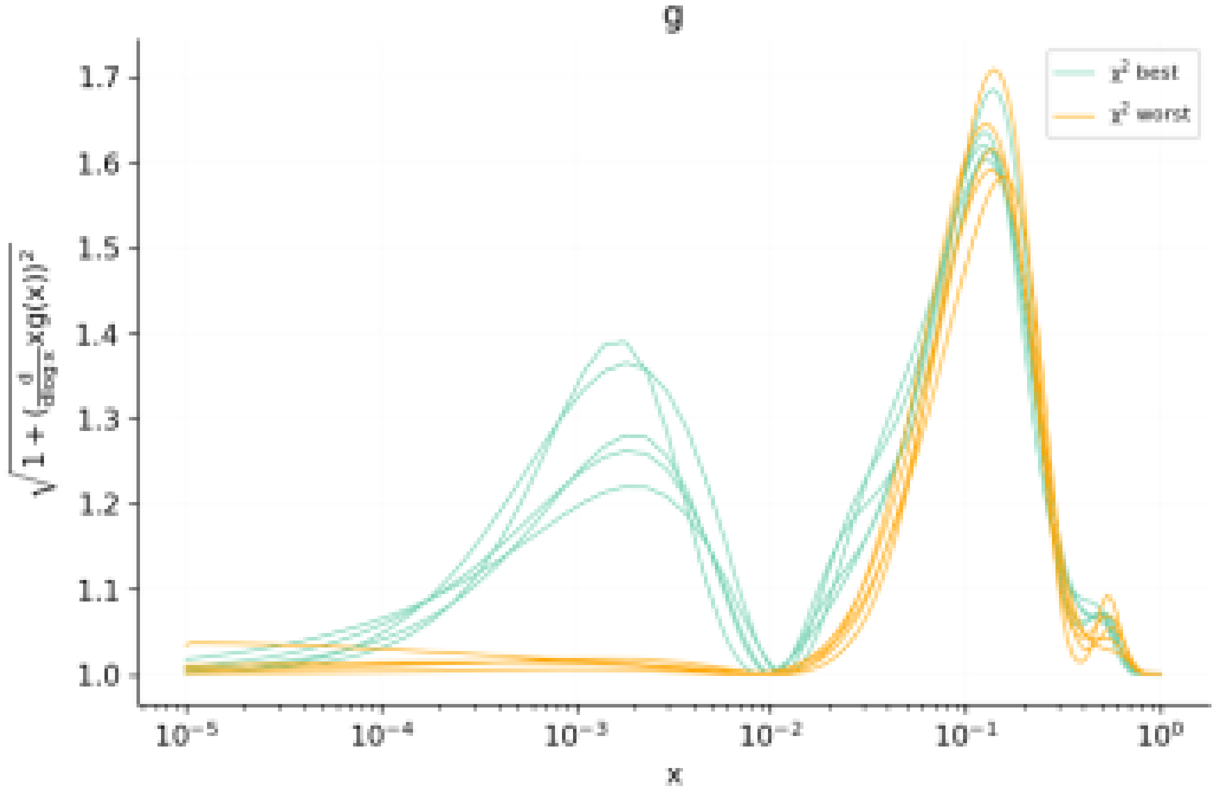
THE PDF KINETIC ENERGY

REPLICAS WITH LOWEST & HIGHEST χ^2 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

GENERALIZATION

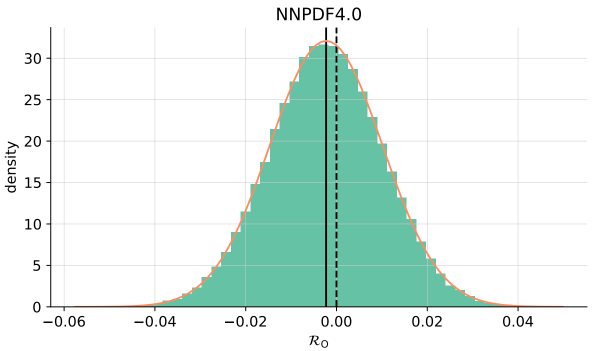
EXPLANATION

OVERLEARNING?

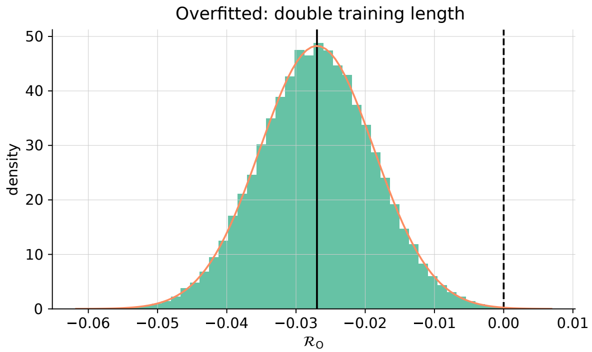
- INDUCE OVERLEARING: DOUBLE TRAINING LENGTH

THE OVERFIT METRIC

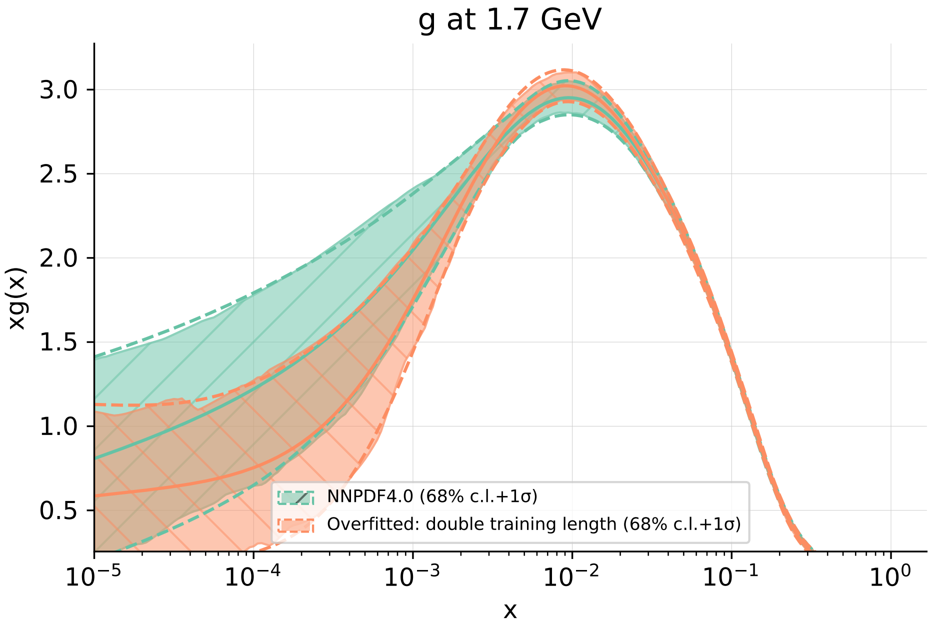
DEFAULT NNPDF4.0



OVERFIT VARIANT



THE GLUON



- LOOK AT THE OUTPUT ⇒ MORE STRUCTURE IN GLUON

EXPLANATION

A PARADOX?

- BEST FIT TO CENTRAL DATA CORRELATED TO HIGH ARCLENGTH
- HIGH ARCLENGTH CORRELATED TO OVERLEARNING
- TRAINING/VALIDATION LOSS

  UNCORRELATED TO QUALITY OF FIT TO CENTRAL DATA

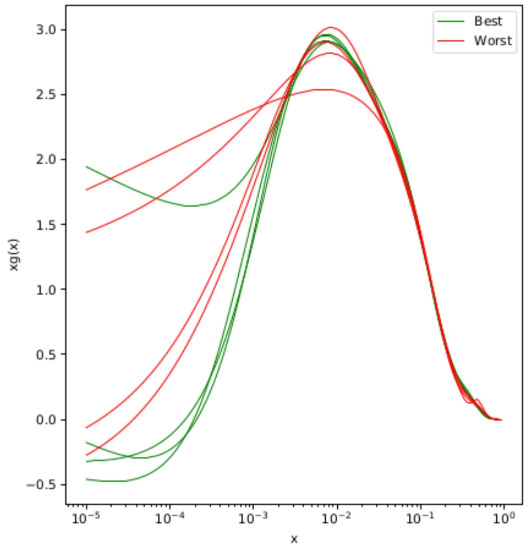
EXPLANATION

GENERALIZATION!

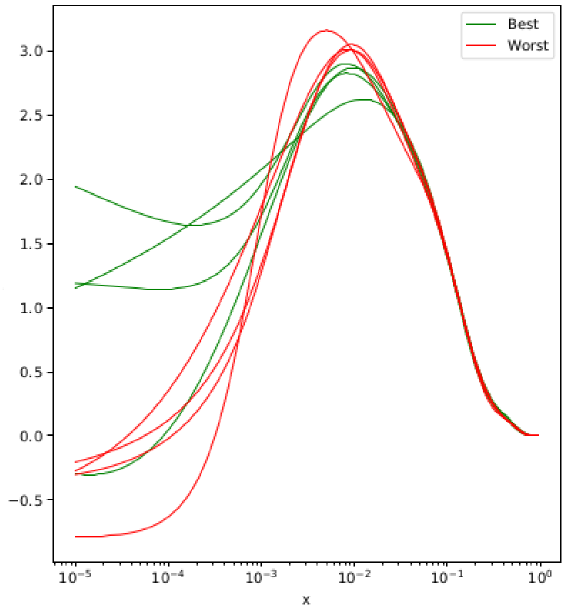
- OVERFITTING CAN MEAN POOR GENERALIZATION
- KEPT IN CHECK BY K-FOLDING (NOT CROSS-VALIDATION)
- LOOK AT BEST χ^2 TO FITTED VS. EXCLUDED FOLDS

THE GLUON

FITTED FOLDS



EXCLUDED FOLD



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

THE FUTURE

WHY ML?

- MULTIDIMENSIONAL REGRESSION
- UNCERTAINTY CONTROL
- EXPLAINABILITY
- GENERALIZATION

EIC AND MACHINE LEARNING

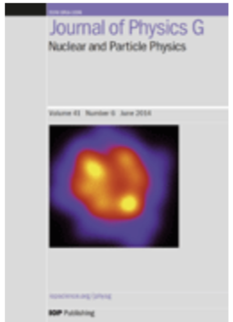
Artificial Intelligence for the Electron Ion Collider (AI4EIC)

93v1 [physics.acc-ph] 17 Jul 2023

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

Journal of Physics G: Nuclear and Particle Physics



ISSN: 1361-6471

SUPPORTS OPEN ACCESS

Journal of Physics G: Nuclear and Particle Physics publishes theoretical and experimental research in nuclear and particle physics including all interface areas between these fields. The journal also publishes articles on nuclear and particle astrophysics.

Submit an article Track my article

RSS Sign up for new issue notifications

Current volume
Number 10, October 2023 Go

Journal archive
Vol 50, 2023 Go

Focus issues
Focus on Neutrino-Nucleus Inter Go

9 days	**62 days**	**3.5**	**7.2**
Median submission to first decision before peer review	Median submission to first decision after peer review	Impact factor	Citescore

Full list of journal metrics

Most read
Latest articles
Review articles
Accepted manuscripts
Open Access

Open all abstracts

OPEN ACCESS

The exploration of hot and dense nuclear matter: introduction to relativistic heavy-ion physics

Hannah Elfner and Berndt Müller 2023 *J. Phys. G: Nucl. Part. Phys.* **50** 103001

+ Open abstract View article PDF

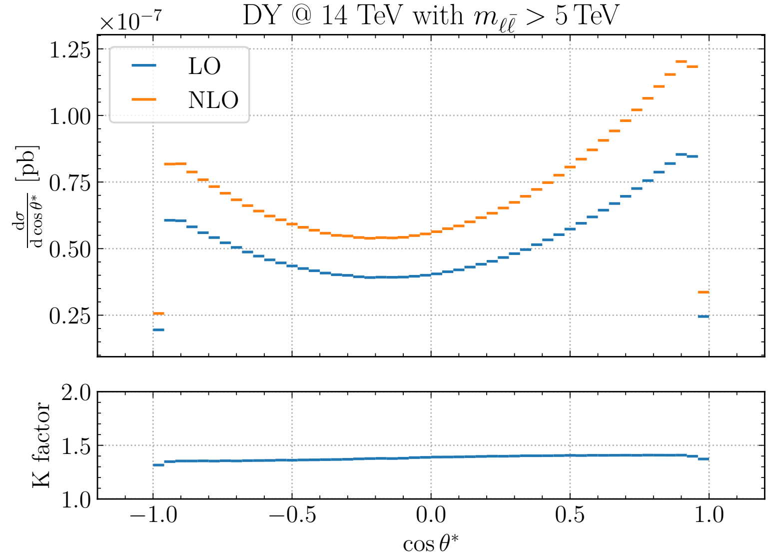
EXTRAS

QUALITATIVE BEHAVIOR: NLO VS. LO
LO INTEGRATED RESULT (HIGH MASS)

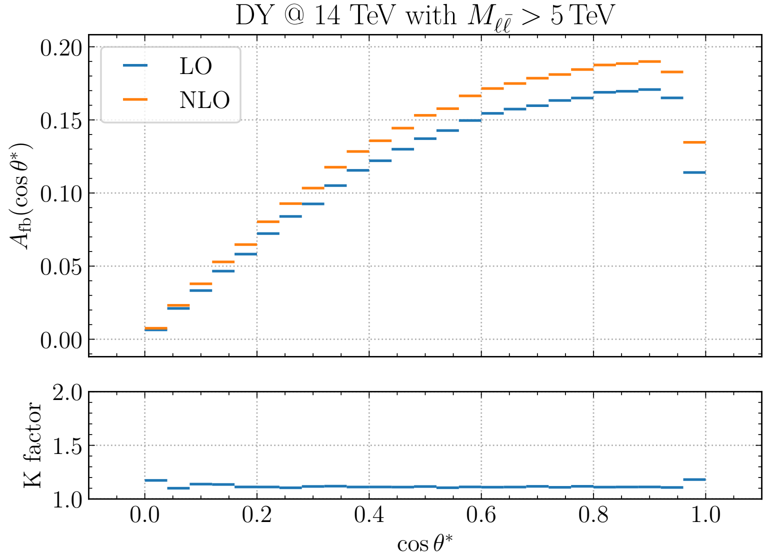
$$A_{\rm fb}(\cos\theta^*) = \frac{\cos\theta^*}{(1+\cos^2(\theta^*))}\,\frac{g_A}{g_{S,q'}}; \quad g_{A,S} \propto \int \frac{dm_{\ell\bar{\ell}}}{m_{\ell\bar{\ell}}}\,\frac{dx_1}{x_1}\,\mathcal{L}_{A,S}(m_{\ell\bar{\ell}}, x_1)$$

LO VS. NLO FOR $M_{z'} = 5$TeV

CROSS-SECTION



FORWARD-BACKWARD ASYMMETRY



- AT LO, $A_{\rm fb} \propto \cos\theta$, EFFECTIVE COUPLING DETERMINED BY PDF LUMINOSITY
- NLO K-FACTOR ALMOST θ-INDEPENDENT

QUALITATIVE BEHAVIOR:

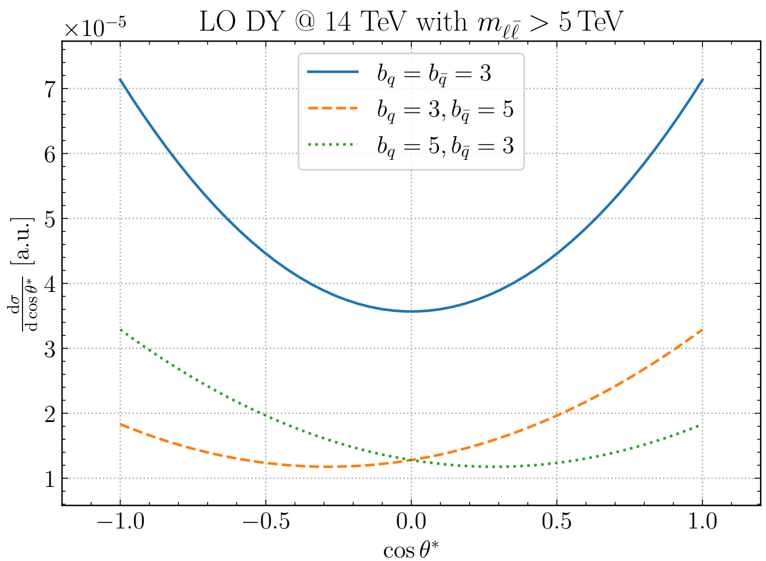
$$A_{\rm fb} \propto \mathcal{L}_{A,q} = f_q(x_1) f_{\bar q}(x_2,) - f_q(x_2) f_{\bar q}(x_1) = \frac{1}{2}\left(f_q^-(x_1) f_q^+(x_2) - f_q^-(x_2) f_q^+(x_1,)\right)$$

$$f_q^\pm = f_q \pm f_{\bar q};\ f^- \to \text{VALENCE};\ f^+ \to \text{SEA}$$

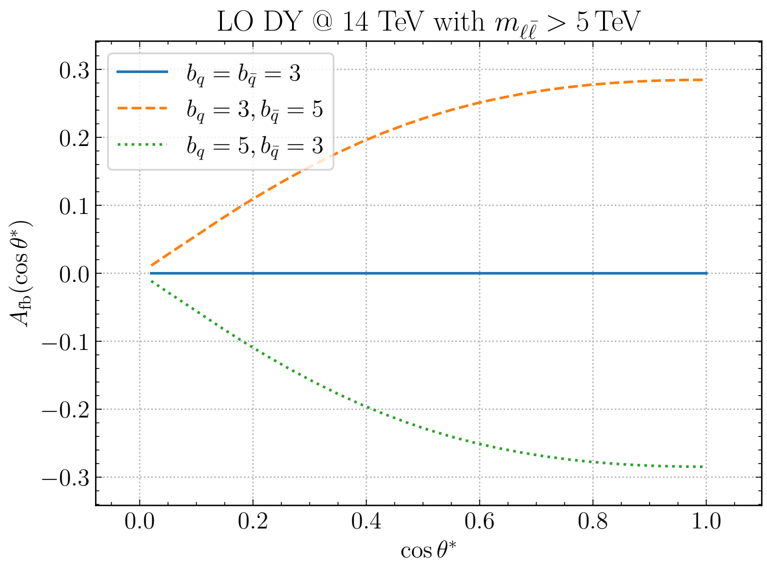
TOY PDFS

$$x f_q^\pm = x^{-1}\left[(1-x)^{b_q} \pm (1-x)^{b_{\bar q}}\right]$$

CROSS-SECTION



FORWARD-BACKWARD ASYMMETRY



- TOY: SIGN OF ASYM $\Leftrightarrow$ SIGN OF VALENCE
- GENERAL: SIGN OF ASYM $\Leftrightarrow$ DROP OF VALENCE VS. SEA

$$\text{sign}\left[\mathcal{L}_{A,q}\right] = \text{sign}\left[\frac{f_q^+(x_2)}{f_q^+(x_1)} - \frac{f_q^-(x_2)}{f_q^-(x_1)}\right] = \text{sign}\left[\frac{f_q(x_2)}{f_q(x_1)} - \frac{f_{\bar q}(x_2)}{f_{\bar q}(x_1)}\right],\ x_1 > x_2$$

- VALENCE DROPS FASTER $\Rightarrow$ NEGATIVE ASYM

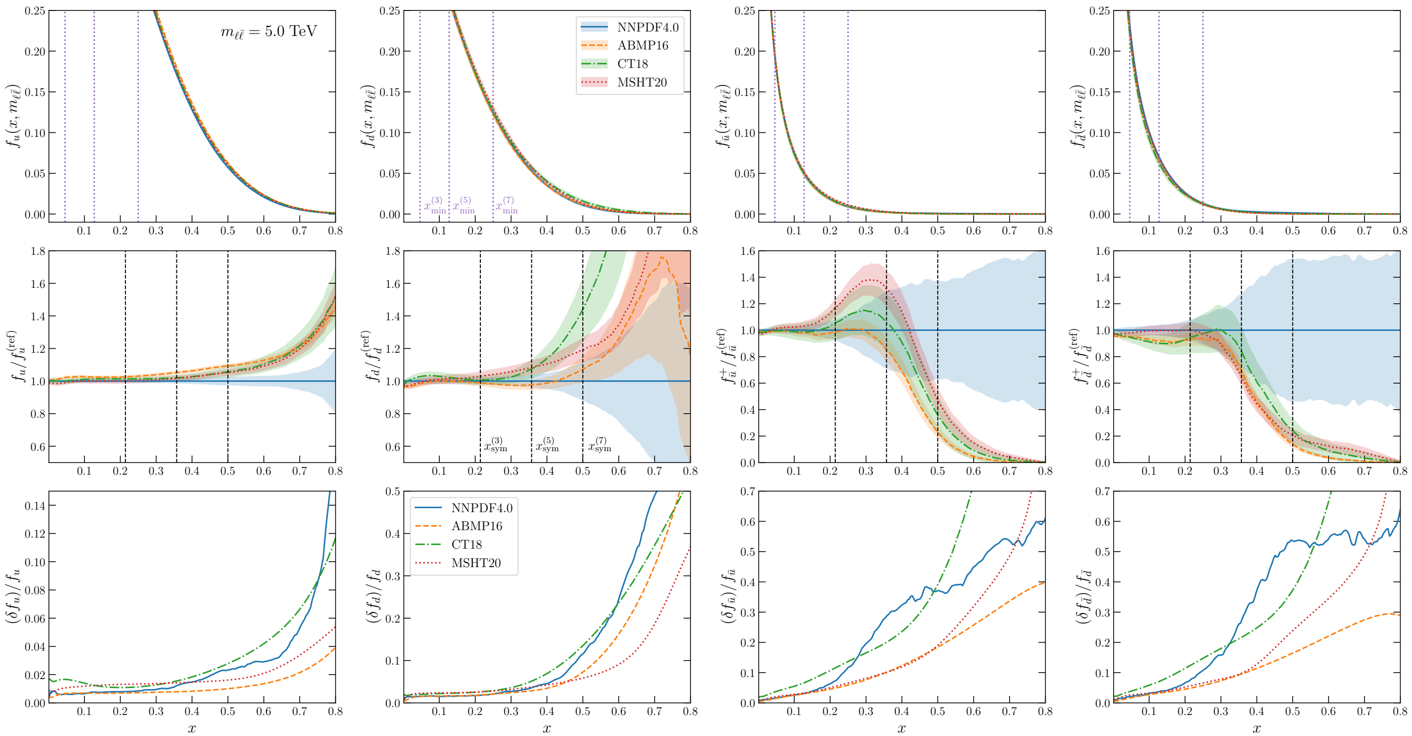
CANNOT HAVE NEGATIVE VALENCE, BUT FAST-DROPPING VALENCE ALLOWED

QUALITATIVE BEHAVIOR:
EXISTING PDF SETS

- DOMINANT CONTRIBUTION $\Rightarrow$ up AND down QUARKS, ANTIQUARKS
- AS Z' MASS CHANGES, x RANGE CHANGES: $x_1 x_2 = \frac{m_{\ell\bar{\ell}}}{\sqrt{s}}$
  BUT PDFS (LARGE x) CHANGE VERY LITTLE

x RANGES & PDF SETS

UP DOWN ANTIUP ANTIDOWN

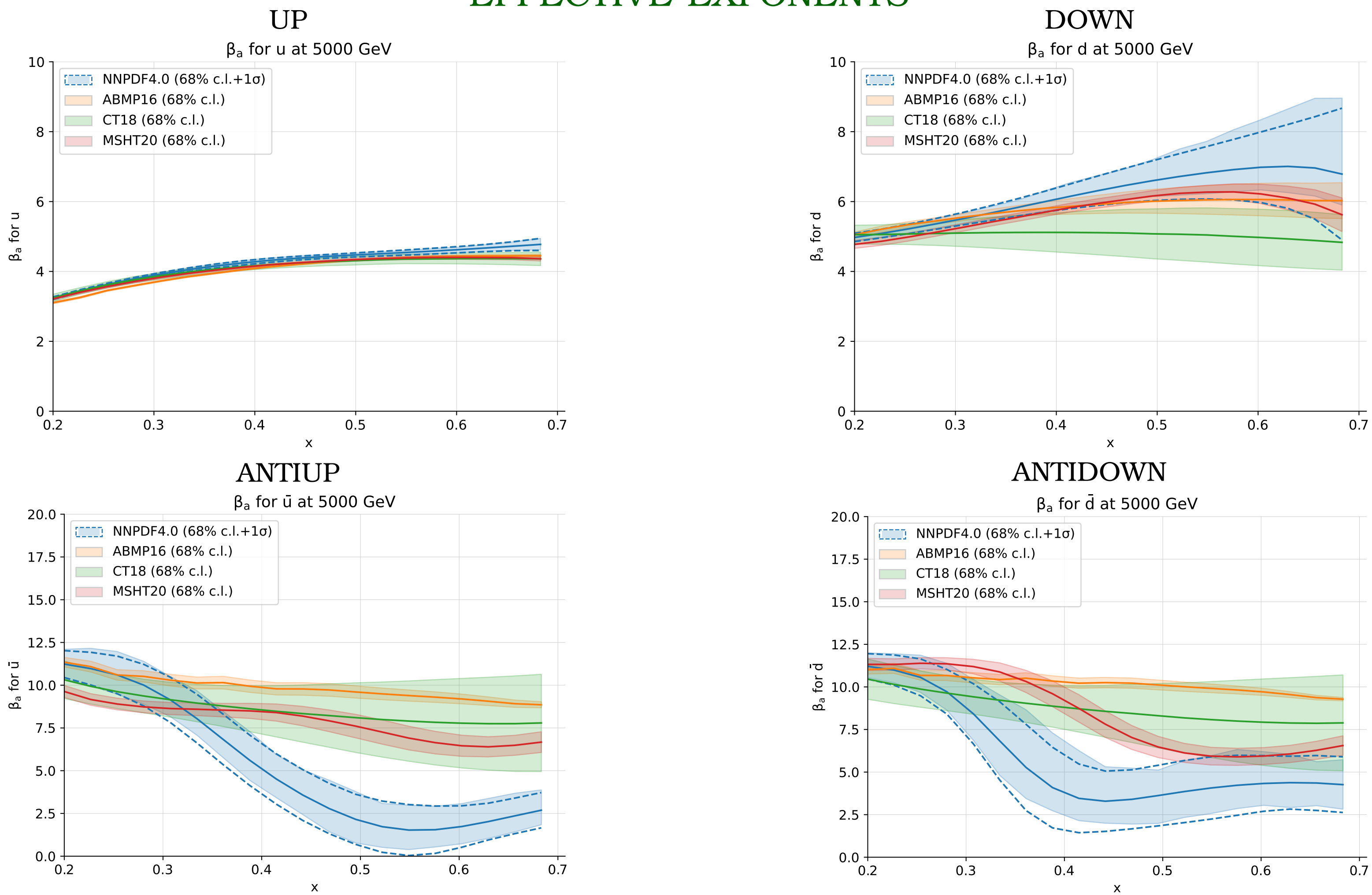


- $M \lesssim 3$ TeV $\Rightarrow$ DATA REGION, ALL PDF SETS AGREE
- $M \gtrsim 5$ TeV $\Rightarrow$ EXTRAPOLATION, NNPDF DISAGREES
  - DIFFERENT CENTRAL VALUE
  - LARGER UNCERTAINTY

PDF BEHAVIOR: WHAT'S GOING ON?

- CT, MSHT, ABMP PARAMETRIZATION: $f(x) = x^{\alpha}(1-x)^{\beta} g(x)$; NNPDF NEURAL NETWORK
- DEFINE EFFECTIVE EXPONENT $\beta(x) \equiv \frac{\partial \ln |x f(x)|}{\partial \ln(1-x)}$

EFFECTIVE EXPONENTS

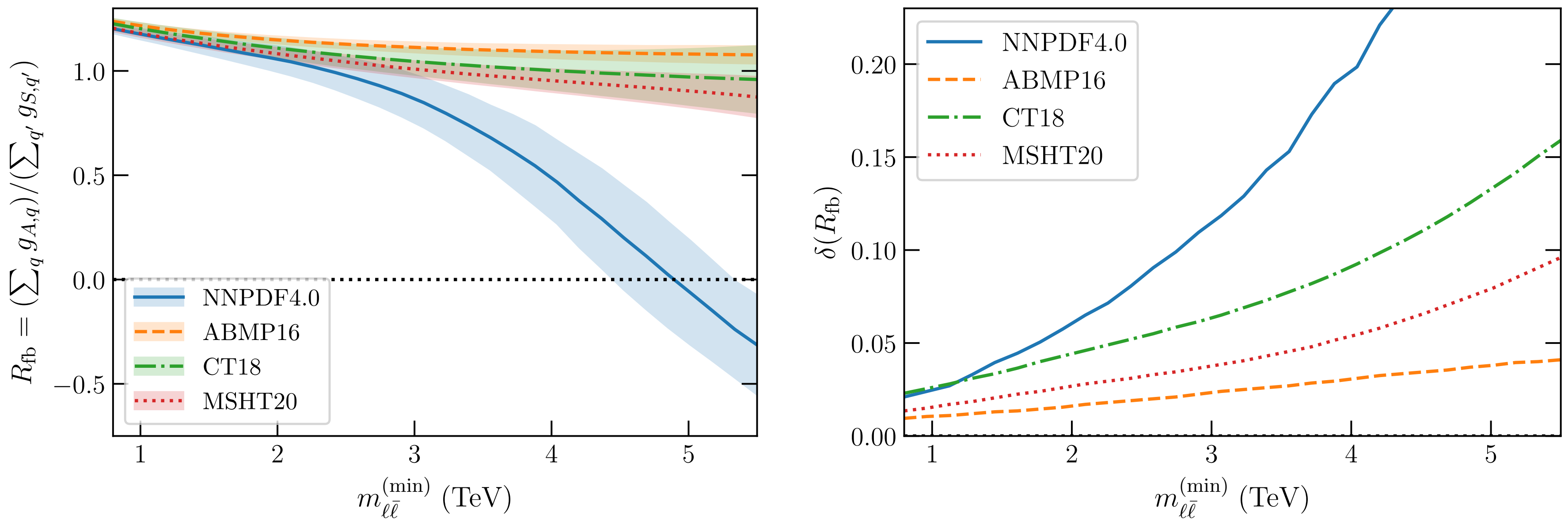


- CT, MSHT, ABMP: LARGE x β APPROX. CONSTANT
- NNPDF: β NOT FIXED BY PARAMETRIZATION

WHAT'S GOING ON?
THE EFFECTIVE COUPLING

RECALL $A_{\rm fb}(\cos\theta^*) = \frac{\cos\theta^*}{(1+\cos^2(\theta^*))}\frac{g_A}{g_{S,q'}}$; $\quad g_{A,S} \propto \int \frac{dm_{\ell\bar\ell}}{m_{\ell\bar\ell}}\frac{dx_1}{x_1}\mathcal{L}_{A,S}(m_{\ell\bar\ell},x_1)$

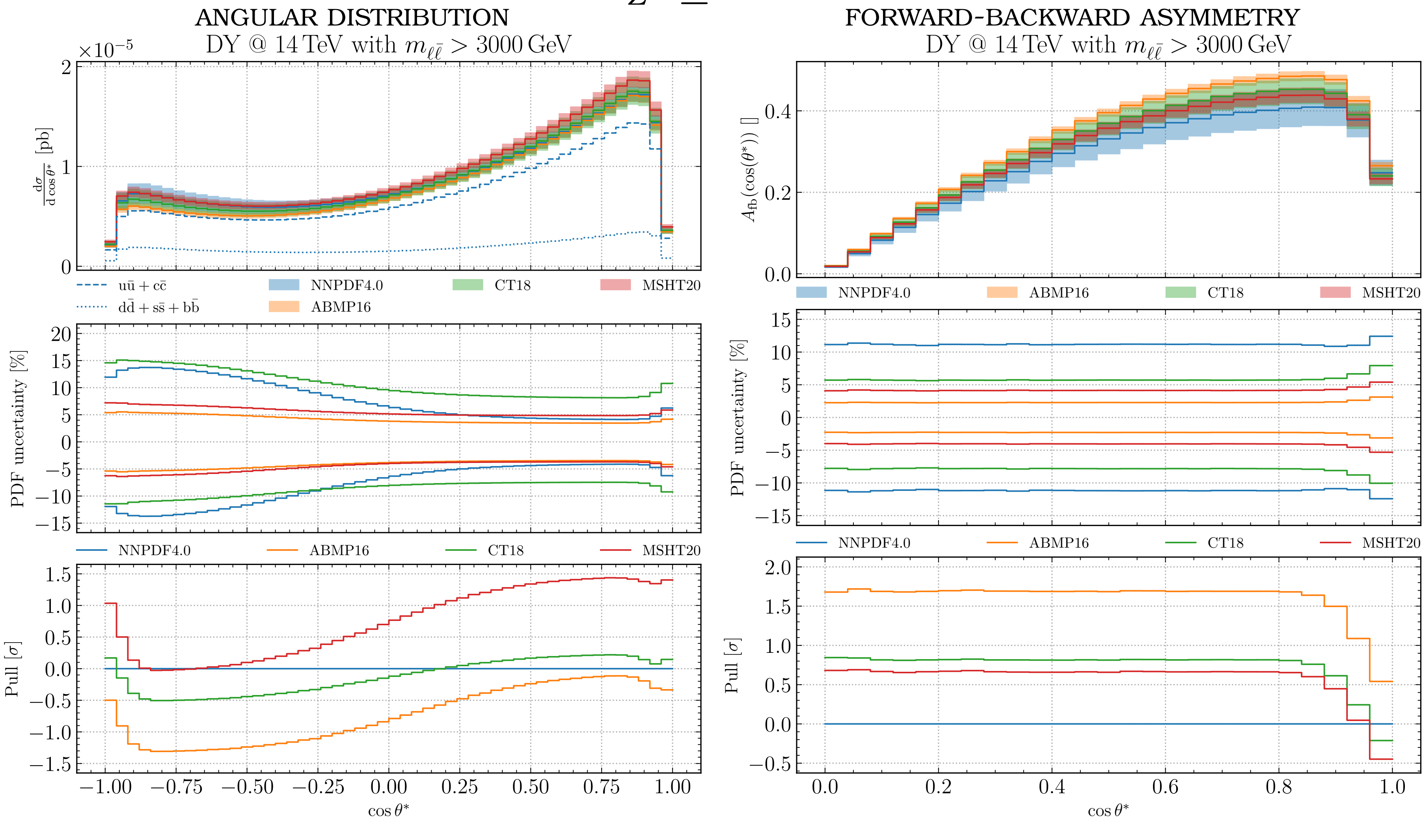
COUPLING — ABSOLUTE UNCERTAINTY



- AS SCALE INCREASES, LARGER x PROBED
- CT, MSHT, ABMP: COUPLING APPROX. SCALE INDEP.
- NNPDF: COUPLING DEPENDS ON SCALE, LARGER UNCERTAINTY

THE FORWARD-BACKWARD ASYMMETRY

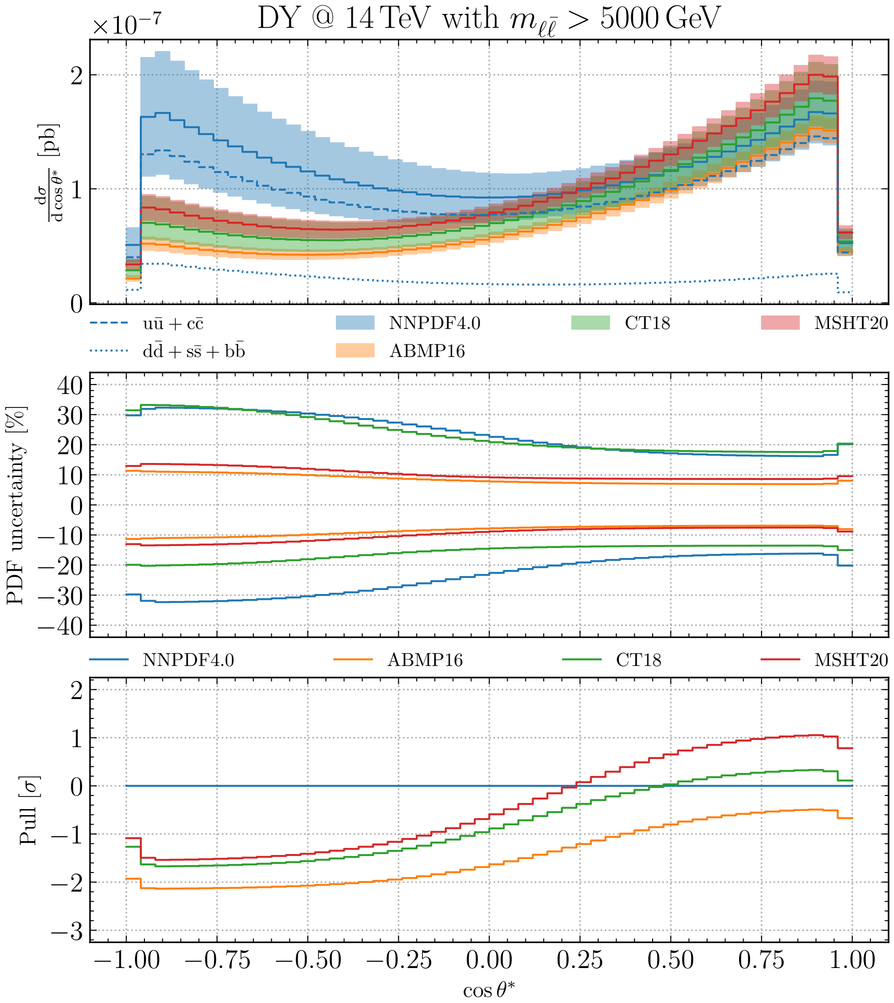
$M_{Z'} \geq 3$ TeV



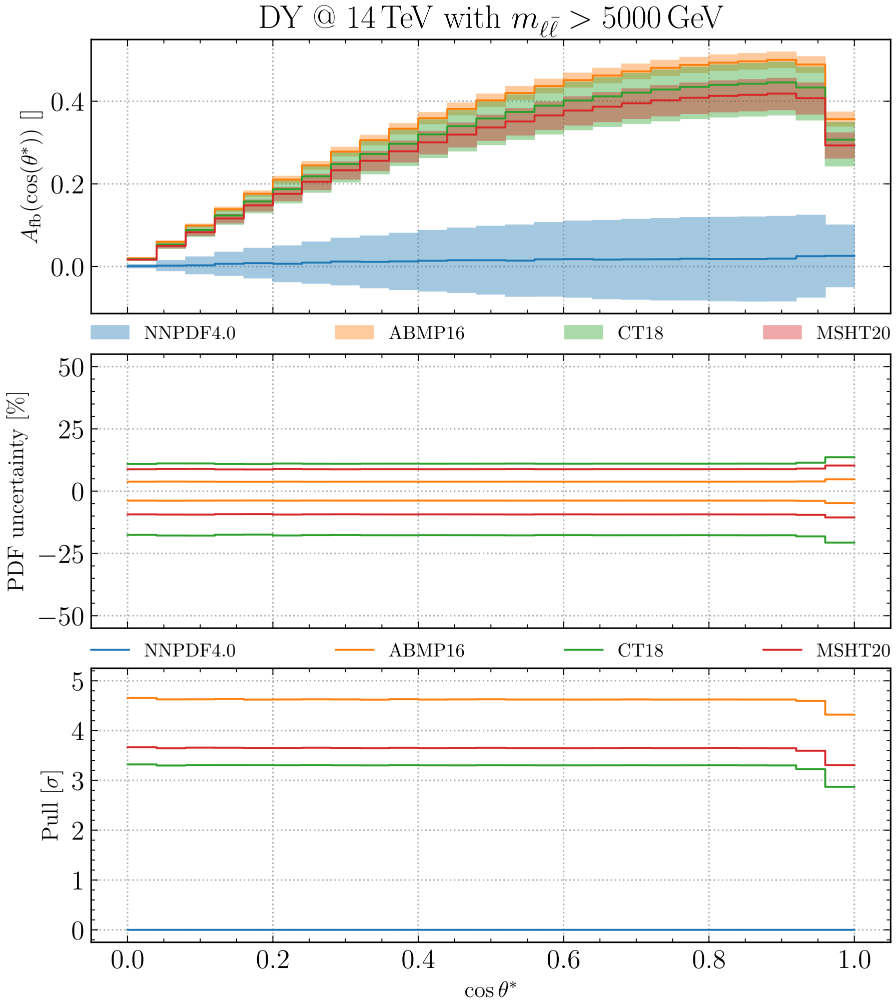
- $M_{Z'} \geq 3$ TeV: DATA REGION, ALL PDF SETS AGREE
- .

THE FORWARD-BACKWARD ASYMMETRY
$M_{Z'} \geq 5$ TeV

ANGULAR DISTRIBUTION

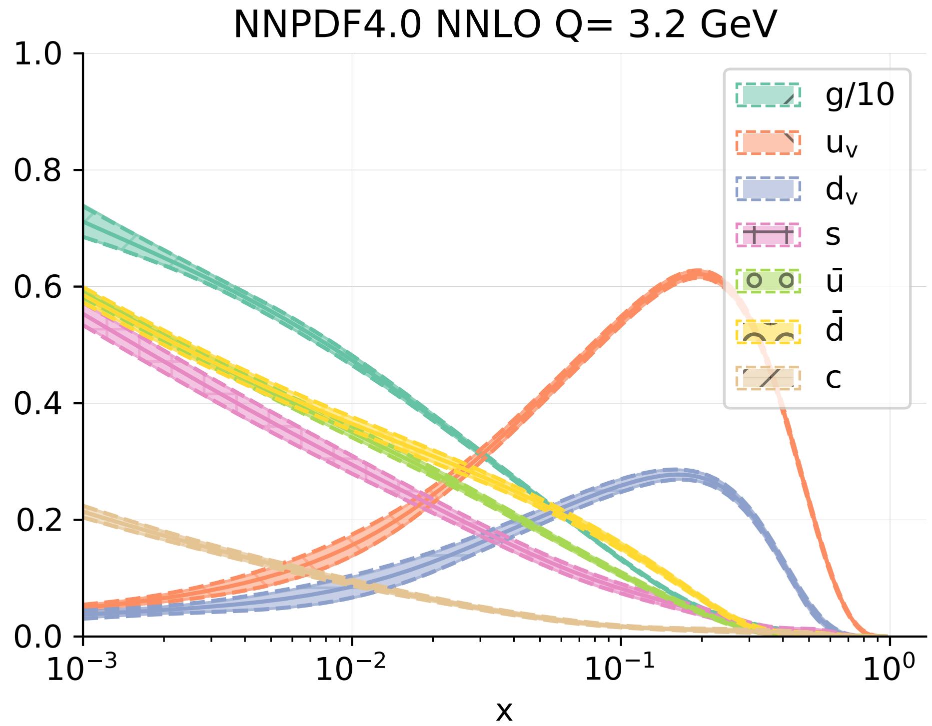


FORWARD-BACKWARD ASYMMETRY



- $M_{Z'} \geq 3$ TeV: DATA REGION, ALL PDF SETS AGREE
- $M_{Z'} \geq 5$ TeV
  - CT, MSHT, ABMP $\Rightarrow$ ASYMMETRY UNCHANGED WITH INCREASING SCALE
  - NNPDF $\Rightarrow$ ASYMMETRY DISAPPEARS AS SCALE INCREASES

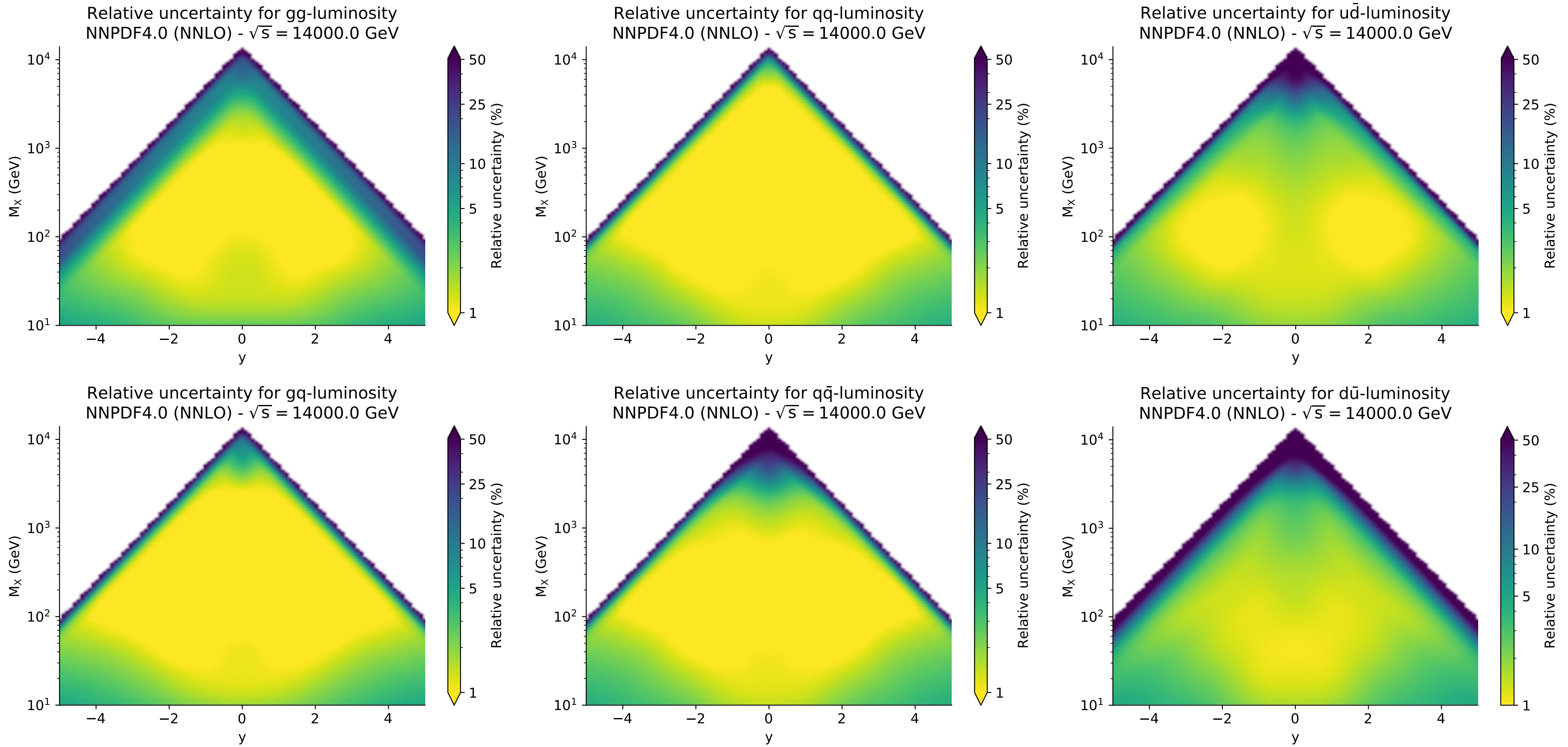
PDFs: THE STATE OF THE ART (NNPDF4.0, 2021)



- A SET OF PROBABILITY DISTRIBUTIONS OF PROBABILITY DISTRIBUTIONS
- FULL (INFINITE DIMENSIONAL) COVARIANCE MATRIX
- MUST BE DETERMINED FROM FINITE SET OF DISCRETE DATA

UNCERTAINTIES: STATE OF THE ART (NNPDF4.0, 2021)

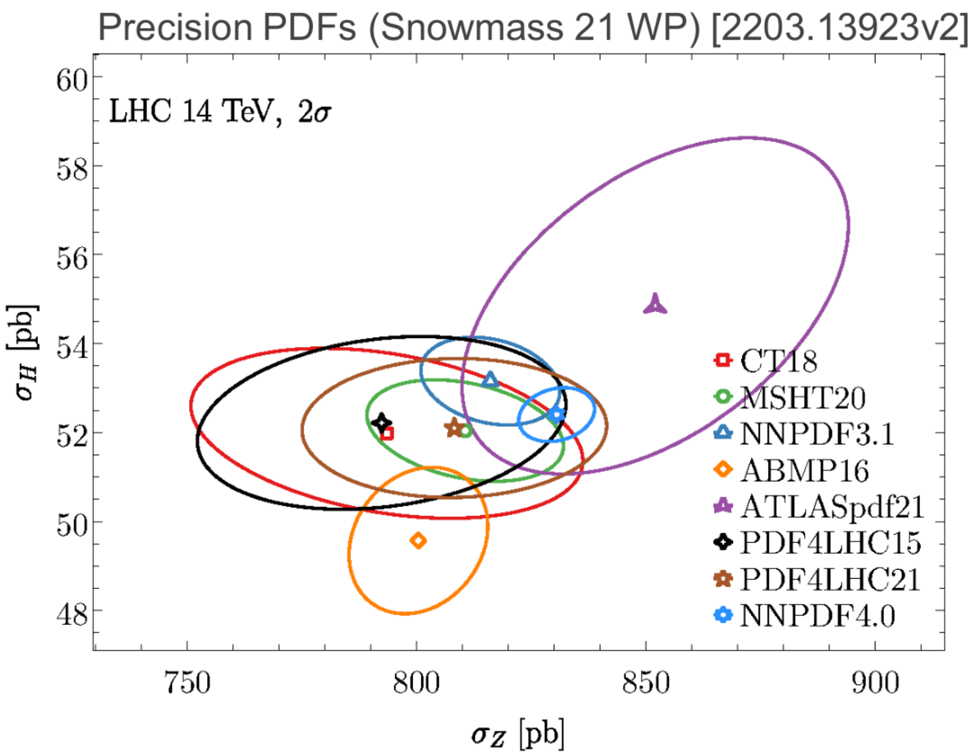
GLUON SINGLET FLAVORS

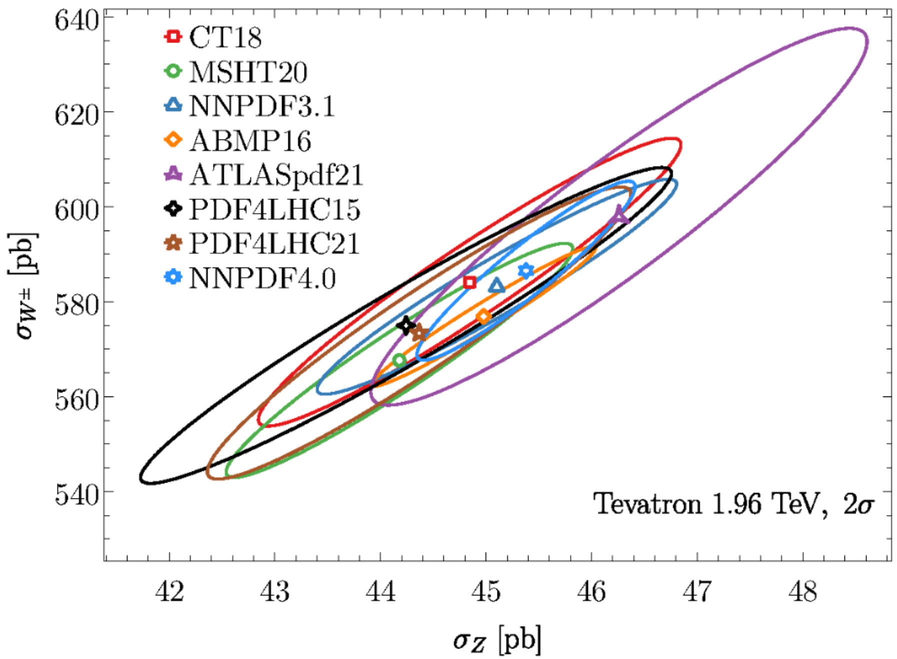


- TYPICAL UNCERTAINTIES IN DATA REGION: SINGLET $\sim 1\%$, NONSINGLET $\sim 2-3\%$
- DATA REGION: $10 \lesssim M_X \lesssim 3 \cdot 10^3$ TeV, $-4 \lesssim y \lesssim 4$

UNCERTAINTIES: STATE OF THE ART (NNPDF4.0, 2021)

Higgs, W, Z production

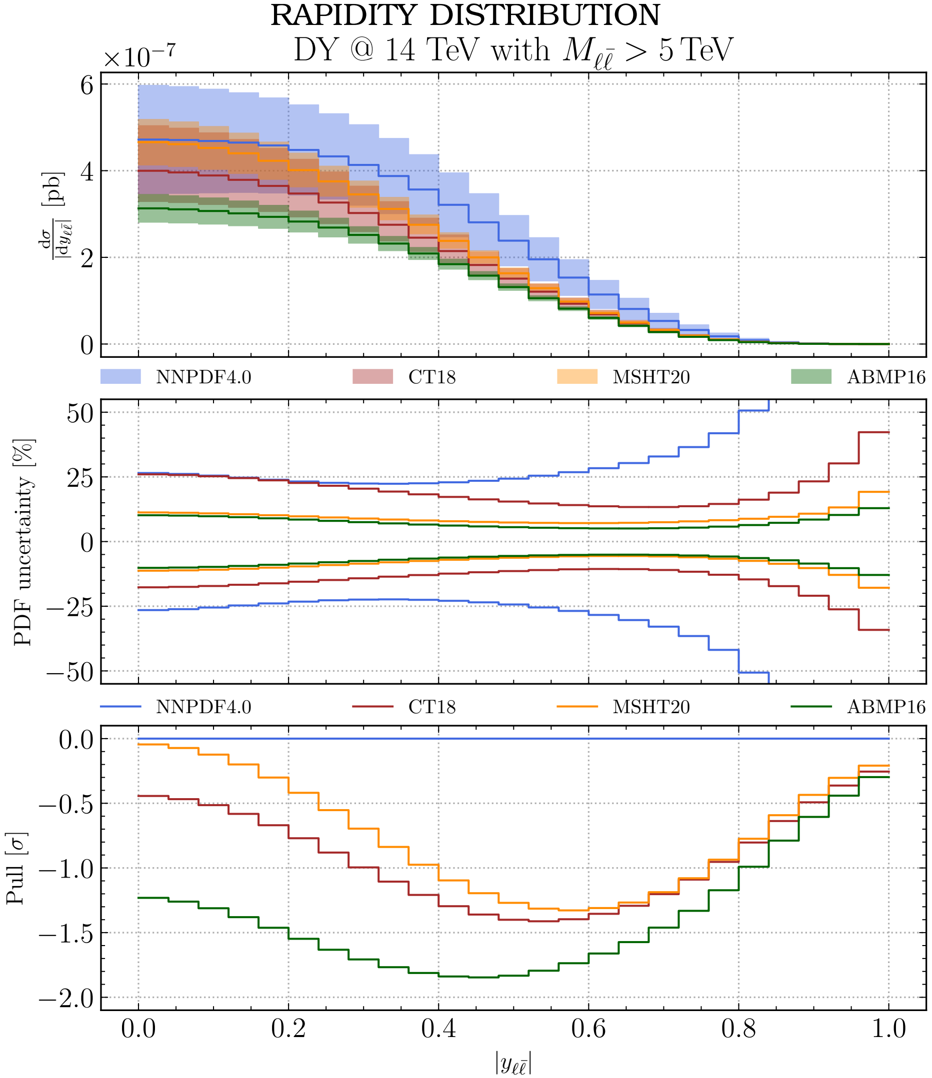


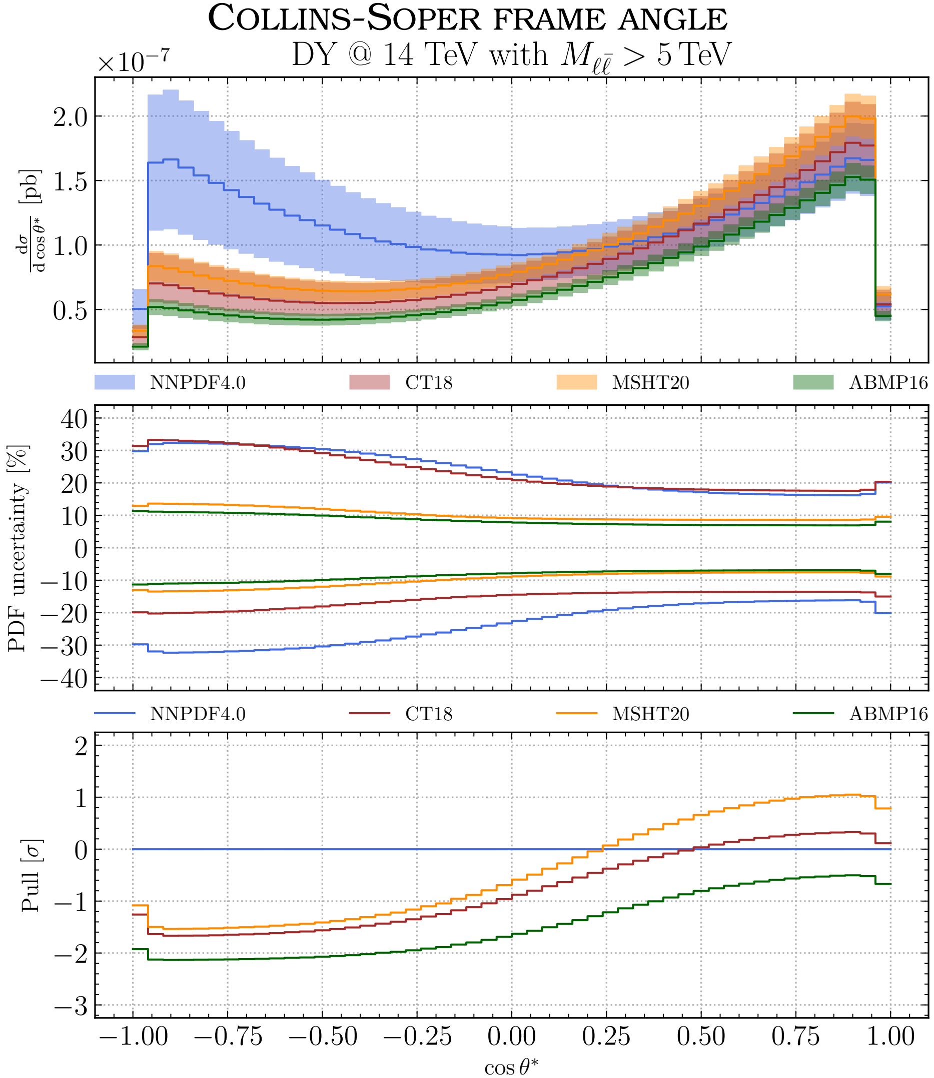


DATA REGION: NNPDF4.0 UNCERTAINTIES $\Rightarrow$ SMALLER
THAN OTHER PDF SETS
BACKWARD CONSISTENT WITH NNPDF3.1

UNCERTAINTIES: STATE OF THE ART (NNPDF4.0, 2021)

Z' production





EXTRAPOLATION REGION: NNPDF4.0 UNCERTAINTIES $\Rightarrow$ LARGER THAN OTHER PDF SETS

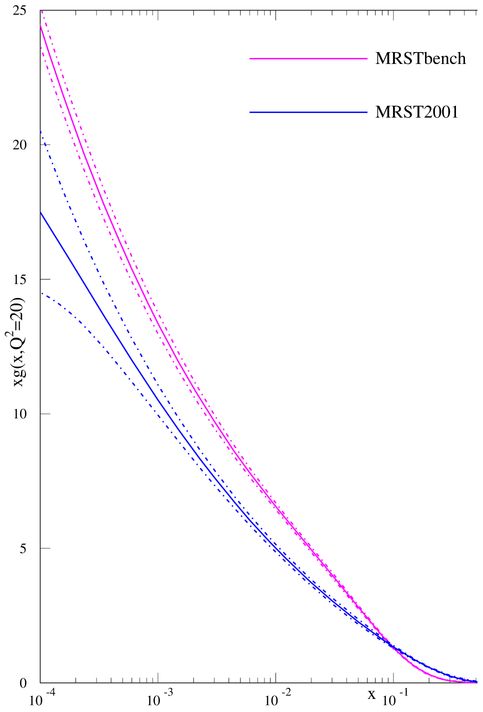
UNBIASED PDF BEHAVIOR

UNCERTAINTY ESTIMATION

THE PDF UNCERTAINTY PROBLEM: THE HERA-LHC BENCHMARK (2005)

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

BENCHMARK VS DEFAULT GLUON

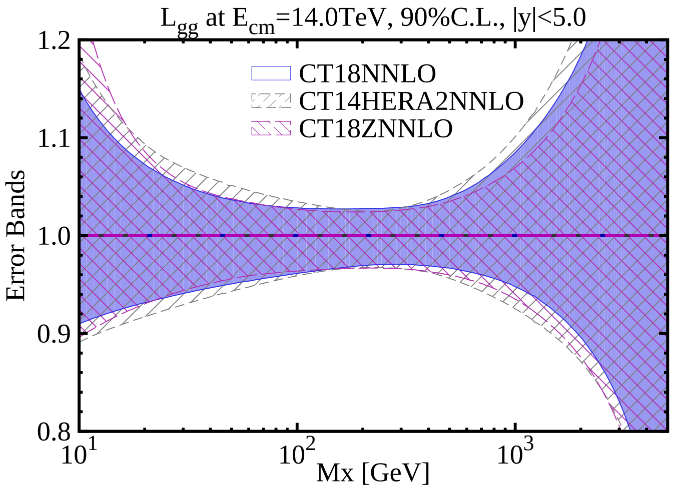


"...the partons extracted using a very limited data set are completely incompatible, even allowing for the uncertainties, with those obtained from a global fit with an identical treatment of errors...The comparison illustrates the problems in determining the true uncertainty on parton distributions."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

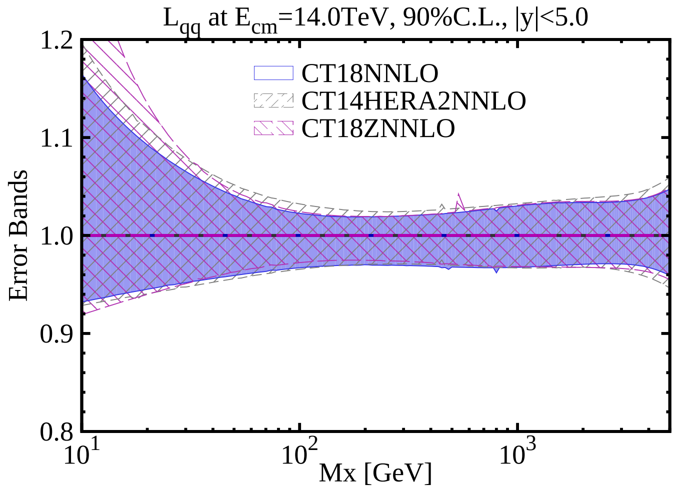
THE PDF UNCERTAINTY PROBLEM:
THE PRE-ML APPROACH

CT18 VS. CT14: PARTON LUMINOSITY UNCERTAINTIES

GLUON-GLUON



QUARK-QUARK



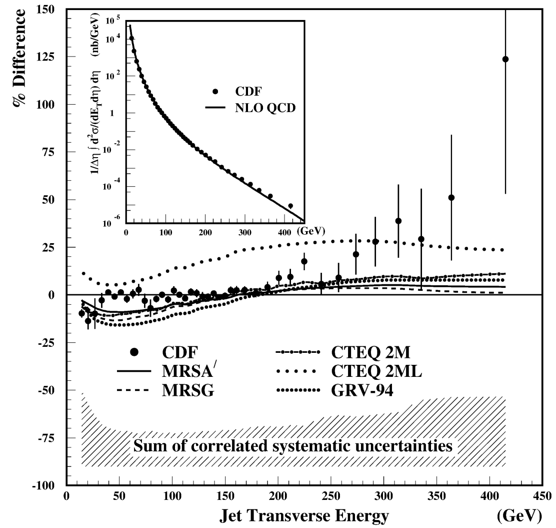
MORE DATA $\Rightarrow$ BIGGER UNCERTAINTIES (?!)

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: $g(x, Q = Q_0) = x^{a_1 - 1} (1-x)^{a2} \left[a_3 (1-y)^3 + a_4 3y(1-y)^2 + a_5 3y^2 (1-y) + y^3\right]$;
  $y = \sqrt{x}$; $a_5 = (3 + 2a_1)/3$.

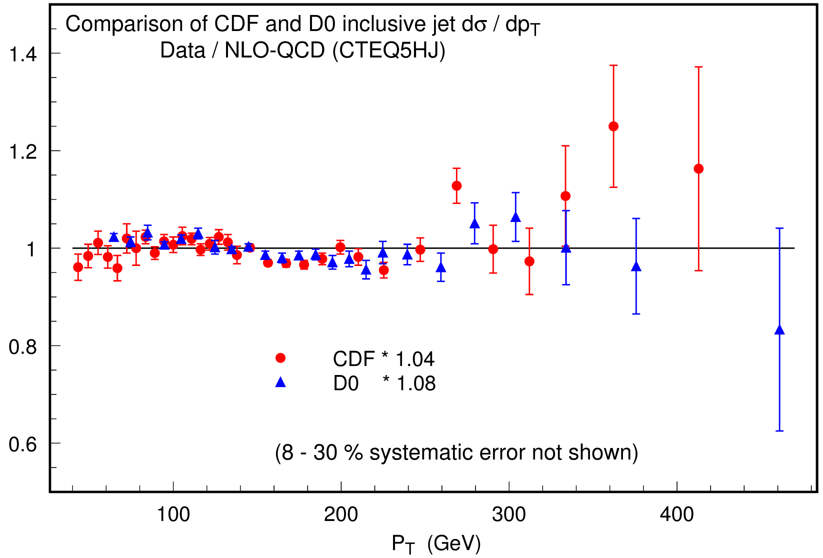
PDF UNCERTAINTIES AND NEW PHYSICS

- DISCREPANCY BETWEEN QCD CALCULATION AND CDF JET DATA (1995)
- EVIDENCE FOR QUARK COMPOSITENESS?
- RESULT STRONGLY DEPENDS ON GLUON AT $x \gtrsim 0.1$
- PDF MUST VANISH AT $x = 0$, BUT (THEN) NO DATA FOR $x \gtrsim 0.05$!



DISCREPANCY REMOVED IF JET DATA USED FOR GLUON DETERMINATION

NEW CTEQ GLUON (1998)

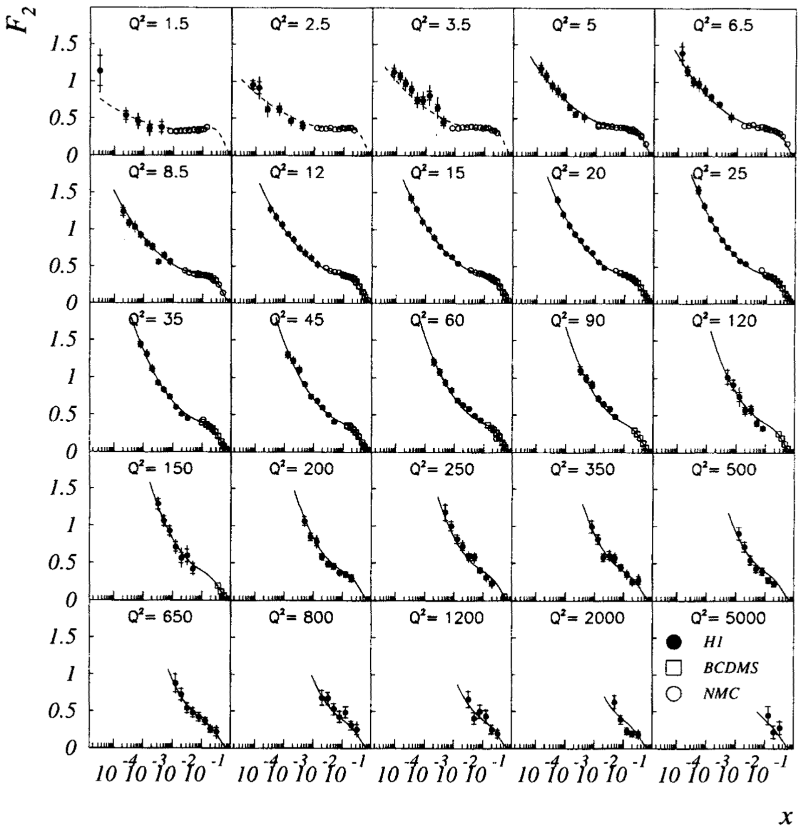


NOW: NO DATA FOR $x \gtrsim 0.5 \Rightarrow$ DISCOVERY (THRESHOLD) REGION!

EXTRAPOLATION AND THEORY BIAS

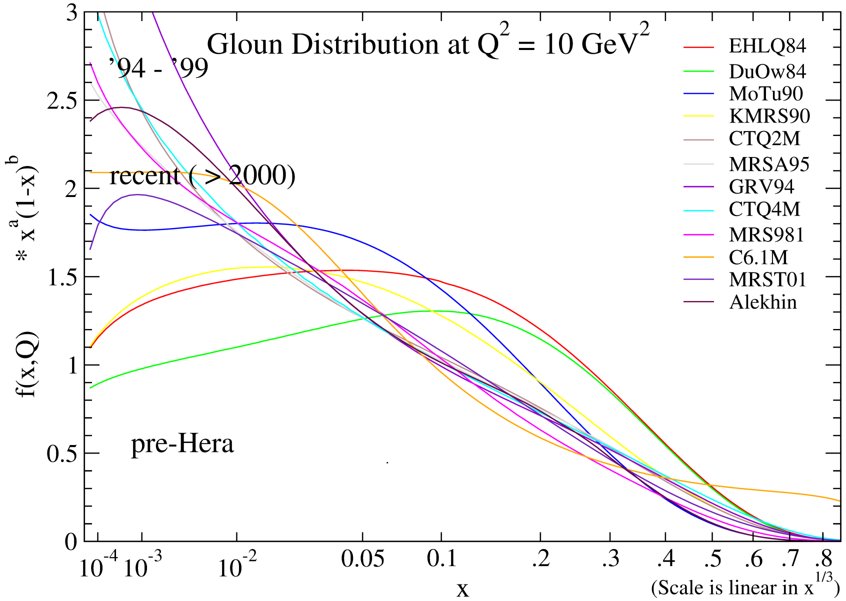
1995: THE RISE OF STRUCTURE FUNCTIONS AT HERA

FIRST HERA DATA VS OLDER DATA



A. de Roeck, Cracow epiphany conf. 1996

HISTORICAL COMPILATION OF GLUON PDFs

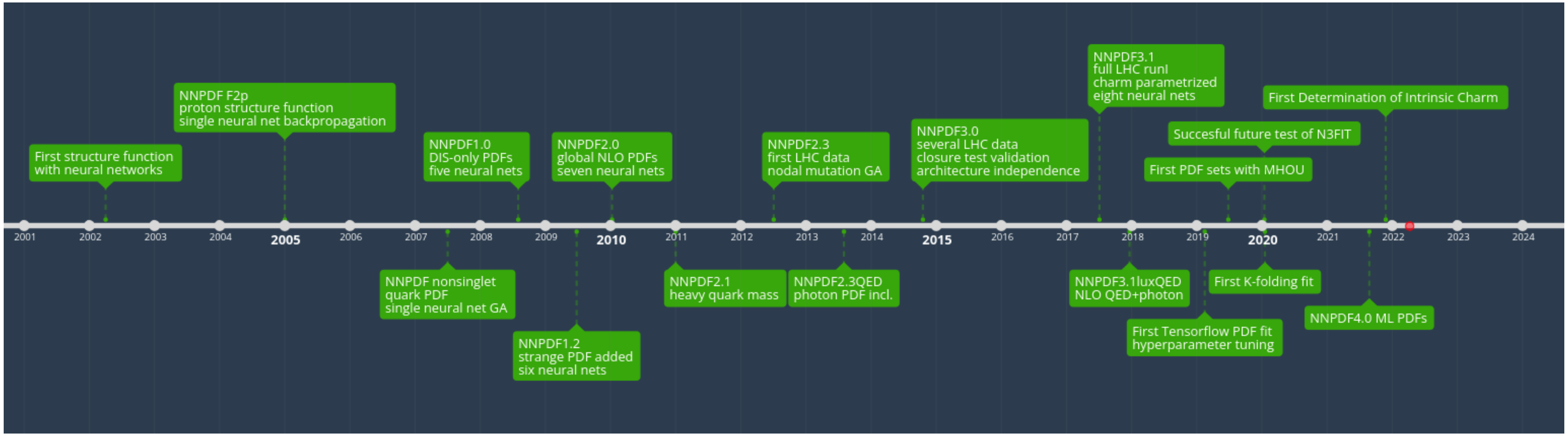


W.K.Tung, DIS 2004

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

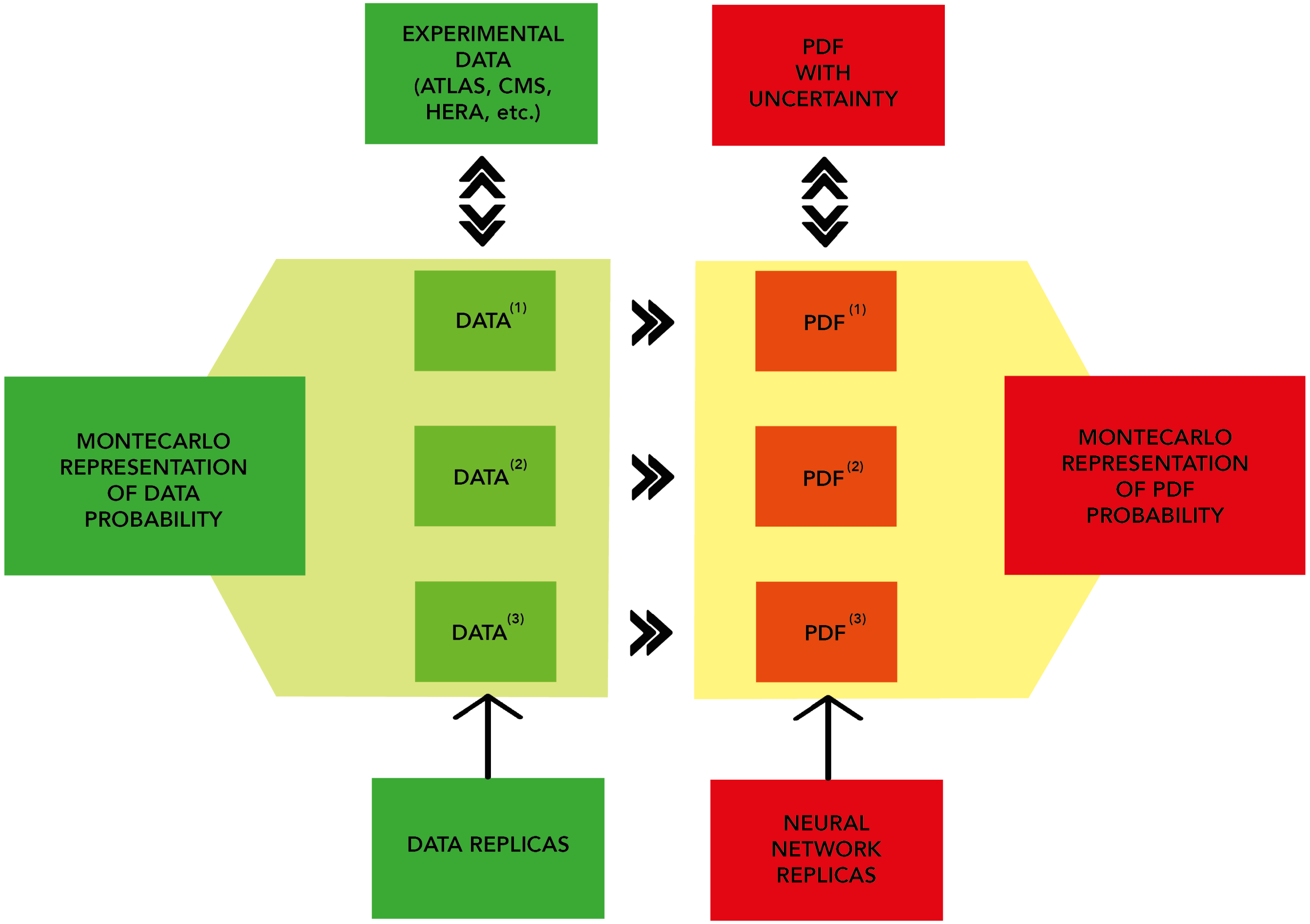
THE NNPDF METHODOLOGY

PROTON STRUCTURE AS AN AI PROBLEM:
NNPDF



THE FUNCTIONAL MONTE CARLO

REPLICA SAMPLE OF FUNCTIONS $\Leftrightarrow$ PROBABILITY DENSITY IN FUNCTION SPACE
KNOWLEDGE OF LIKELIHHOD SHAPE (FUNCTIONAL FORM) NOT NECESSARY

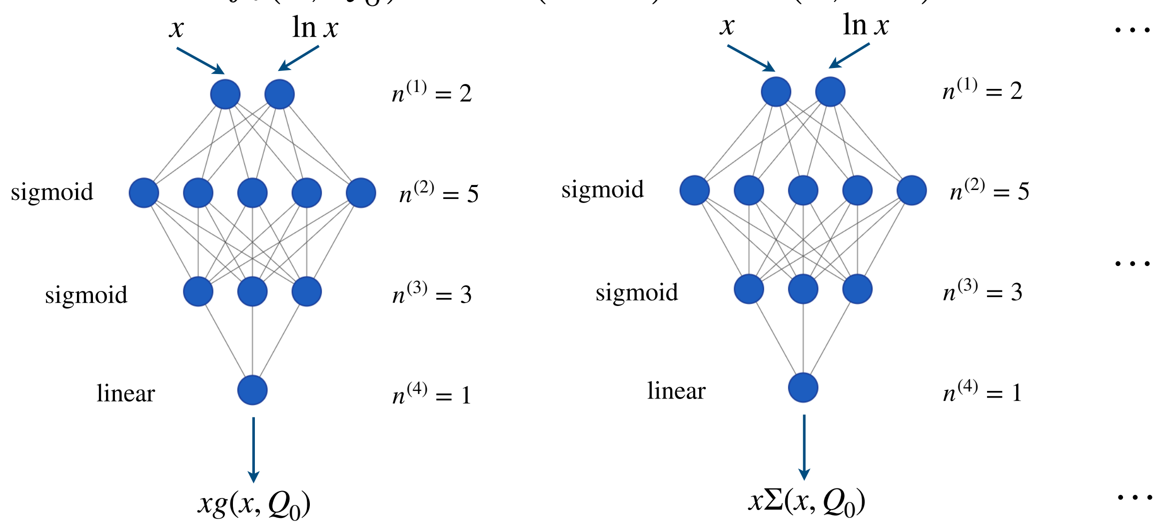


FINAL PDF SET: $f_i^{(a)}(x,\mu)$;
i =up, antiup, down, antidown, strange, antistrange, charm, gluon; $j = 1, 2, \ldots N_{\rm rep}$

OLD NNPDF (UP TO 2021)

NEURAL NETWORKS

$$f_i(x, Q_0^2) = x^{\alpha_i}(1-x)^{\beta_i} NN(x, \ln x)$$



- 37 PARAMETERS $\times$ 8 PDFs
- PREPROCESSING EXPONENTS RANDOMIZED REPLICA PER REPLICA, RANGE DETERMINED SELF-CONSISTENTLY

GENETIC MINIMIZATION

- NODAL MUTATION REPLACED POINT MUTATION
- SINGLE EPOCH WITH NO REWEIGHTING UNLIKE PREVIOUS

OPTIMAL FIT

- CROSS-VALIDATION WITH 50-50 TRAINING & VALIDATION FRACTIONS
- LOOKBACK STOPPING REPLACED THRESHOLD ON DERIVATIVE

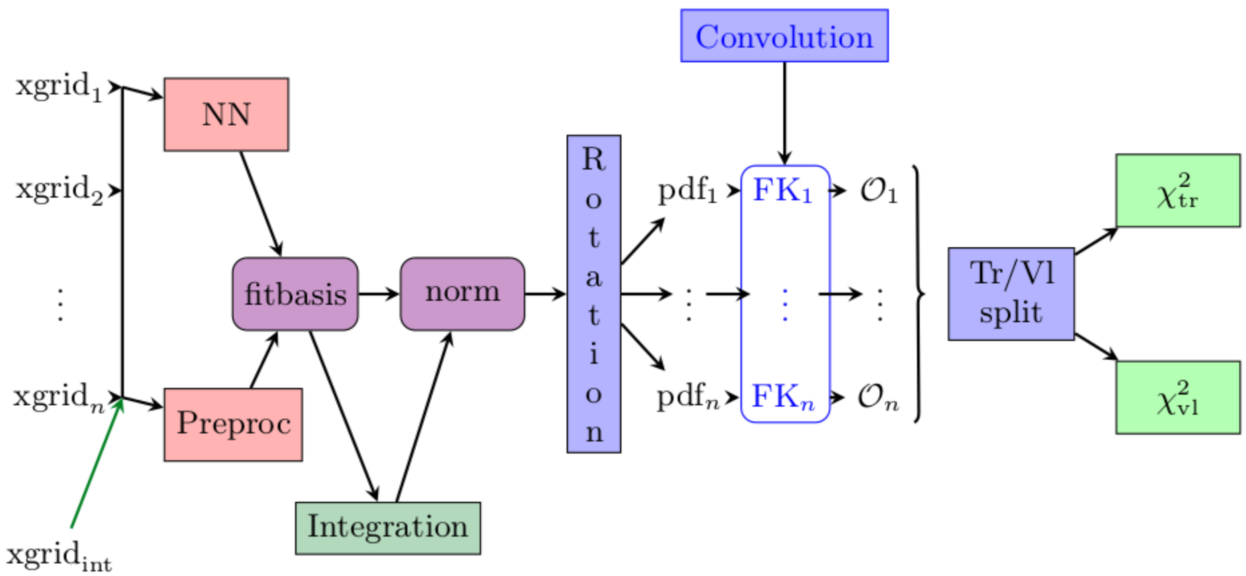
MACHINE LEARNING PDFs

LEARNING THE METHODOLOGY

THE N3FIT PROJECT

HOW DO WE KNOW THAT THE METHODOLOGY IS THE BEST?
"ACCUMULATED WISDOM" INEFFICIENT AND SLOW

CHANGE OF PHILOSOPHY ⇒ DETERMINISTIC MINIMIZATION (GRADIENT DESCENT)
GO FOR THE ABSOLUTE MINIMUM, AND (HYPER)OPTIMIZE



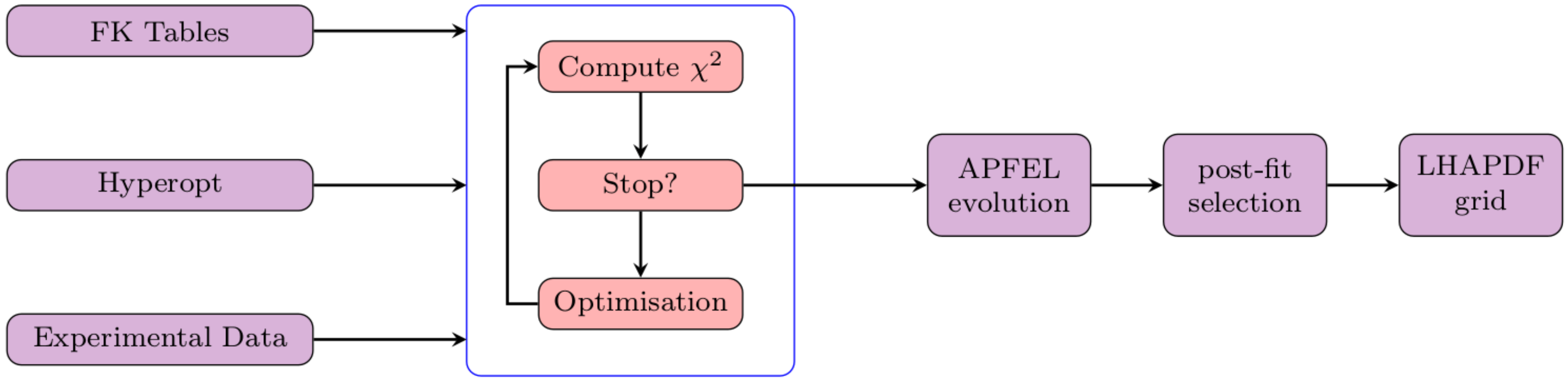
- PYTHON-BASED KERAS + TENSORFLOW FRAMEWORK
- EACH BLOCK INDEPENDENT LAYER
- CAN VARY ALL ASPECTS OF METHODOLOGY

THE NNPDF CODE STRUCTURE

- MODULAR PYTHON-BASED CODE
- HIGH DEGREE PARALLELIZATION & HARDWARE ACCELERATION

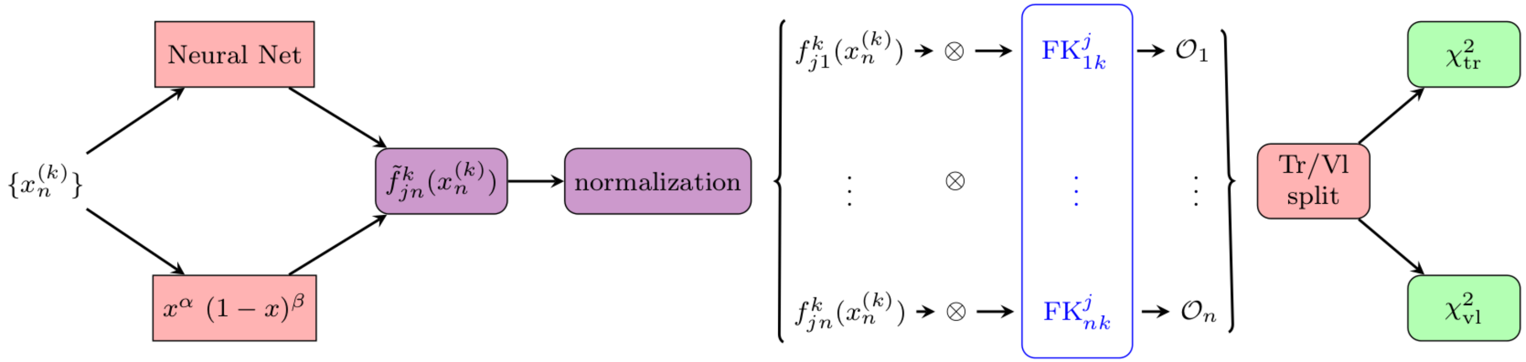
AVERAGE FITTING TIME PER REPLICA AND USE OF RESOURCES
SAME DATASET FOR OLD AND NEW METHODOLOGIES IN CPU AND GPU
CPU: INTEL(R) CORE(TM) I7-4770 AT 3.40GHZ; GPU: NVIDIA TITAN V

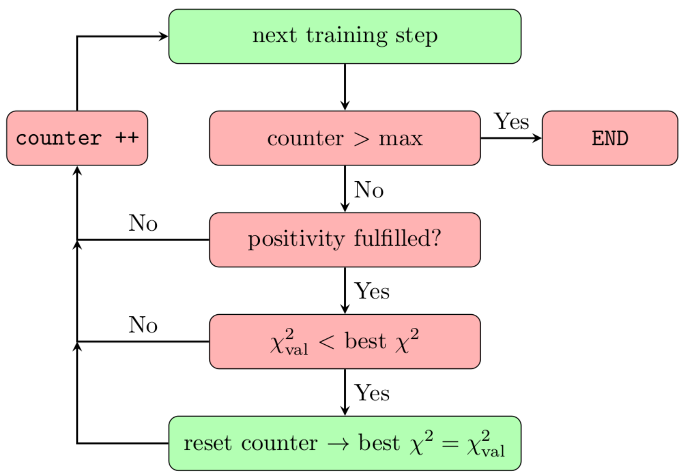
	NNPDF31 CODEBASE	NNPDF40 CODEBASE IN CPU	NNPDF40 CODEBASE IN GPU
TIME	15.2 H.	38 ± 5 MIN.	6.6 MIN.
RAM USE	1.5 GB	6.1 GB	NA



MINIMIZATION AND CROSS-VALIDATION

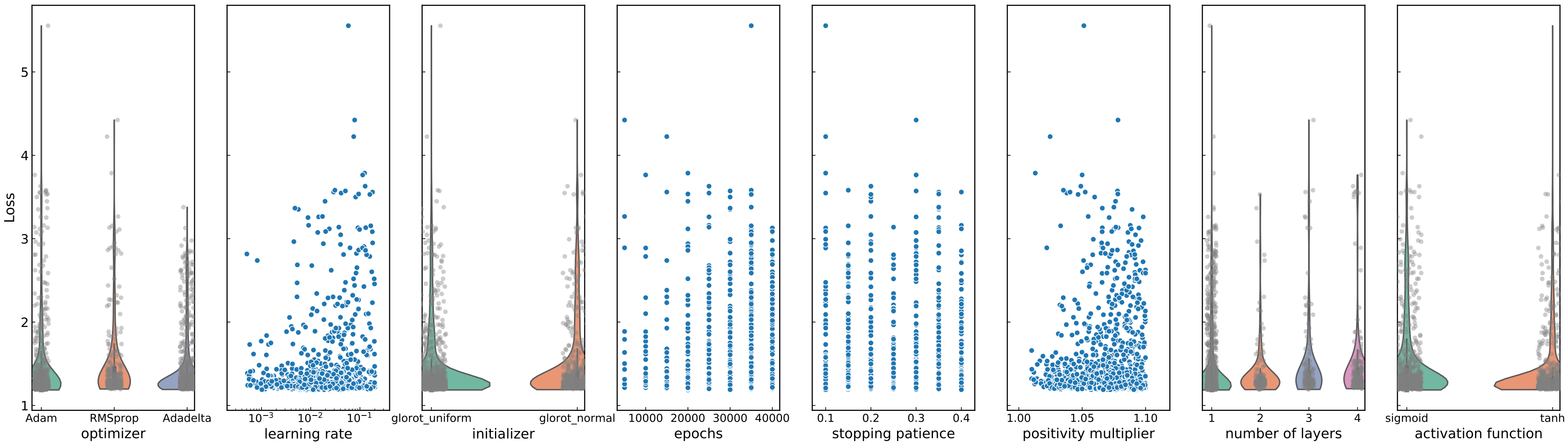
- DATA REPLICAS $\Rightarrow$ PDF REPLICAS
- EACH PDF REPLICA: PREPROCESSED NEURAL NET
- NEURAL NET $\Rightarrow$ OBSERVABLES
- RANDOM TRAINING-VALIDATION SPLIT, χ^2 TO TRAINING DATA REPLICAS MINIMIZED
- TRAINING STOPS IF VALIDATION χ^2 GROWS FOR A WHILE (PATIENCE)
- LOWEST VALIDATION χ^2 $\Rightarrow$ OPTIMAL FIT





FITTING THE METHODOLOGY

HYPEROPTIMIZATION SCANS



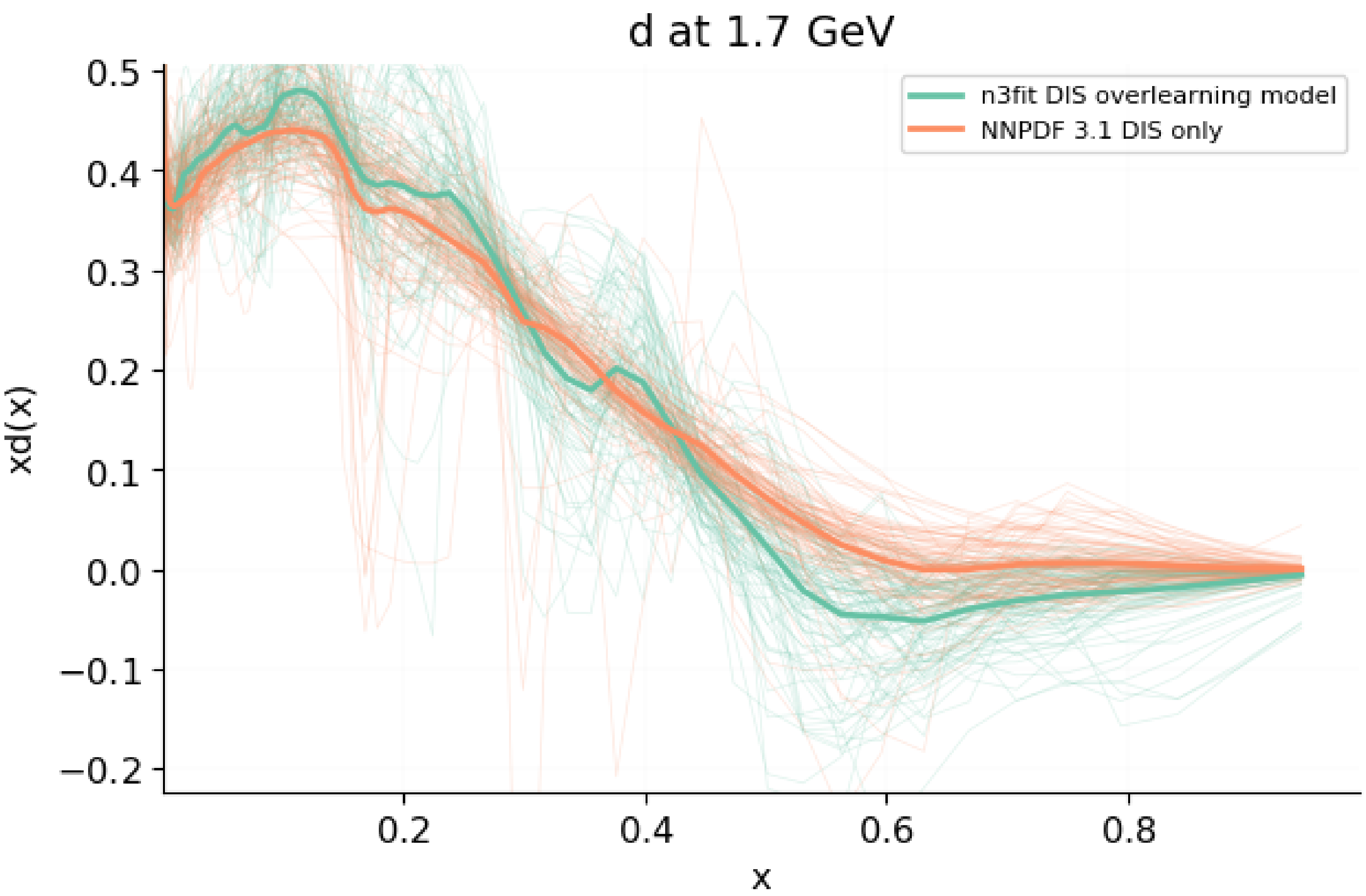
HYPEROPT PARAMETERS

NEURAL NETWORK	FIT OPTIONS
NUMBER OF LAYERS (*)	OPTIMIZER (*)
SIZE OF EACH LAYER	INITIAL LEARNING RATE (*)
DROPOUT	MAXIMUM NUMBER OF EPOCHS (*)
ACTIVATION FUNCTIONS (*)	STOPPING PATIENCE (*)
INITIALIZATION FUNCTIONS (*)	POSITIVITY MULTIPLIER (*)

- SCAN PARAMETER SPACE
- OPTIMIZE FIGURE OF MERIT: VALIDATION χ^2
- BAYESIAN UPDATING

HYPEROPTIMIZATION: OVERFITTING

DOWN QUARK: HYPEROPTIMIZED VS. HAND-PICKED



- NOT HYPEROPTIMIZED: WIGGLES: FINITE SIZE $\Rightarrow$ WILL GO AWAY AS $N_{\rm rep}$ GROWS
- N3FIT: WIGGLY PDFS $\Leftrightarrow$ OVERFITTING $\Rightarrow$ WILL NOT GO AWAY ($\chi^2_{\rm train} \ll \chi^2_{\rm valid}$!!)

WHAT HAPPENED?

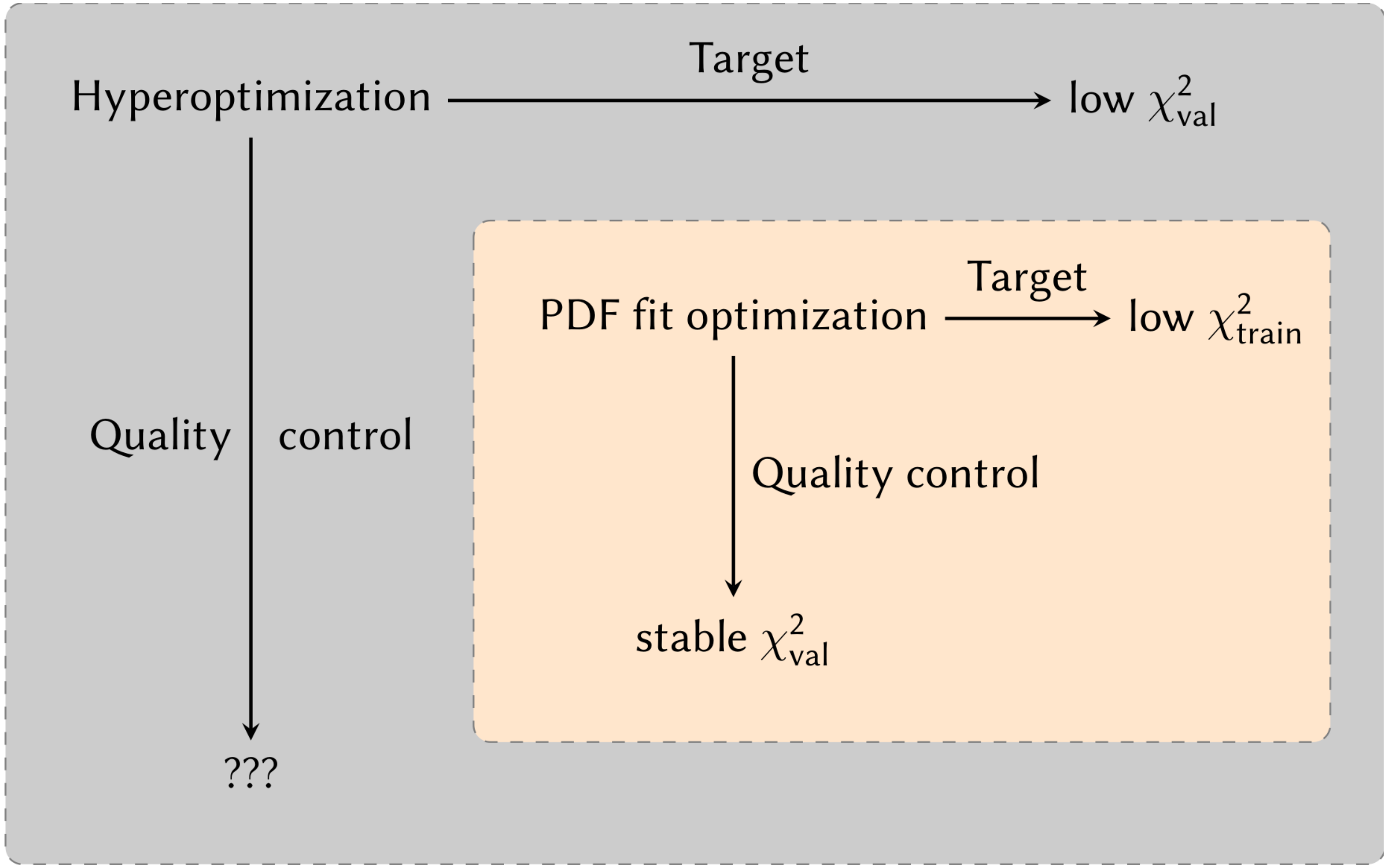
OPTIMIZATION

$$\begin{tikzcd} \text{PDF fit optimization} \arrow[r, "\text{Target}"] \arrow[d, "\text{Quality control}"] & \text{low } \chi^2_{\text{train}} \\ \text{stable } \chi^2_{\text{val}} & \end{tikzcd}$$

CROSS-VALIDATION SELECTS THE OPTIMAL MINIMUM

WHAT HAPPENED?

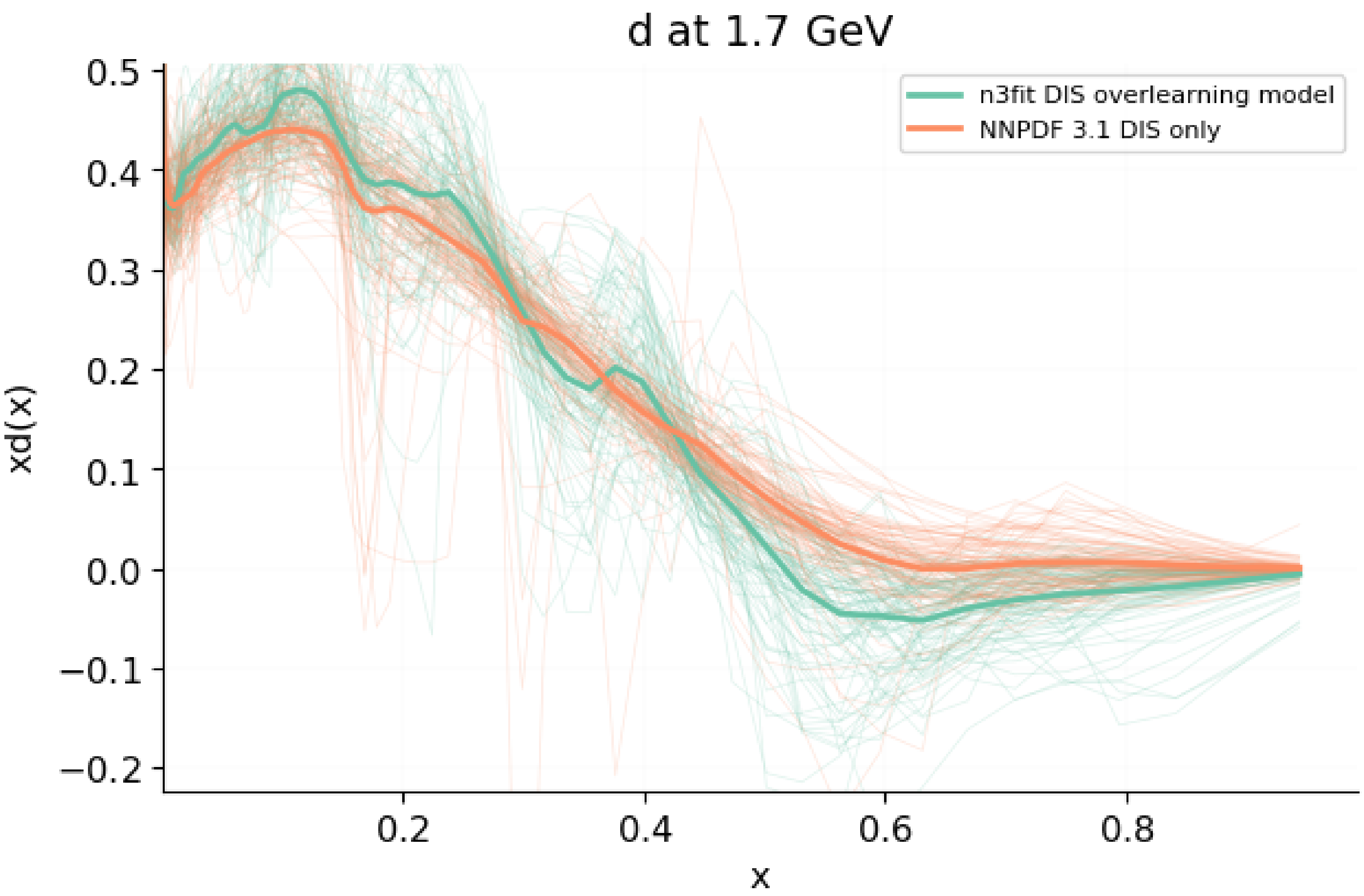
HYPEROPTIMIZATION



WE ARE MISSING A SELECTION CRITERION

HYPEROPTIMIZATION: OVERFITTING

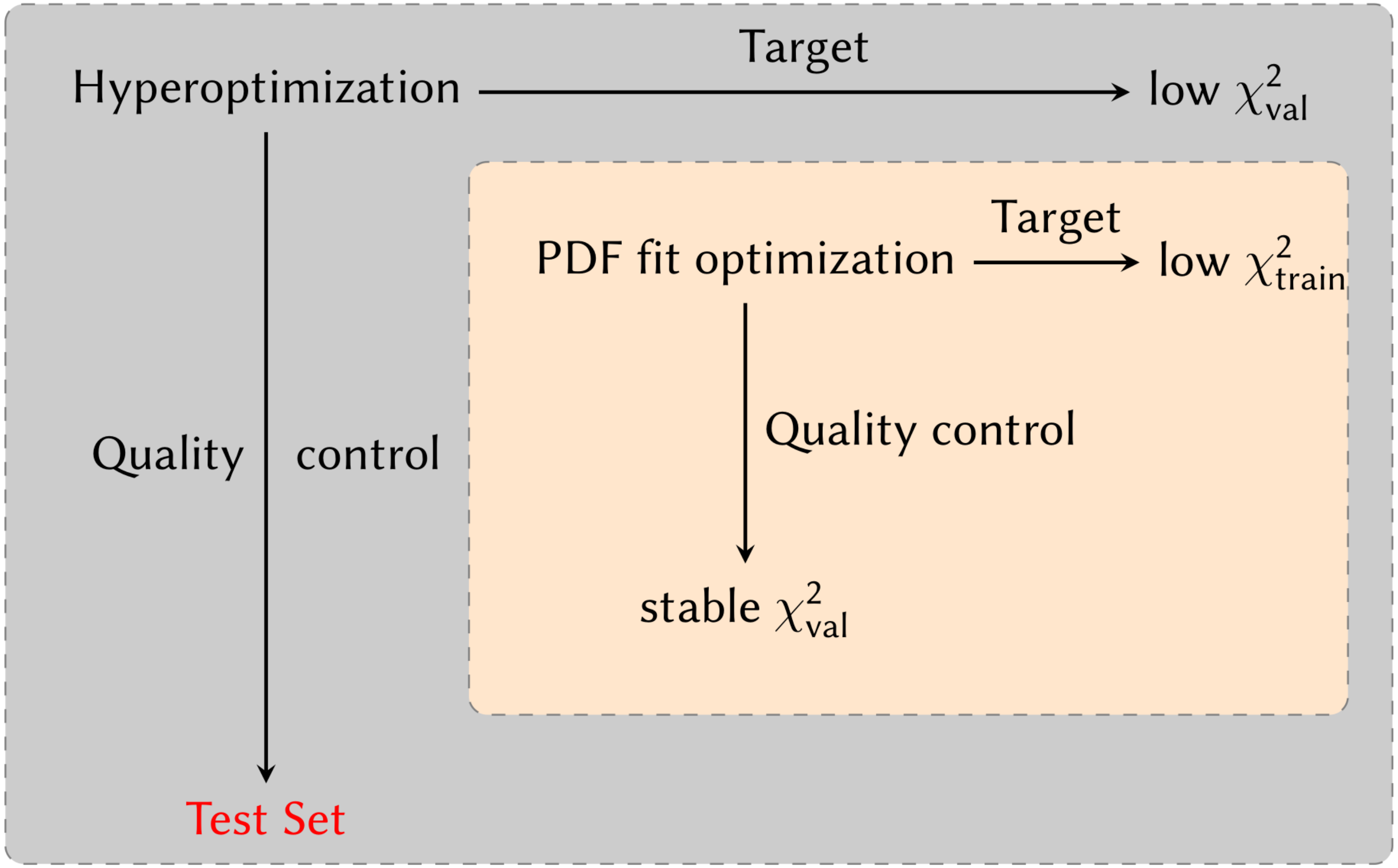
DOWN QUARK: HYPEROPTIMIZED VS. HANDPICKED



- HANDPICKED: WIGGLES: FINITE SIZE $\Rightarrow$ WILL GO AWAY AS $N_{\rm rep}$ GROWS
- N3FIT: WIGGLY PDFS $\Leftrightarrow$ OVERFITTING $\Rightarrow$ WILL NOT GO AWAY ($\chi^2_{\rm train} \ll \chi^2_{\rm valid}$!!)
- CORRELATIONS BETWEEN TRAINING AND VALIDATION DATA

THE SOLUTION

TUNED HYPEROPTIMIZATION



COMPARE TO A A TEST SET (NEW SET OF DATA PREVIOUSLY NOT USED AT ALL)
TESTS GENERALIZATION POWER

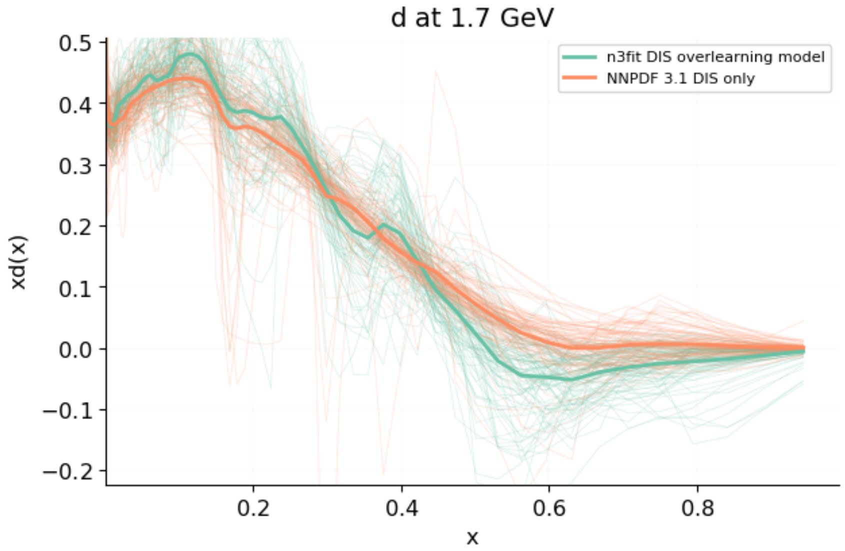
THE TEST SET METHOD

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

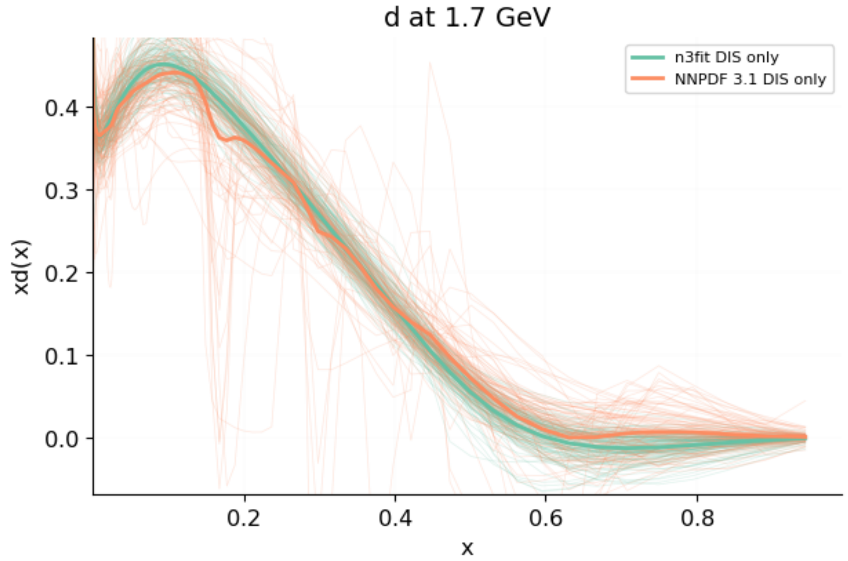
HYPEROPTIMIZED PDFs

DOWN QUARK

OVERFIT VS HANDPICKED



HYPEROPT VS HANDPICKED



- NO OVERFITTING
- COMPARED TO HANDPICKED
  - MUCH GREATER STABILITY $\Rightarrow$ FEWER REPLICAS FOR EQUAL ACCURACY
  - UNCERTAINTIES SOMEWHAT REDUCED

K-FOLDING

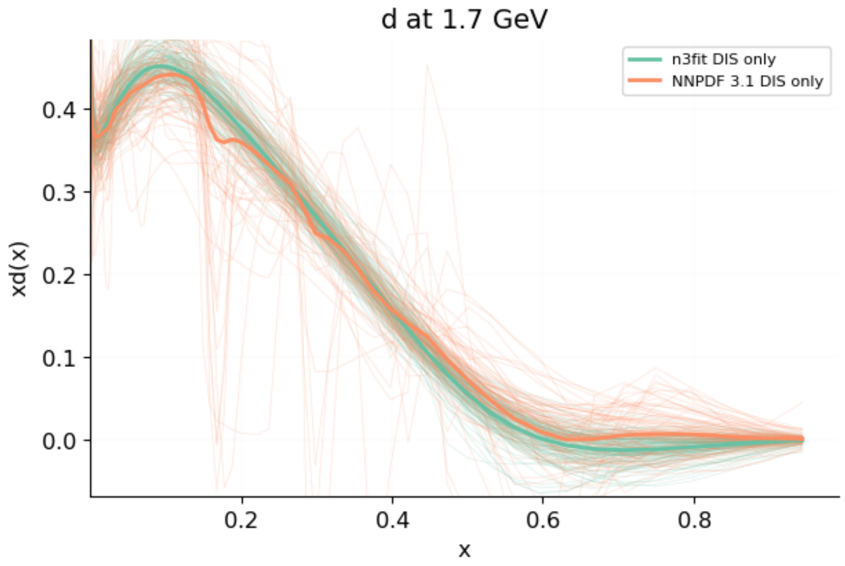
THE BASIC IDEA:

- DIVIDE THE DATA INTO n REPRESENTATIVE SUBSETS
  EACH CONTAINING PROCESS TYPES, KINEMATIC RANGE OF FULL SET
- FIT $n-1$ SETS AND USE n-TH SET AS TEST
  $\Rightarrow n$ VALUES OF $\chi^2_{\text{test, i}}$
- HYPEROPTIMIZE ON NON FITTED $\chi^2_{\text{test, i}}$
  $\rightarrow$ GOOD & STABLE GENERALIZATION

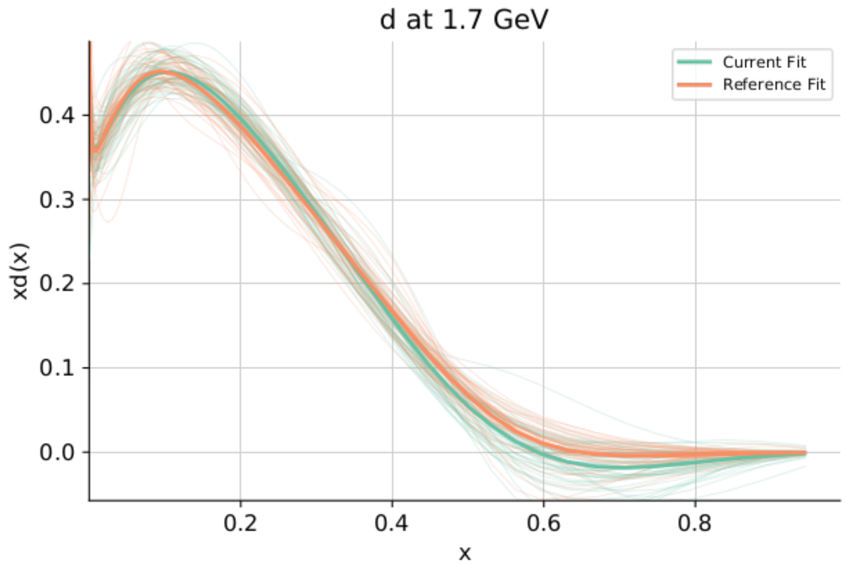
FOLDED PDFS

DOWN QUARK

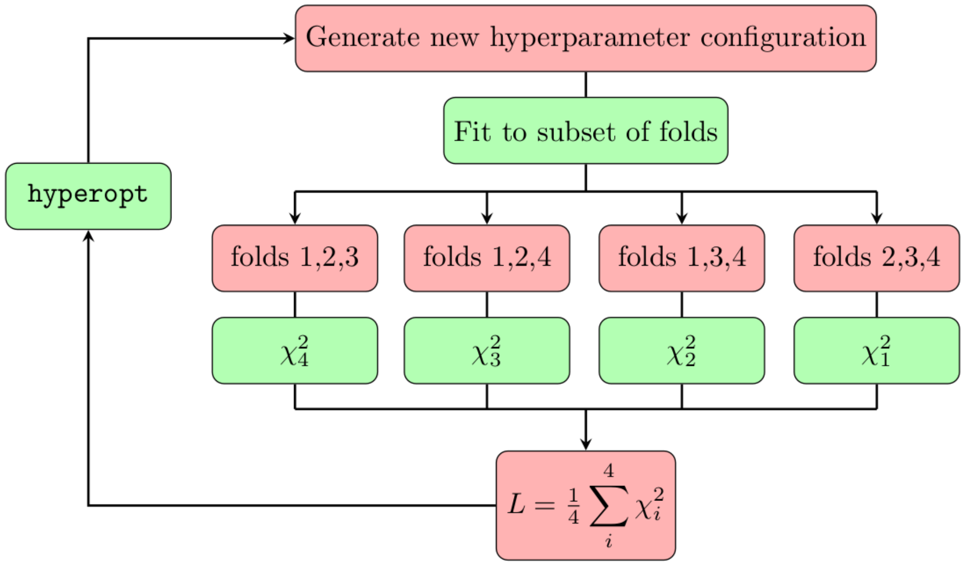
TEST-SET HYPER VS HANDPICKED



K-FOLD HYPER VS. TEST-SET HYPER



K-FOLDING IMPLEMENTATION



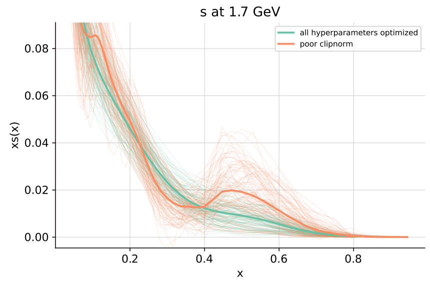
Fold 1		
CHORUS σ^{ν}_{CC}	HERA I+II inc NC e^+p 920 GeV	BCDMS p
LHCb Z 940 pb	ATLAS W, Z 7 TeV 2010	CMS Z p_T 8 TeV (p^{ll}_T, y_{ll})
DY E605 $\sigma^p_{\rm DY}$	CMS Drell-Yan 2D 7 TeV 2011	CMS 3D dijets 8 TeV
ATLAS single-$\bar{t}$ y (normalised)	ATLAS single top R_t 7 TeV	CMS $t\bar{t}$ rapidity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

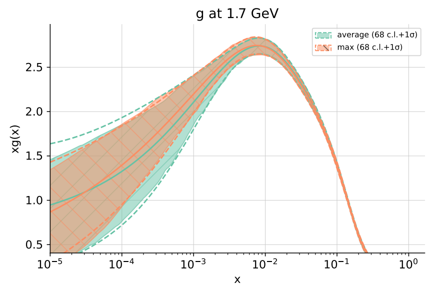
Fold 3		
HERA I+II inc CC e^+p	HERA I+II inc NC e^+p 575 GeV	NMC d/p
NuTeV σ^{ν}_c	LHCb $W, Z \to \mu$ 7 TeV	LHCb $Z \to ee$
ATLAS W, Z 7 TeV 2011 Central selection	ATLAS W^++jet 8 TeV	ATLAS HM DY 7 TeV
CMS W asymmetry 4.7 fb	DYE 866 $\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$	CDF Z rapidity (new)
ATLAS $\sigma^{\rm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

NO K-FOLDING



K-FOLDING VARIATION



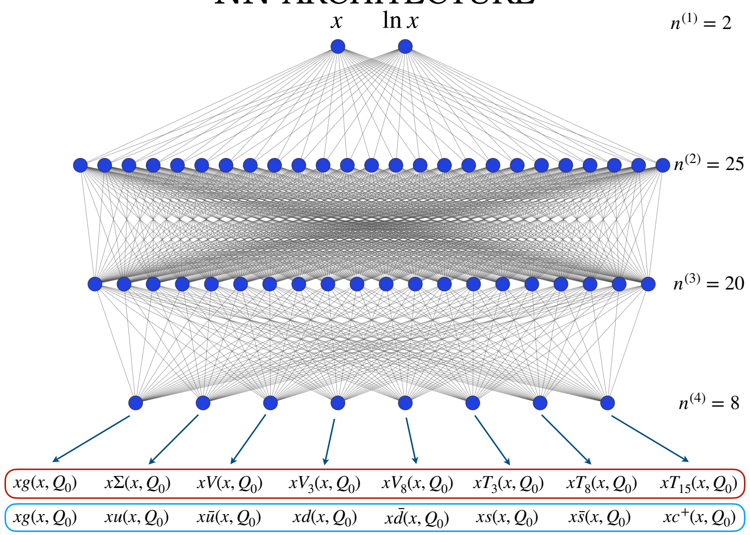
- EACH FOLD REPRODUCES FEATURES OF FULL DATASET
- DIFFERENT CHOICES POSSIBLE FOR LOSS (NON-FITTED)
  - BEST WORST
  - BEST AVERAGE
- RESULTS STABLE

THE ML METHODOLOGY

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

NN ARCHITECTURE



- HYPEROPT ADAPTS TO EXTERNAL CHOICES (E.G. PARAMETRIZATION BASIS)
- SIMILAR RESULTS CAN BE OBTAINED WITH RATHER DIFFERENT SETTINGS
- ~ 800 FREE PARAMETERS

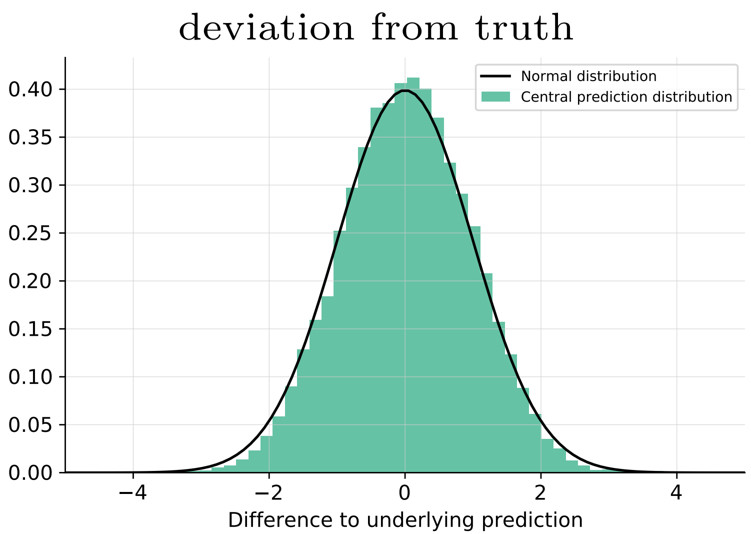
VALIDATION

CLOSURE TESTS

FAITHFUL UNCERTAINTIES IN DATA REGION?

- ASSUME "TRUE" UNDERLYING PDF $\Rightarrow$ E.G. SOME RANDOM PDF REPLICA
- GENERATE DATA DISTRIBUTED ACCORDING TO EXPERIMENTAL COVARIANCE MATRIX
- RUN WHOLE METHDOLOGY ON THESE DATA
- DO STATISTICS ON "RUNS OF THE UNIVERSE", POSSIBLE THANKS TO EFFICIENT METHDOLOGY: COMPARE TO TRUE PDFS, OR TO TRUE VALUES OF OBSERVABLES (NOT FITTED)
  - BIAS/VARIANCE: MEAN SQUARE DEVIATION WR TO TRUTH VS UNCERTAINTY
  - IS TRUTH WITHIN ONE SIGMA 68% OF TIMES?

RESULTS

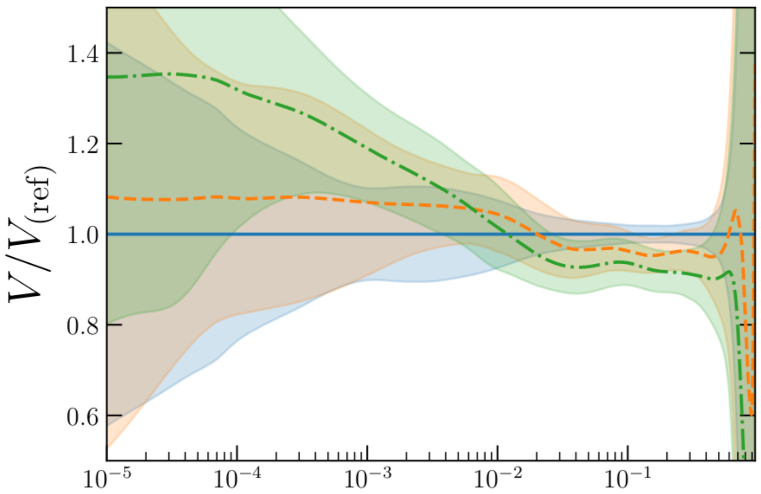


Dataset	$\sqrt{\text{bias/variance}}$	$\xi_{1\sigma}^{(\text{data})}$
DY	0.99 ± 0.08	0.69 ± 0.02
Top-pair	0.75 ± 0.06	0.75 ± 0.03
Jets	1.14 ± 0.05	0.63 ± 0.03
Dijets	0.99 ± 0.07	0.70 ± 0.03
Direct photon	0.71 ± 0.06	0.81 ± 0.03
Single top	0.87 ± 0.07	0.69 ± 0.04
Total	1.03 ± 0.05	0.68 ± 0.02

FUTURE TESTS
FAITHFUL UNCERTAINTIES IN EXTRAPOLATION?

- DETERMINE PDFs FROM A SUBSET OF CURRENT DATA: “PRE-HERA”, “PRE-LHC”,. . .
- COMPUTE χ^2 TO THE FULL CURRENT DATASET:
  - WITHOUT PDF UNCERTAINTIES $\Rightarrow$ IF $\gg 1$, MISSING INFORMATION
  - WITH PDF UNCERTAINTY $\Rightarrow$ IF ~ 1, MISSING INFO REPRODUCED BY UNCERTAINTY

VALENCE: PRE-HERA VS PRE-LHC VS NNPDF4.0



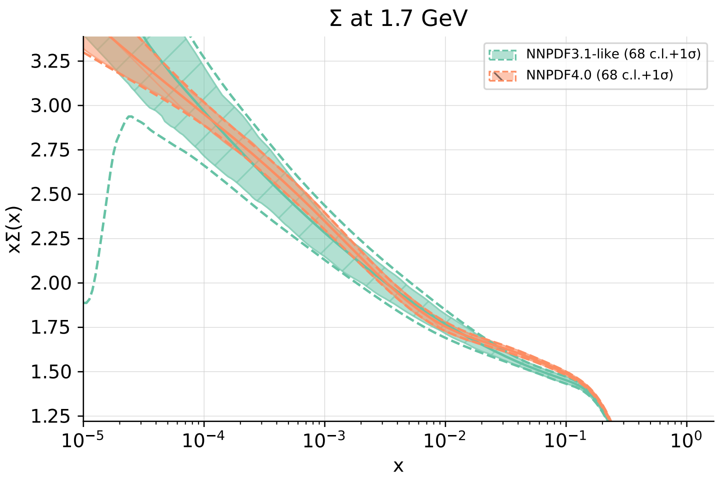
χ^2 WITHOUT/**WITH PDF UNC.**

PROCESS	PRE-HERA	PRE-LHC	3.1-LIKE	4.0-GLOB
FT DIS (NC)	1.04	1.17	1.17	1.26
FT DIS (CC)	0.80	0.86	0.88	0.90
FT DY	0.93	1.27	1.43	1.59
HERA	24.01/**1.12**	1.22	1.21	1.21
COLL. DY (TEV.)	5.31/**1.08**	0.96	0.95	1.13
COLL. DY (LHC)	15.50/**1.37**	2.64/**1.54**	1.39	1.54
TOP QUARK	23.35/**1.08**	1.29/**0.86**	0.82	0.98
JETS	6.18/**1.21**	3.66/**1.29**	2.07/**1.37**	1.26
TOTAL	9.70	1.44	1.22	1.17

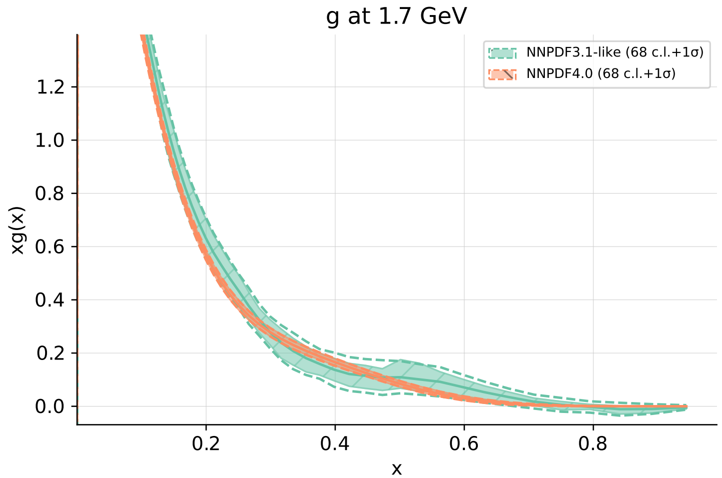
NNPDF4.0 vs. NNPDF3.1

- FULL BACKWARD COMPATIBILITY
- SUBSTANTIAL REDUCTION IN UNCERTAINTY

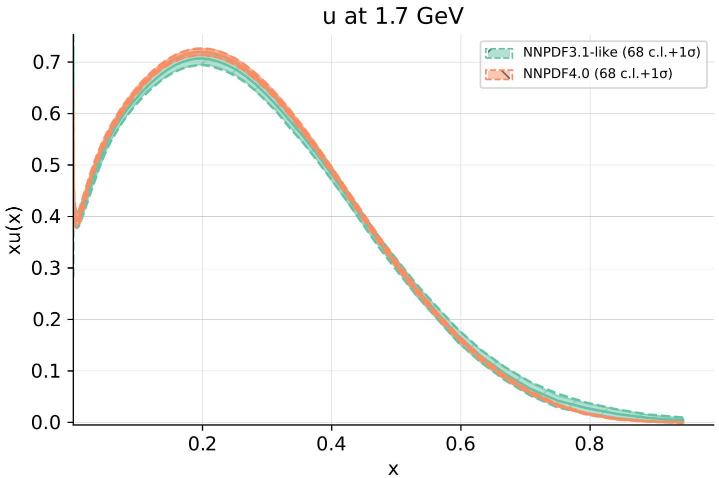
SINGLET



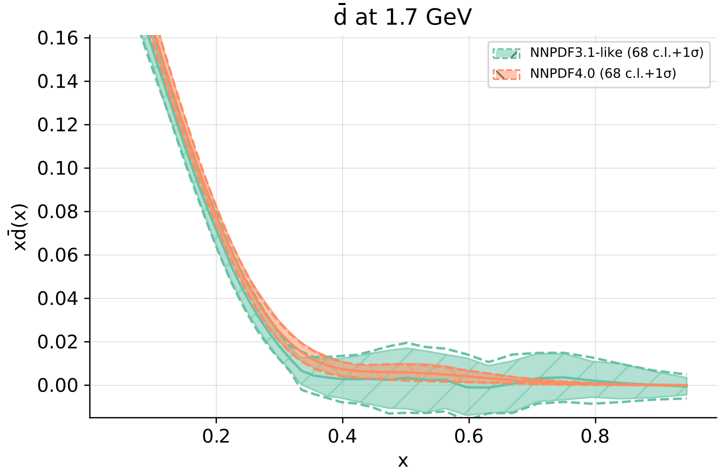
GLUON



UP



ANTIDOWN

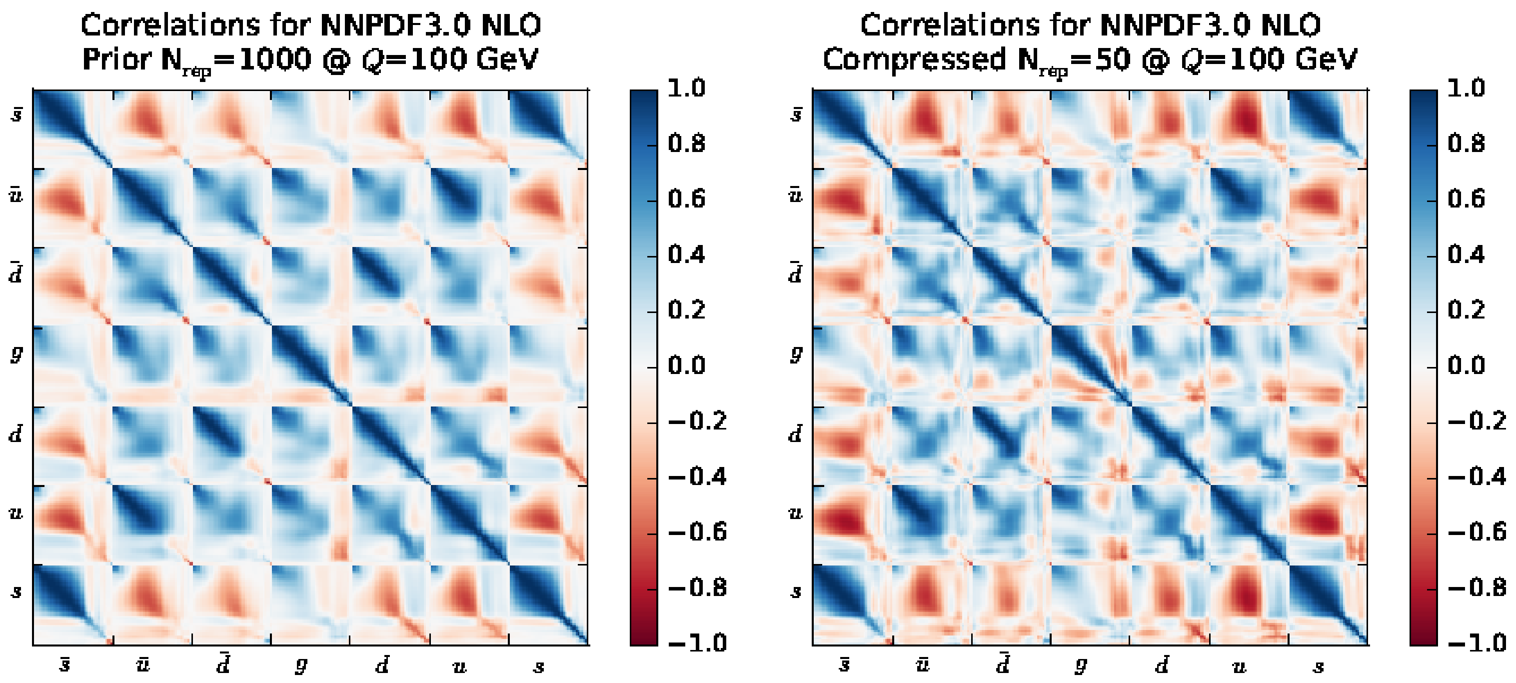


DELIVERY

MONTECARLO COMPRESSION

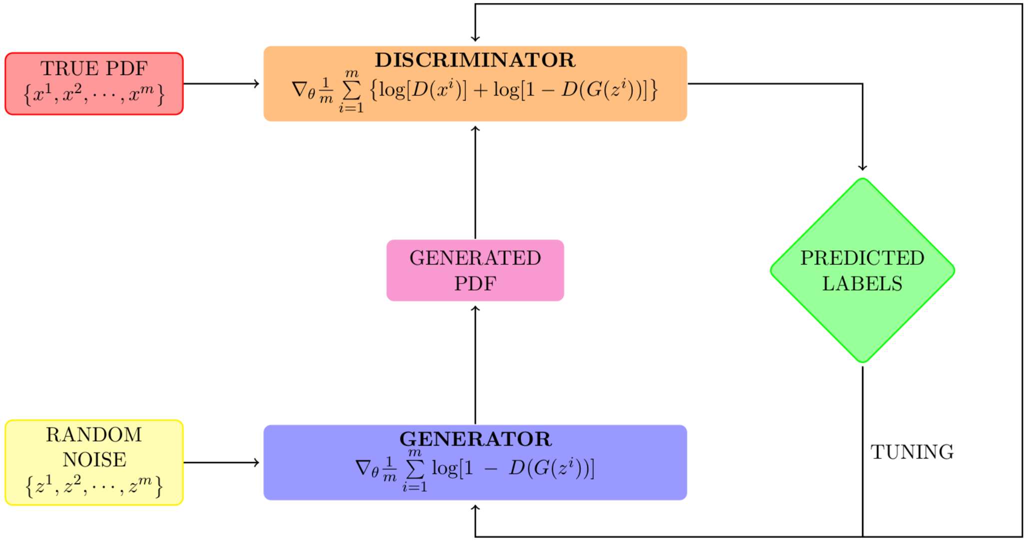
CAN WE REDUCE THE NUMBER OF REPLICAS?

- START WITH LARGE REPLICA SAMPLE
- SELECT BY GENETIC ALGORITHM SUBSET OF REPLICAS $\Rightarrow$ STATISTICAL FEATURES OPTIMIZED TO PRIOR
- MINIMIZE LOSS: DIFFERENCE OF MOMENTS, KL DIVERGENCE, ...
- 50 COMPRESSED REPLICA REPRODUCE 1000 REPLICA SET TO PRECENT ACCURACY



GAN ENHANCEMENT

CAN WE FURTHER REDUCE THE NUMBER OF COMPRESSED REPLICAS WITHOUT LOSS OF INFORMATION? GENERATIVE ADVERSARIAL NETWORKS

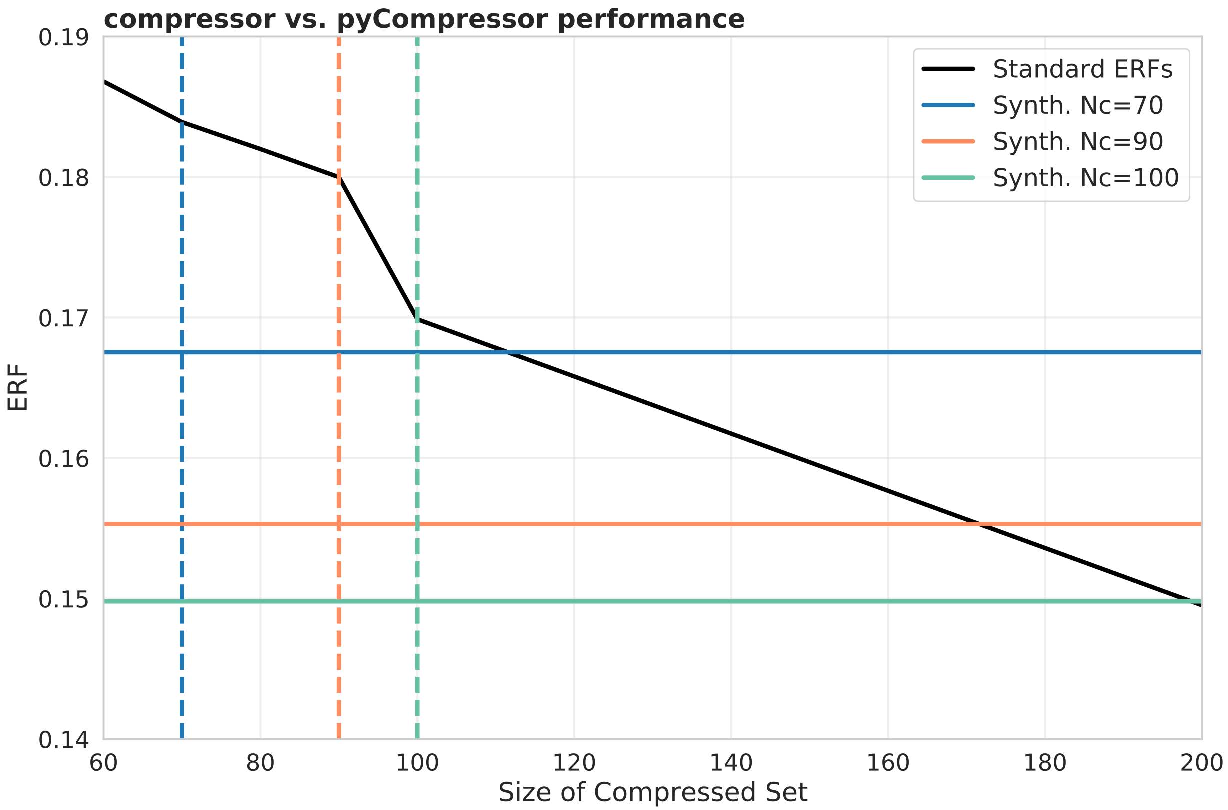


- TRAIN A NETWORK TO SIMULATE THE TRUE DISTRIBUTION (GENERATOR)
- TRAIN A NETWORK TO DISCRIMINATE TRUTH FROM SIMULATION (DISCRIMINATOR)
- TRAIN THE GENERATOR TO TRICK THE DISCRIMINATOR

GAN ENHANCEMENT

- ENHANCE THE STARTING PDF SET BY ADDING GAN-PDFS TO IT
- PERFORM COMPRESSION OF THE ENHANCED SET

PERFORMANCE



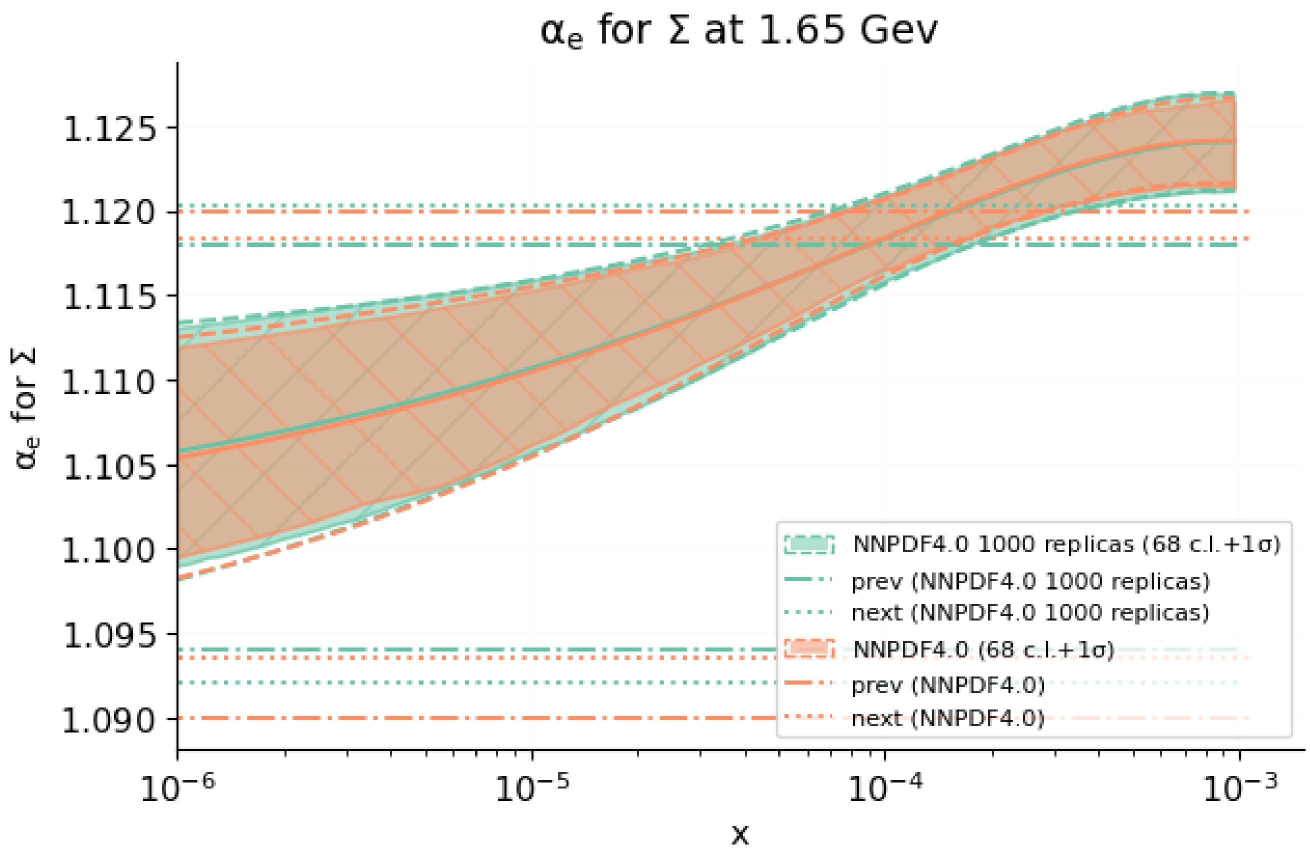
ENHANCED: NUMBER OF REPLICAS CUT IN HALF FOR SAME TARGET ACCURACY

NEW IDEAS

PREPROCESSING

- PDF EQUAL TO PREPROCESSED NN: $f_i = x^{\alpha_i}(1-x)_i^{\beta} NN(i, x, \ln x)$
- PREPROCESSING EXPONENTS α_i, β_i VARIED RANDOMLY REPLICA BY REPLICA
- RANGE OF VARIATION DETERMINED SELF-CONSISTENTLY
- PDF TAKES x, $\ln x$ AS INPUTS

QUARK SINGLET EFFECTIVE α

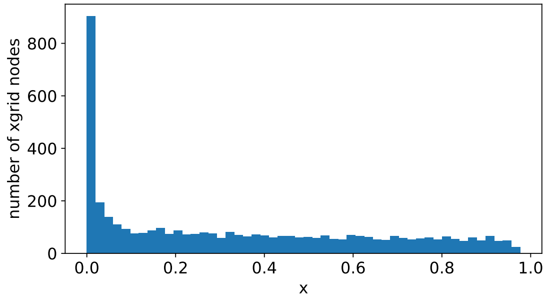


- NEED TO ITERATE FITS
- POSSIBLE SOURCE OF BIAS?

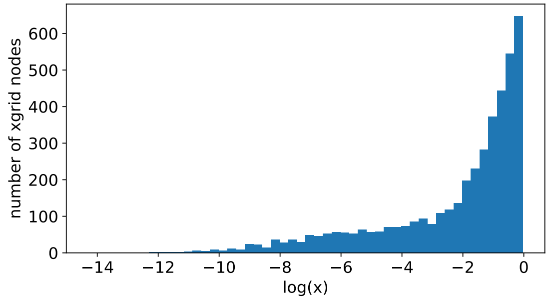
FEATURE SCALING

INPUT NODE DISTRIBUTION

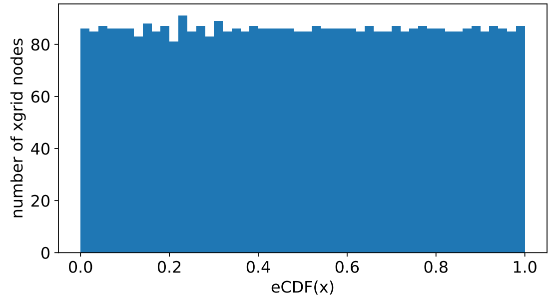
x DISTN



$\ln x$ DISTN



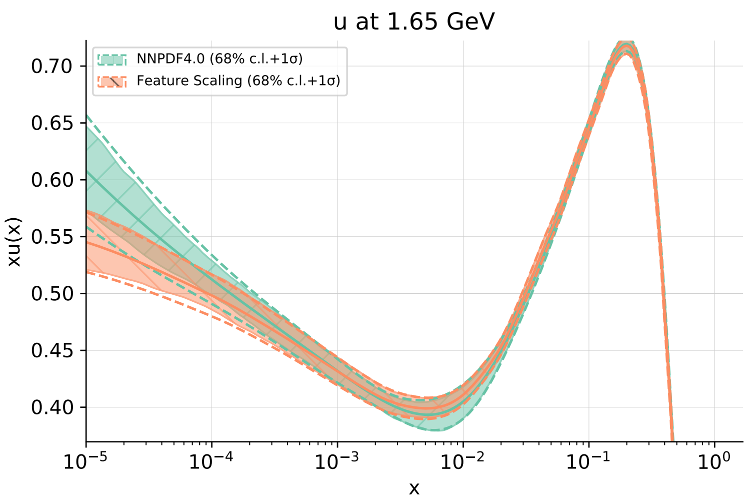
eCDF DISTN



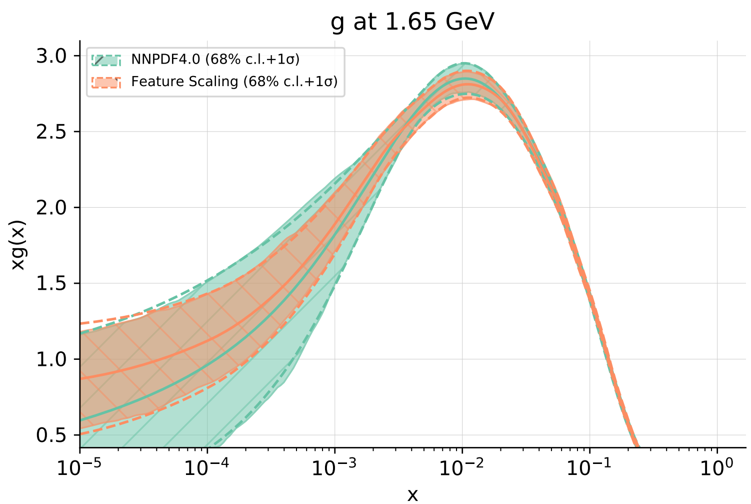
- RESCALE $\Rightarrow$ UNIFORMLY DISTRIBUTED INPUT (eCDF+INTERPOLATION)
- RERUN HYPEROPT
- ONLY ONE INPUT NEEDED

STANDARD VS. FEATURE SCALING

UP



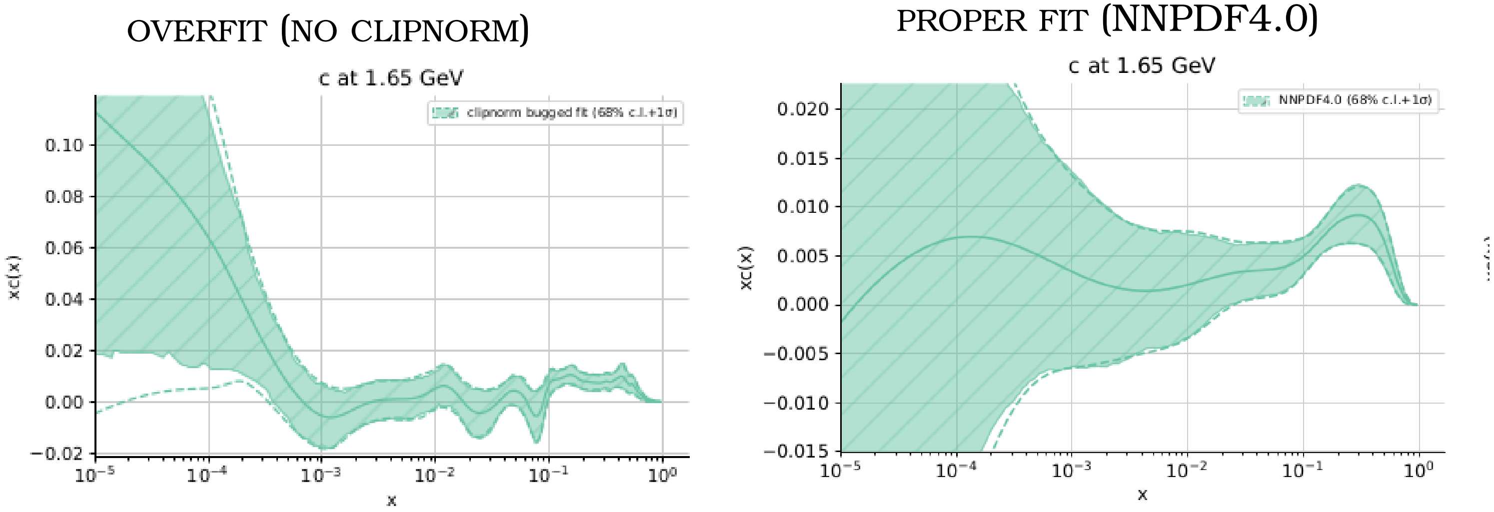
GLUON



OVERFITTING IN HYPEROPT

- HYPEROPT $\Rightarrow$ OVERFITTING?
- WHAT IS A "NATURAL" PDF?

CHARM PDF

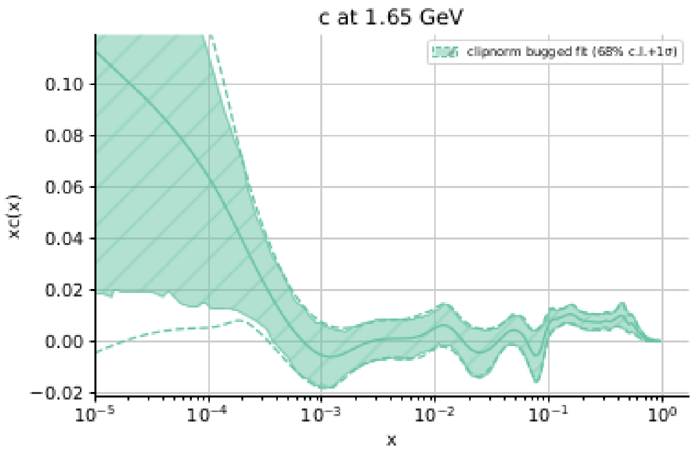


OVERFITTING METRIC

- RECOMPUTE VALIDATION χ^2
  - SAME TRAINING-VALIDATION SPLIT
  - DIFFERENT FLUCTUATED VALIDATION DATA
- COMPUTE AVERAGE χ^2 & DETERMINE DIFFERENCE TO VALIDATION $\mathcal{R}_O = \langle \chi^2_{\mathrm{val}} - \chi^2_{\mathrm{val}'} \rangle$ OVERFITNESS
- NEGATIVE OVERFITNESS $\mathcal{R}_O \Rightarrow$ OVERFIT

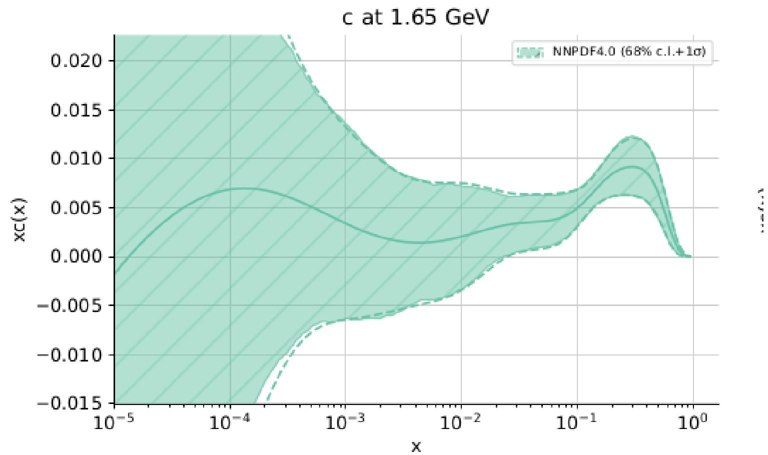
CHARM PDF

OVERFIT (NO CLIPNORM)



$$\mathcal{R}_O = -0.024 \pm 0.012$$

PROPER FIT (NNPDF4.0)



$$\mathcal{R}_O = -0.001 \pm 0.013$$

FOOD FOR THOUGHT

- OPTIMIZED FOLDS
- OVERFITTING AND UNDERFITTING METRICS IN HYPEROPT
- HYPEROPT ON FUTURE TESTS
- BEYOND NEURAL NETS

AN OPEN SOURCE CODE!

- THE FULL NNPDF CODE IS PUBLIC!
- INCLUDING HYPEROPTIMIZATION, EVOLUTION, THEORY, FITTING, VISUALIZATION
- FULLY DOCUMENTED CODE
- LINKS TO CODE (GITHUB), DOCUMENENTATION, INSTALLATION BINARY PACKAGES AVAILABLE FROM
`http://nnpdf.mi.infn.it/nnpdf-open-source-code/`

`https://arxiv.org/abs/2109.02671`

An open-source machine learning framework for global analyses of parton distributions

The NNPDF Collaboration: Richard D. Ball · Stefano Carrazza · Juan Cruz-Martinez · Luigi Del Debbio · Stefano Forte · Tommaso Giani · Shayan Iranipour · Zahari Kassabov · Jose I. Latorre · Emanuele R. Nocera · Rosalyn L. Pearson · Juan Rojo · Roy Stegeman · Christopher Schwan · Maria Ubiali · Cameron Voisey · Michael Wilson

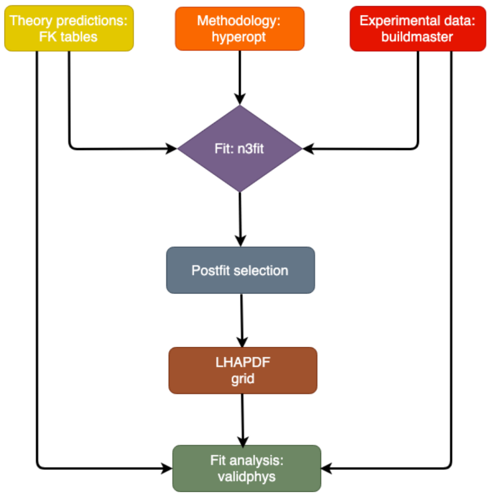


Fig. 2.1. Workflow for an NNPDF fit