NNPDF

MACHINE LEARNING

AND

PARTON DISTRIBUTIONS

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

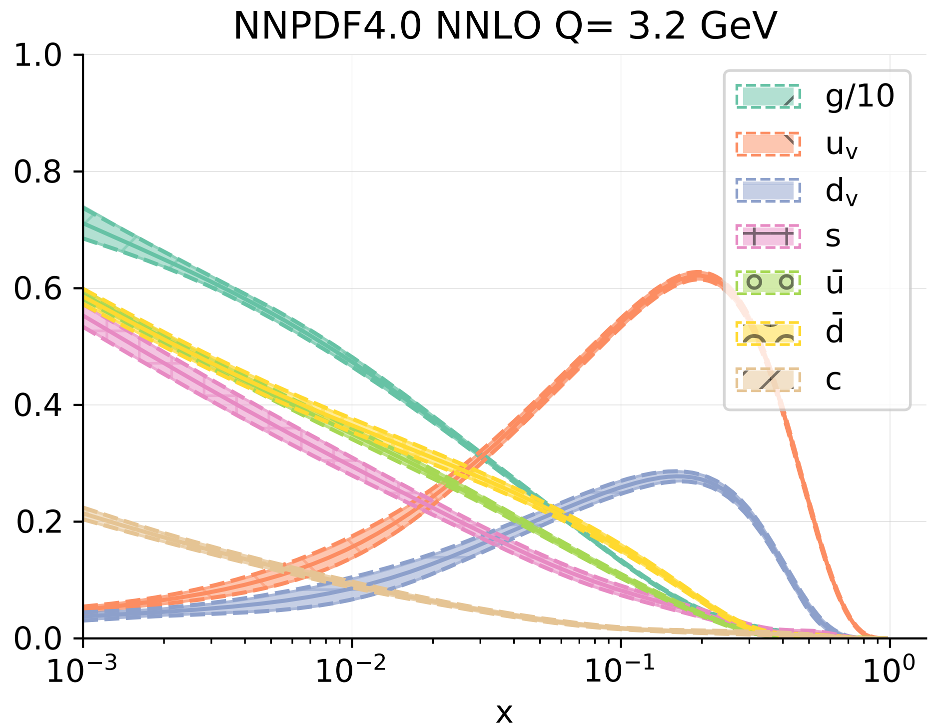
UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

Istituto Nazionale di Fisica Nucleare

PRESENT AND FUTURE PERSPECTIVES IN HADRON PHYSICS

FRASCATI, JUNE 19, 2024

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



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

WHY MACHINE LEARNING?

WHY WE NEED MACHINE LEARNING I

ALTERNATIVE: A MODEL-DEPENDENT APPROACH

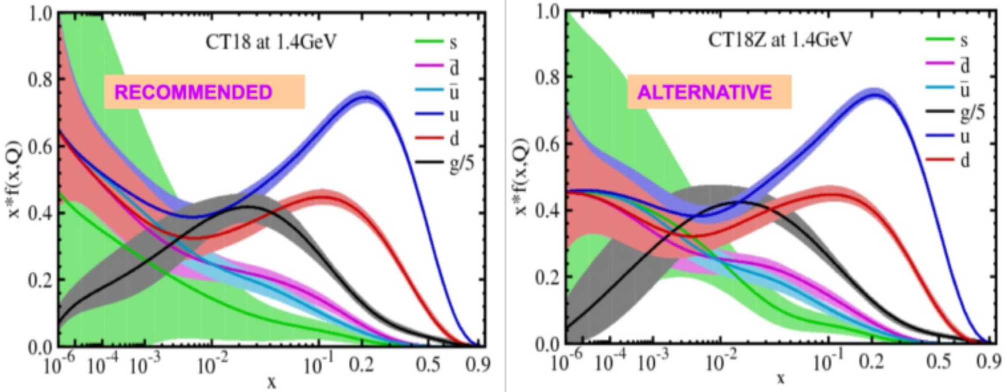
PARAMETRIZATIONS

- 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: $g(x, Q = Q_0) = x^{a_1 - 1}(1-x)^{a_2}\left[a_3(1-y)^3 + a_4 3y(1-y)^2 + a_5 3y^2(1-y) + y^3\right]$;
  $y = \sqrt{x}$; $a_5 = (3 + 2a_1)/3$.

MORE DATA $\Rightarrow$ BIGGER PARAMETRIZATION (?)

PROLIFERATION OF PDF SETS

Four PDF ensembles: CT18 (default), A, X, and Z



- The CT18 family of PDFs includes LHC data available up to 2018, i.e. mostly 7 and 8 TeV data
- CT18 is the primary PDF; CT18A includes the ATLAS 7 TeV W/Z data (excluded from CT18 due to very poor fit); CT18X includes scale to simulate effects of low x resummation for DIS; CT18Z includes both effects
- CT18As (new) allows a more flexible parametrization for strange
- CT18As_Lat (new) adds lattice constraint

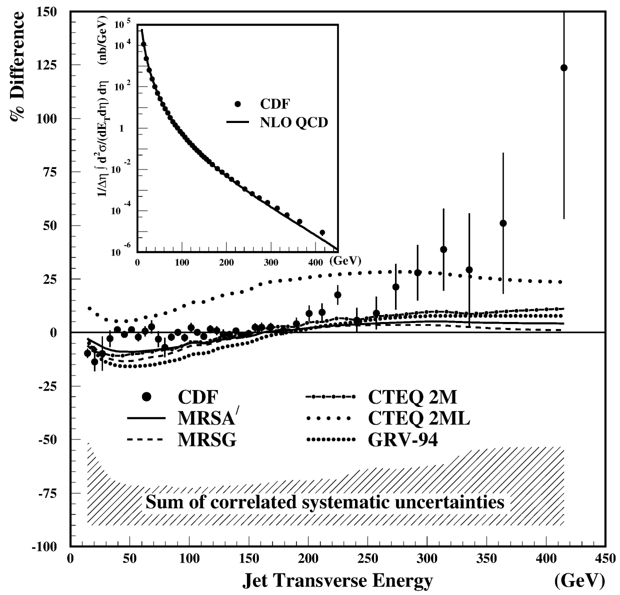
(J. Huston, PDF4LHC 11/2023)

MORE DATA $\Rightarrow$ BIGGER UNCERTAINTIES (!)

WHY WE NEED MACHINE LEARNING II

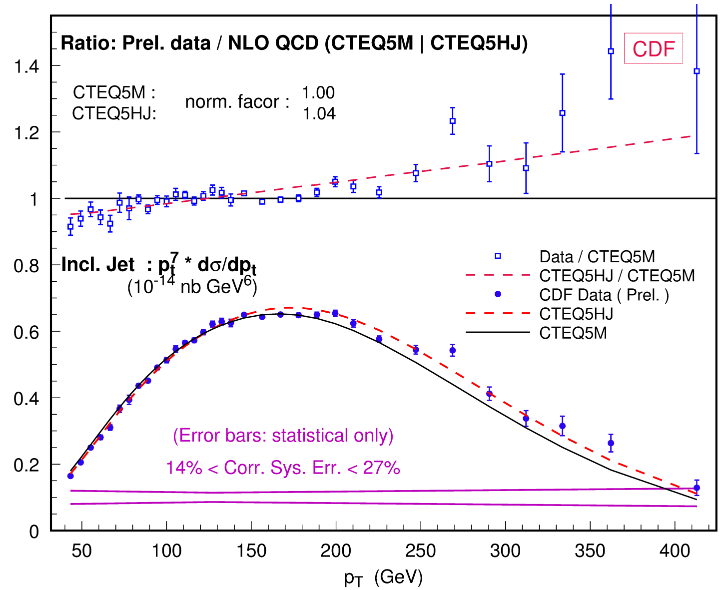
DISCOVERY PHYSICS 1995

- 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 $x \gtrsim 0.1$

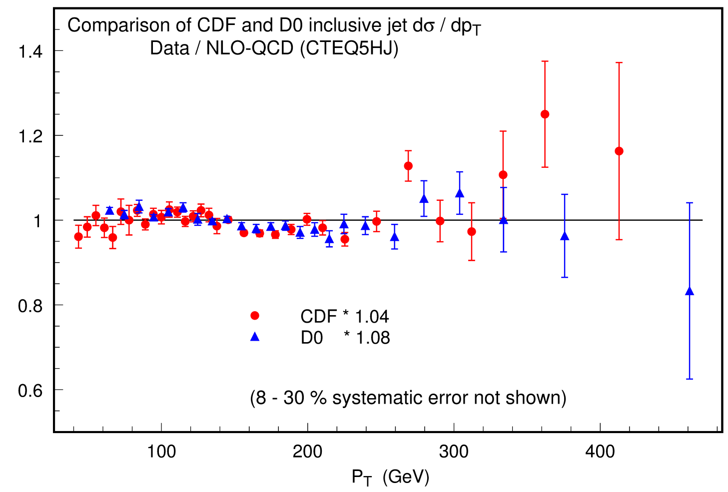


DISCREPANCY REMOVED IF JET DATA INCLUDED IN THE FIT

NEW CTEQ FIT (1996)



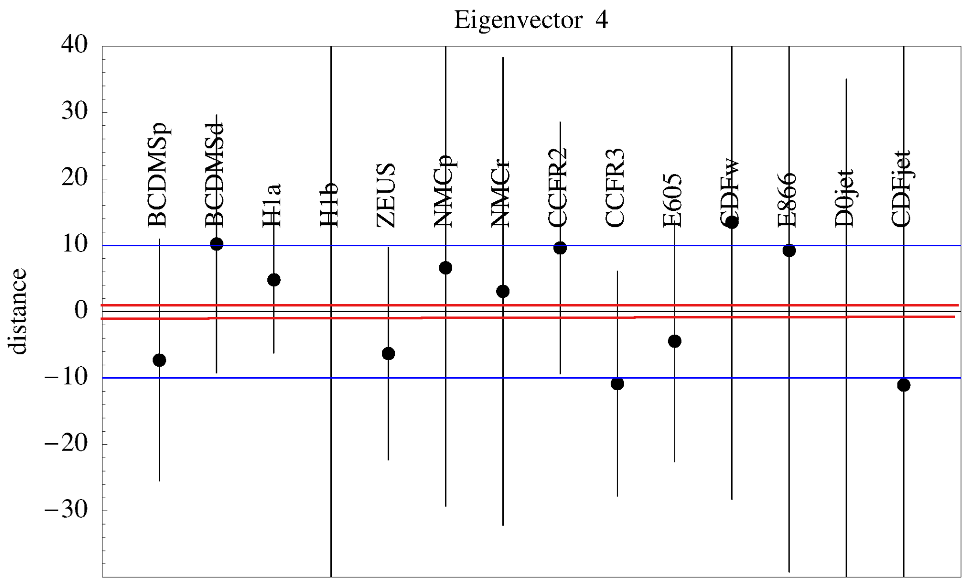
FINAL CTEQ FIT (1998)



WHY WE NEED MACHINE LEARNING III
“TOLERANCE”

FIRST PDFs WITH UNCERTAINTIES (2002)

one sigma & ten sigma intervals for typical
covariance matrix eigenvalue
vs best value and uncertainty from individual experiments



MSHT PDFS (2020)

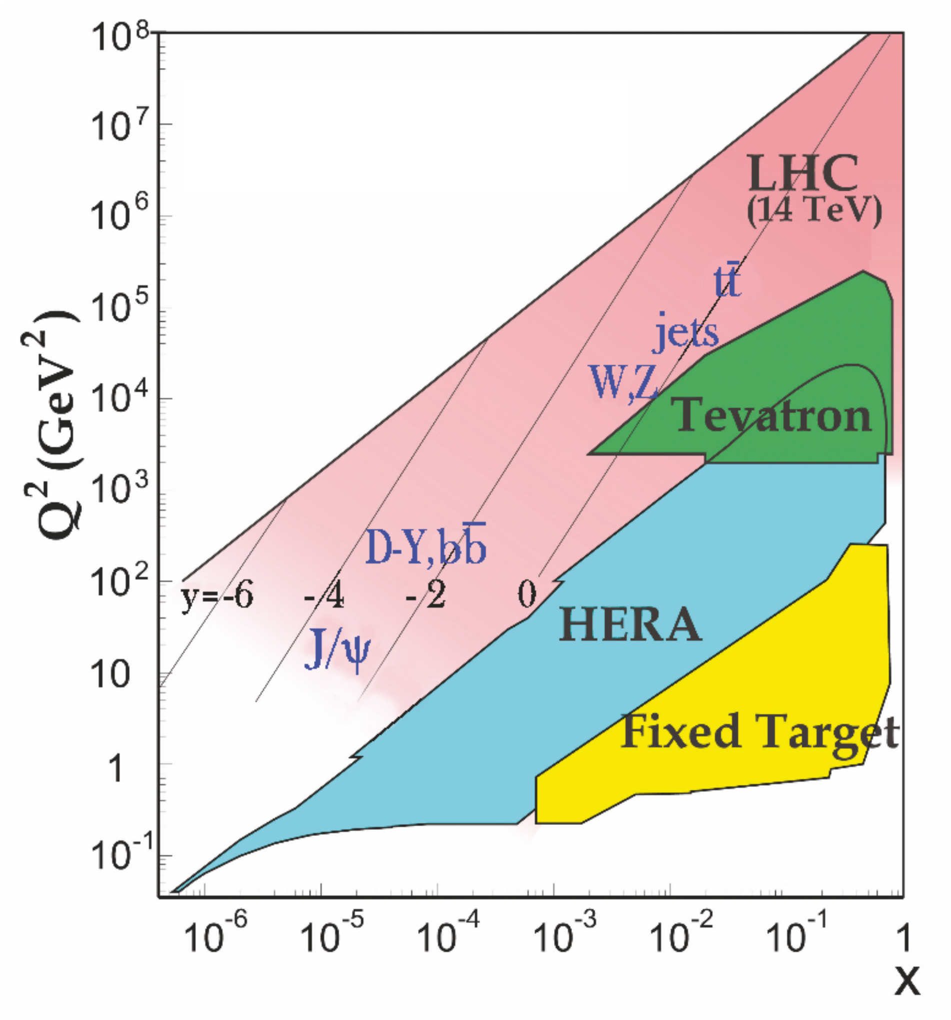
e− vector	+ t	+ T	Most constraining data set	− t
1	3.71	3.75	ATLAS 7 TeV high prec. W,Z	4.76
2	3.12	3.33	NuTeV $\nu N \to \mu\mu X$	2.85
3	2.48	2.58	NuTeV $\nu N \to \mu\mu X$	4.07
4	3.61	3.60	CMS 8 TeV W	2.93
5	2.64	3.00	ATLAS 7 TeV high prec. W,Z	2.72
6	5.22	5.46	ATLAS 8 TeV double dif Z	5.01
7	4.07	4.37	NMC/... F_L	2.90
8	3.90	3.50	LHCb 2015 W,Z	3.90
9	5.48	5.59	LHCb 2015 W,Z	3.73
10	3.55	3.58	BCDMS μp F_2	4.87
11	3.06	2.91	DØ W asym.	4.83
12	1.42	1.71	DØ W asym.	3.40
13	3.87	4.10	CMS asym. $p_T > 25, 30$ GeV	4.38
14	1.36	1.50	E866/NuSea pd/pp DY	3.67
15	5.53	5.89	E866/NuSea pd/pp DY	3.17
16	1.89	0.52	E866/NuSea pd/pp DY	5.64
17	2.51	2.54	E866/NuSea pd/pp DY	2.69
18	1.80	1.88	DØ W asym.	2.47
19	2.47	2.18	CMS 8 TeV W	1.37
20	1.82	2.22	DØ W asym.	4.69
21	4.41	5.36	ATLAS 8 TeV Z p_T	4.68
22	3.49	3.23	DØ W asym.	3.04
23	1.84	2.43	ATLAS 8TeV sing dif $t\bar{t}$ dilep	4.96
24	0.99	1.23	E866/NuSea pd/pp DY	4.61
25	2.01	1.35	DØ W asym.	2.77
26	2.25	2.51	NuTeV νN xF_3	2.06
27	2.83	3.65	ATLAS 8 TeV $t\bar{t}$, dilepton	2.64
28	1.74	1.92	DØ W asym.	2.65
29	2.57	2.85	CMS 7 TeV $W + c$	1.79
30	4.76	3.92	CCFR $\nu N \to \mu\mu X$	2.25
31	2.79	4.81	ATLAS 7TeV high prec W,Z	2.07
32	2.57	4.27	CCFR $\nu N \to \mu\mu X$	2.58

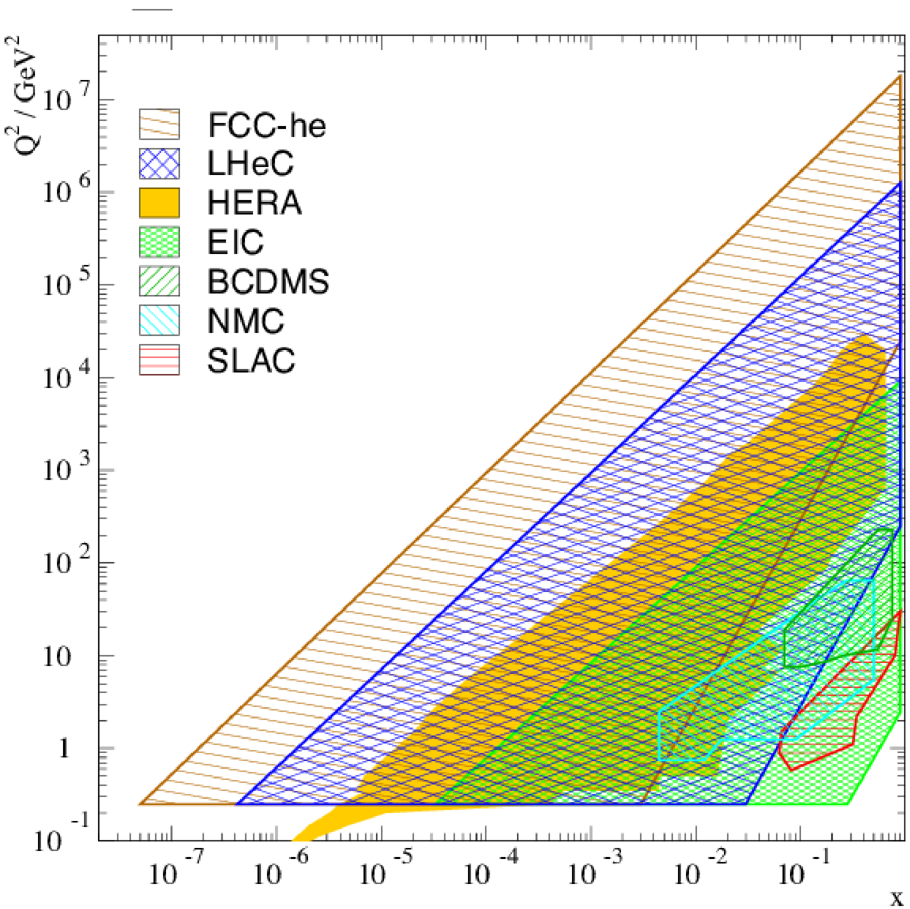
- PDF UNCERTAINTIES RESCALED BY “TOLERANCE” $T \sim 4 \div 10$
- DETERMINED FROM SPREAD OF BEST-FIT FROM DIFFERENT DATA

WHERE IT COMES ALL TOGETHER:
THE DIS-HADRON COLLIDER SYNERGY THE KINEMATIC PLANE

DIS

LHC

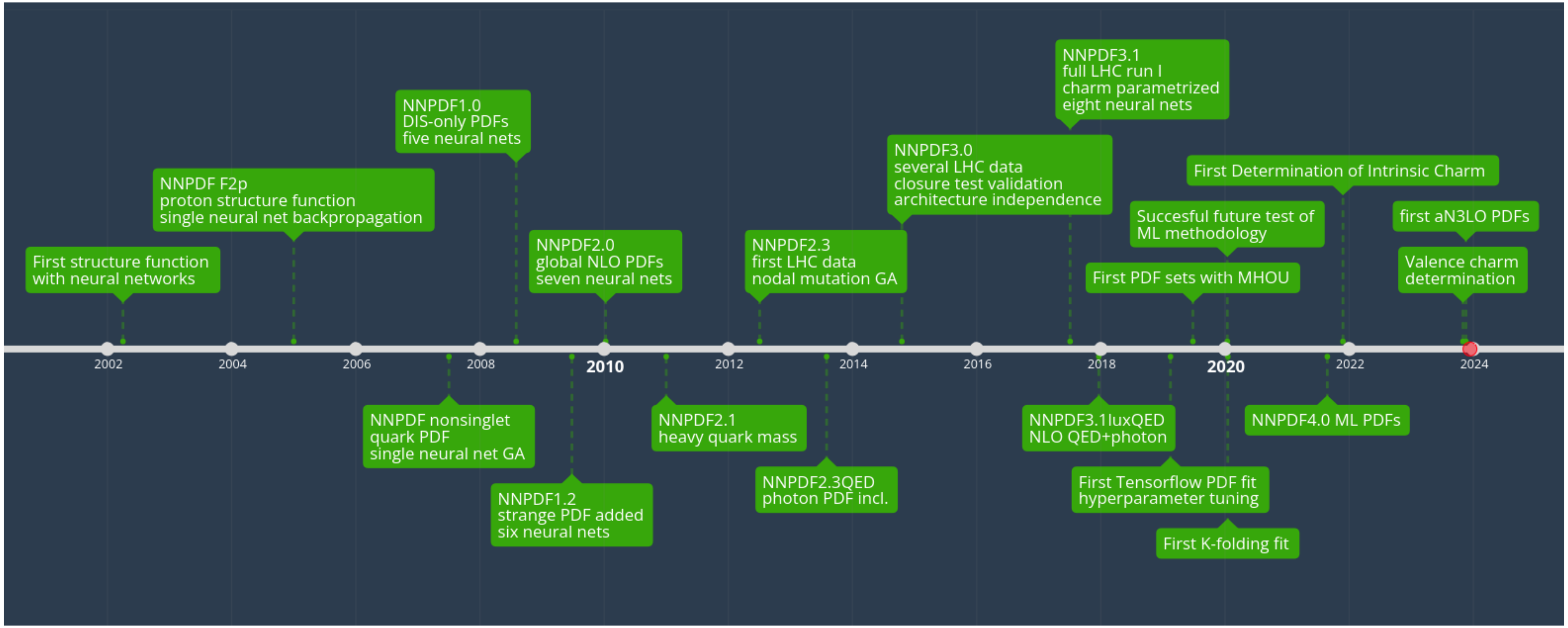




- PDFs LARGELY UNKNOWN IN LARGE x REGION
- DISCOVERY REGION AT HADRON COLLIDERS

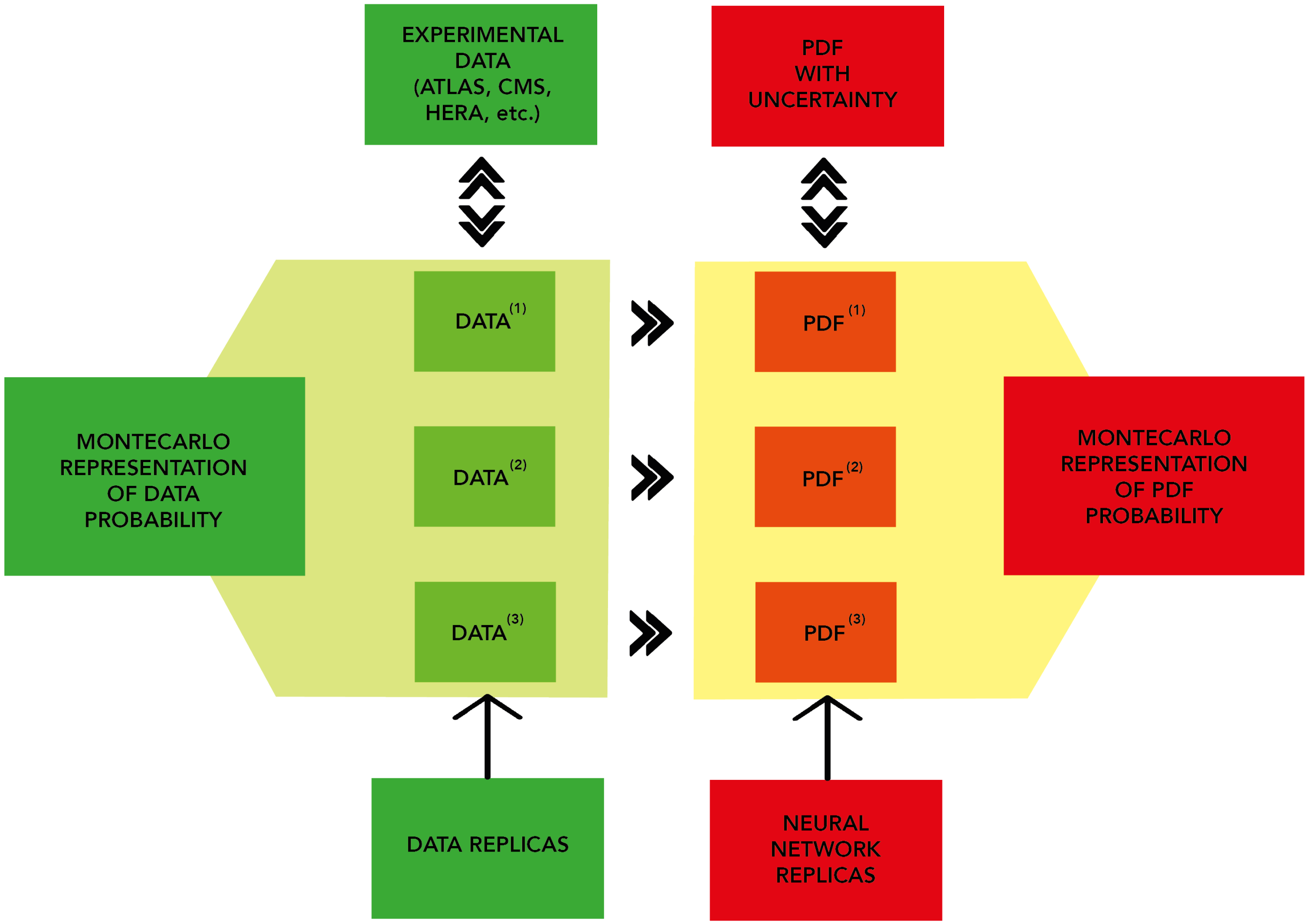
WHICH MACHINE LEARNING?

PROTON STRUCTURE AS A ML PROBLEM
NNPDF



PROBABILITY REGRESSION

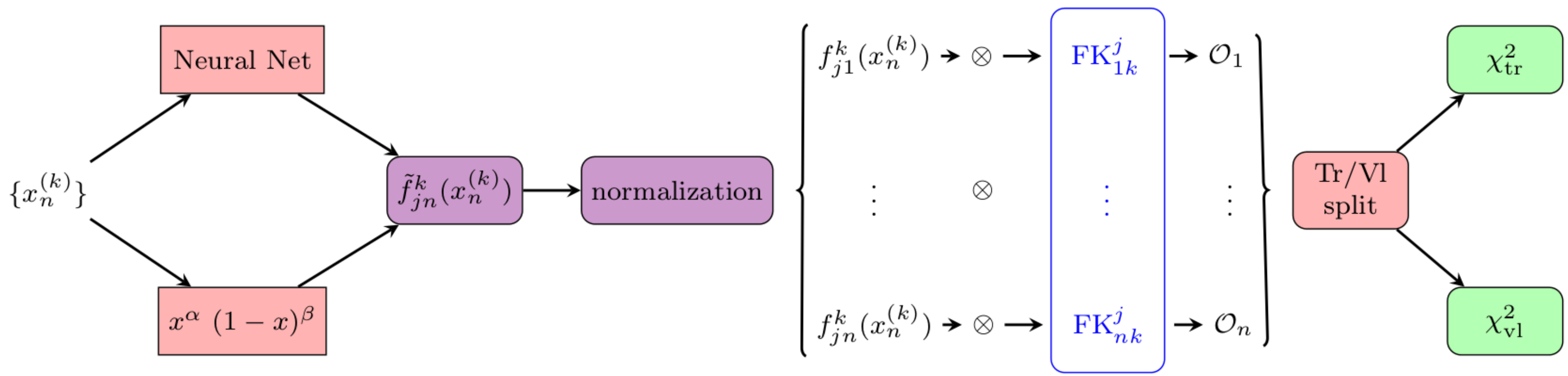
REPLICA SAMPLE OF FUNCTIONS $\Leftrightarrow$ PROBABILITY DENSITY IN FUNCTION SPACE
KNOWLEDGE OF LIKELIHHOD SHAPE (FUNCTIONAL FORM) NOT NECESSARY

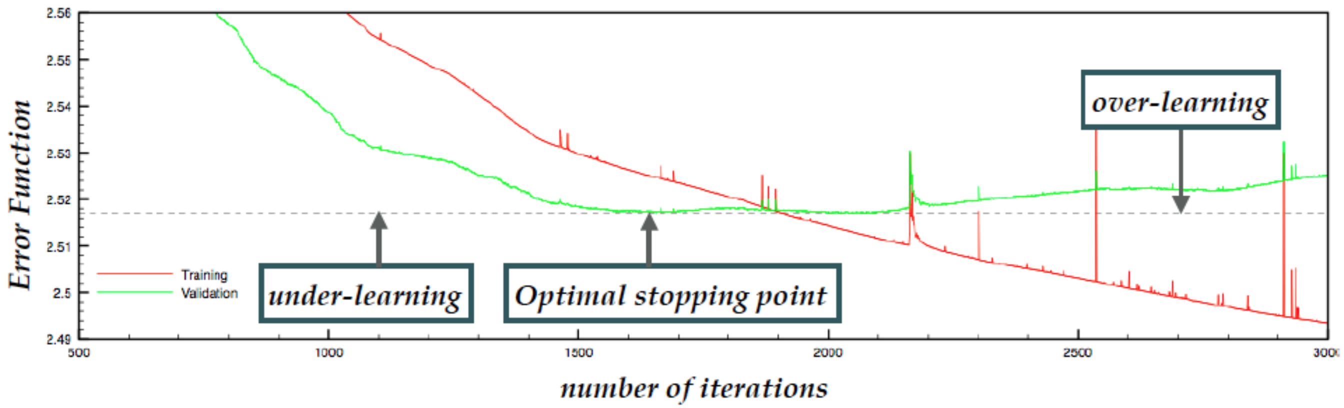


FINAL PDF SET: $f_i^{(a)}(x,\mu)$;
i =up, antiup, down, antidown, strange, antistrange, charm, gluon; $j = 1, 2, \ldots N_{\rm rep}$

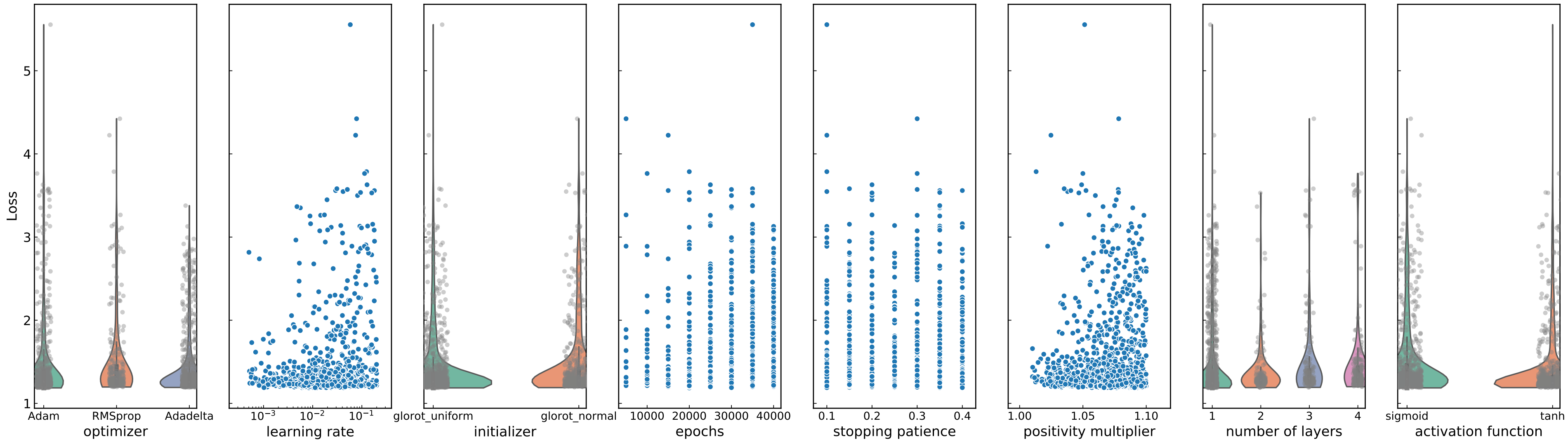
CROSS-VALIDATED LEARNING

- 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 $\chi^2 \Rightarrow$ OPTIMAL FIT





METHODOLOGY HYPEROPTIMIZATION

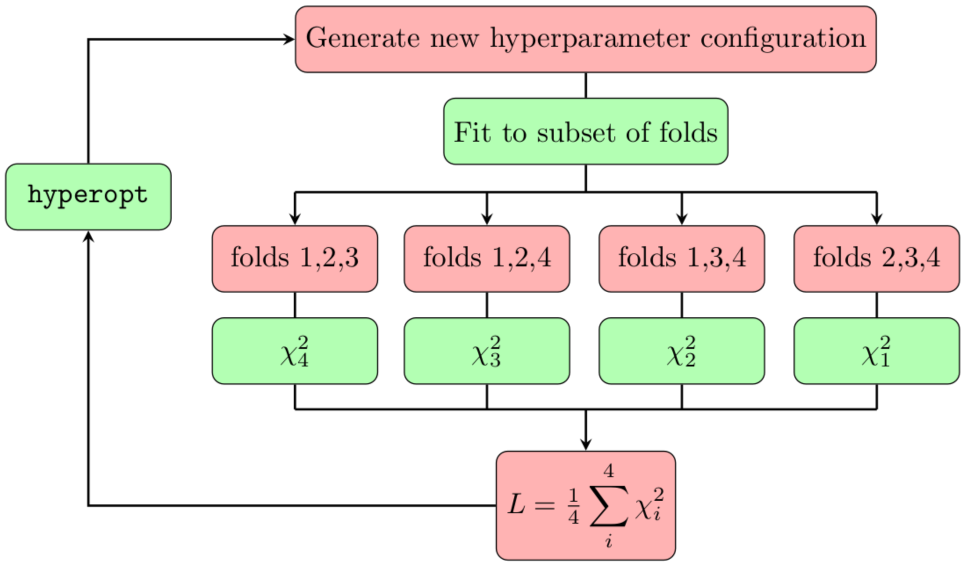


HYPEROPT PARAMETERS

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-FOLD OPTIMIZATION



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_T^{ll}, y_{ll})
DY E605 σ^p_{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		

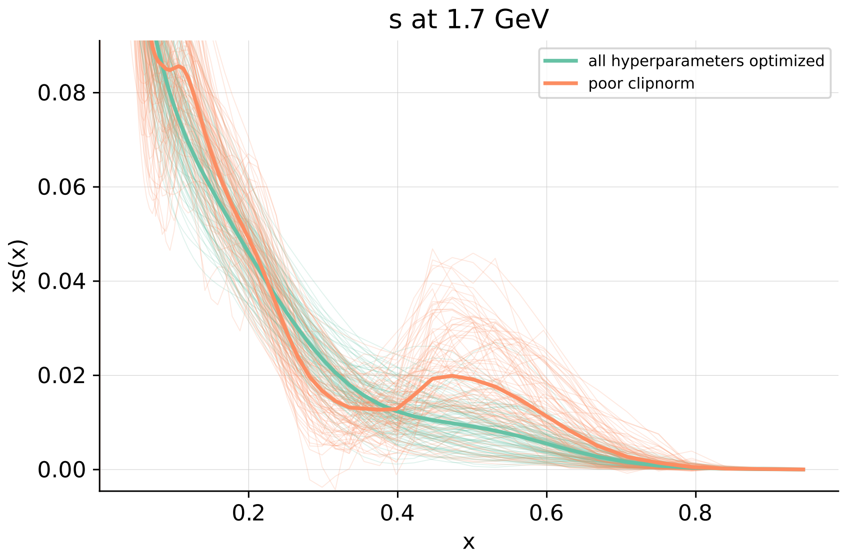
Fold 2		
HERA I+II inc CC e^-p	HERA I+II inc NC e^+p 460 GeV	HERA comb. $\sigma^{\mathrm{red}}_{b\bar{b}}$
NMC p	NuTeV $\sigma^{\bar{\nu}}_c$	LHCb $Z \to ee$ 2 fb
CMS W asymmetry 840 pb	ATLAS Z p_T 8 TeV (p_T^{ll}, M_{ll})	D0 $W \to \mu\nu$ asymmetry
DY E886 σ^p_{DY}	ATLAS direct photon 13 TeV	ATLAS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

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

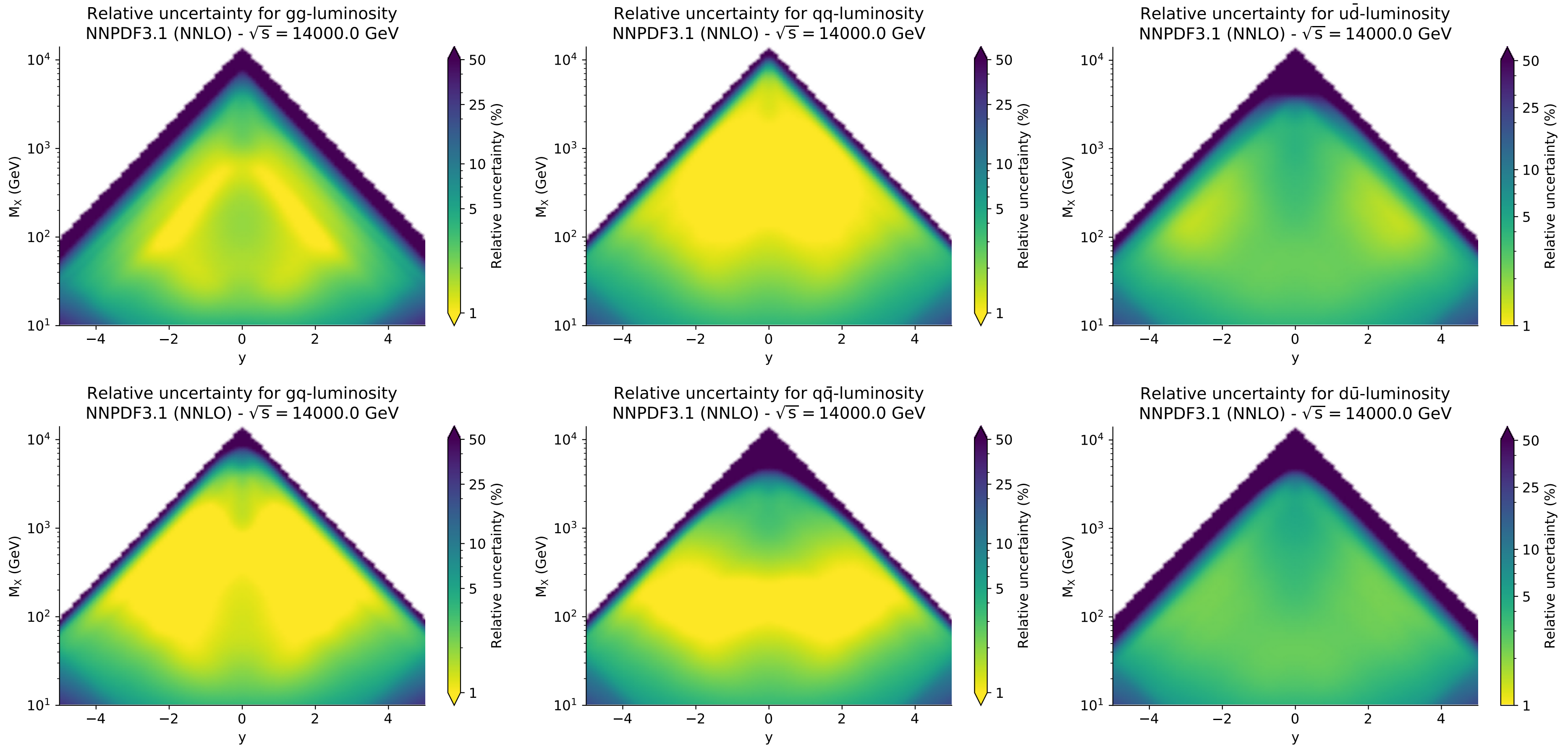
K-FOLDING VS NO K-FOLDING



WHAT IS ML GOOD FOR?

WHAT DOES ML BUY US?: PRECISION UNCERTAINTIES 2016

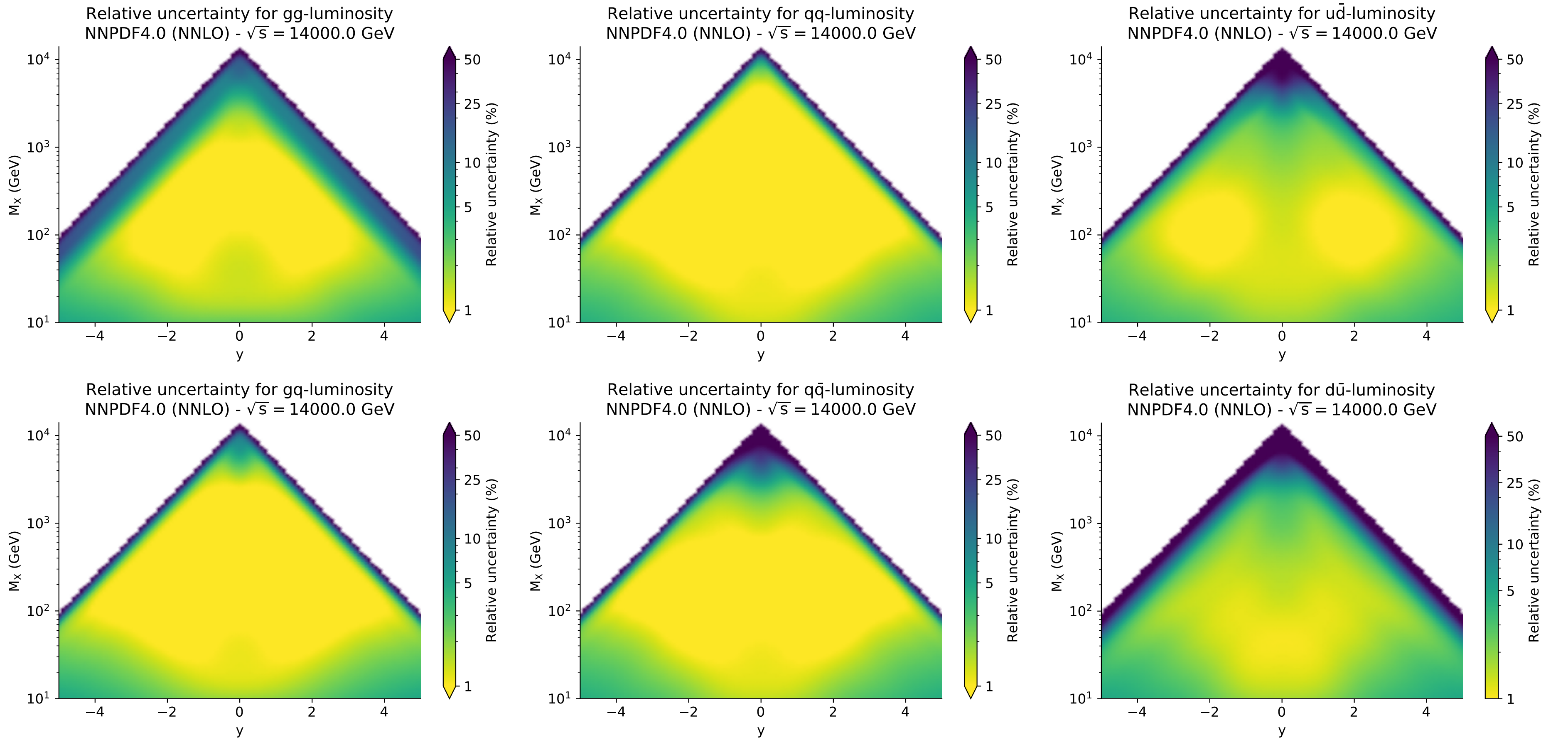
GLUON SINGLET FLAVORS



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

WHAT DOES ML BUY US?: PRECISION UNCERTAINTIES 2022

GLUON SINGLET FLAVORS

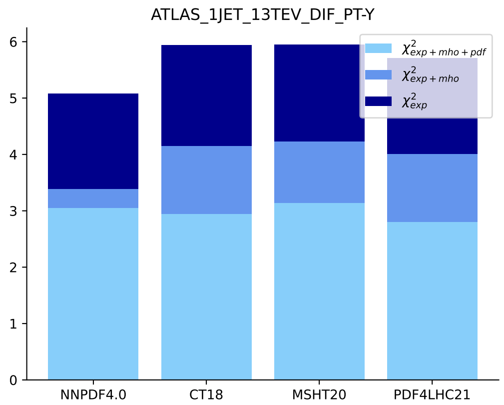


- TYPICAL UNCERTAINTIES IN DATA REGION: SINGLET $\sim 1\%$, NONSINGLET $\sim 2-3\%$
- DATA REGION: $10 \lesssim M_X \lesssim 3 \cdot 10^3$ TeV, $-4 \lesssim y \lesssim 4$

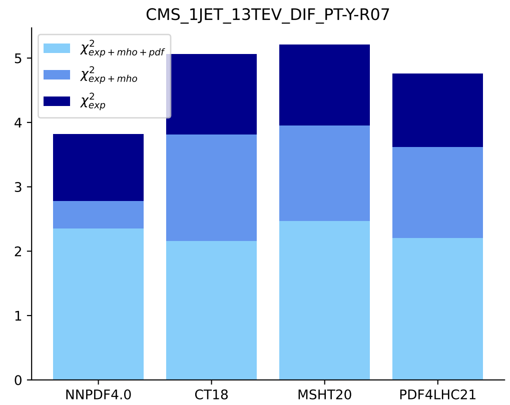
WHAT DOES ML BUY US?: ACCURACY

- COMPARISON TO DATA PUBLISHED AFTER PUBLICATION OF NNPDF4.0
- χ^2 WITH EXP, EXP+TH AND TOTAL (EXP+TH+PDF) UNCERTAINTIES

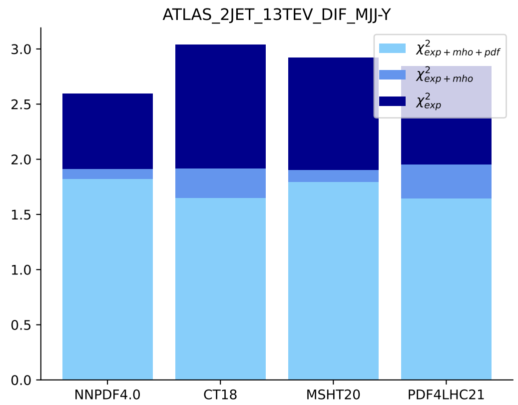
ATLAS JETS



CMS JETS



ATLAS DIJETS

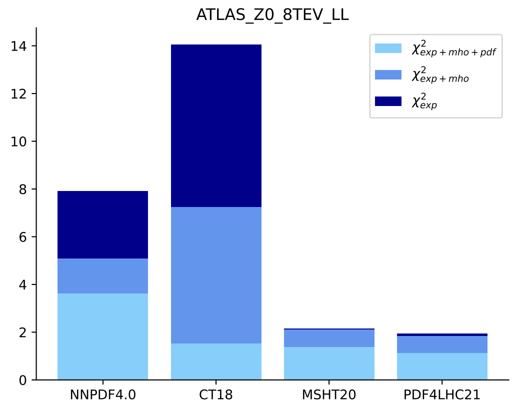


- EXP χ^2 LOWER $\Rightarrow$ NNPDF4.0 CENTRAL VALUE AGREES BETTER WITH DATA
- EXP AND TOTAL χ^2 CLOSER $\Rightarrow$ NNPDF4.0 PDF UNCERTAINTIES SMALLER
- AGREEMENT WITH DATA OF ALL PDF SETS COMPARABLE $\Rightarrow$ ALL UNCERTAINTIES FAITHFUL

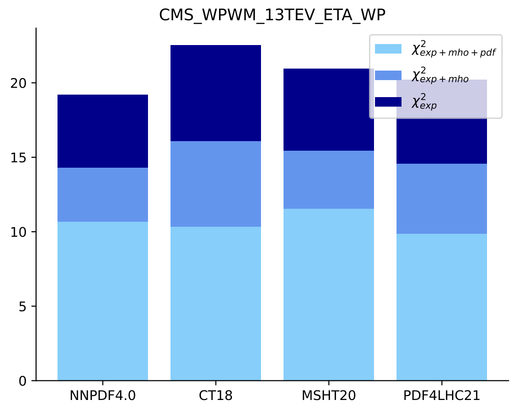
WHAT DOES ML BUY US?: ACCURACY

- COMPARISON TO DATA PUBLISHED AFTER PUBLICATION OF NNPDF4.0
- χ^2 WITH EXP, EXP+TH AND TOTAL (EXP+TH+PDF) UNCERTAINTIES

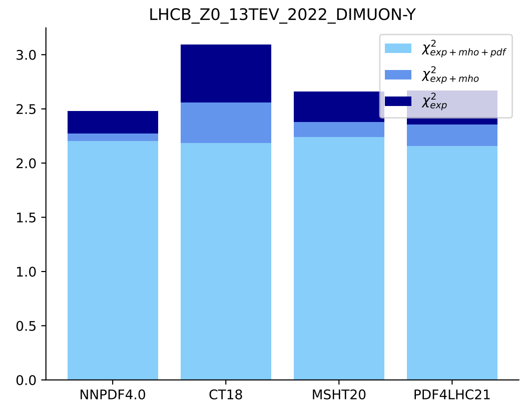
ATLAS Z



CMS W^+



LHCb Z

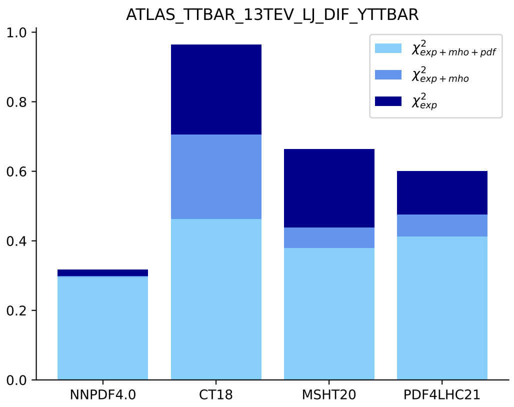


- EXP χ^2 LOWER $\Rightarrow$ NNPDF4.0 CENTRAL VALUE AGREES BETTER WITH DATA
- EXP AND TOTAL χ^2 CLOSER $\Rightarrow$ NNPDF4.0 PDF UNCERTAINTIES SMALLER
- AGREEMENT WITH DATA OF ALL PDF SETS COMPARABLE $\Rightarrow$ ALL UNCERTAINTIES FAITHFUL

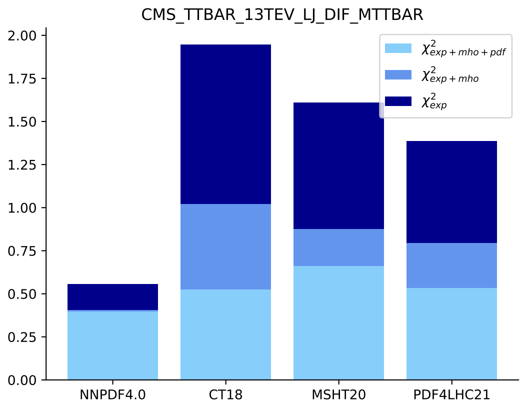
WHAT DOES ML BUY US?: ACCURACY

- COMPARISON TO DATA PUBLISHED AFTER PUBLICATION OF NNPDF4.0
- χ^2 WITH EXP, EXP+TH AND TOTAL (EXP+TH+PDF) UNCERTAINTIES

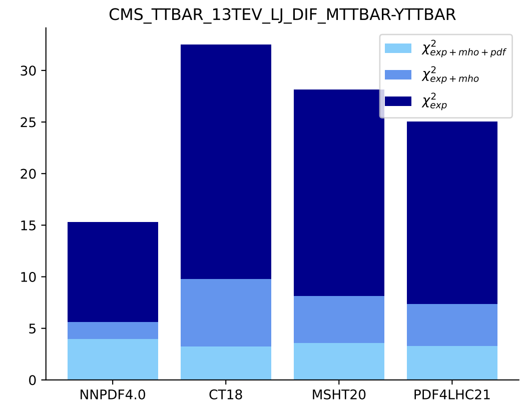
ATLAS TOP PAIRS SINGLE DIFF



CMS TOP PAIRS SINGLE DIFF



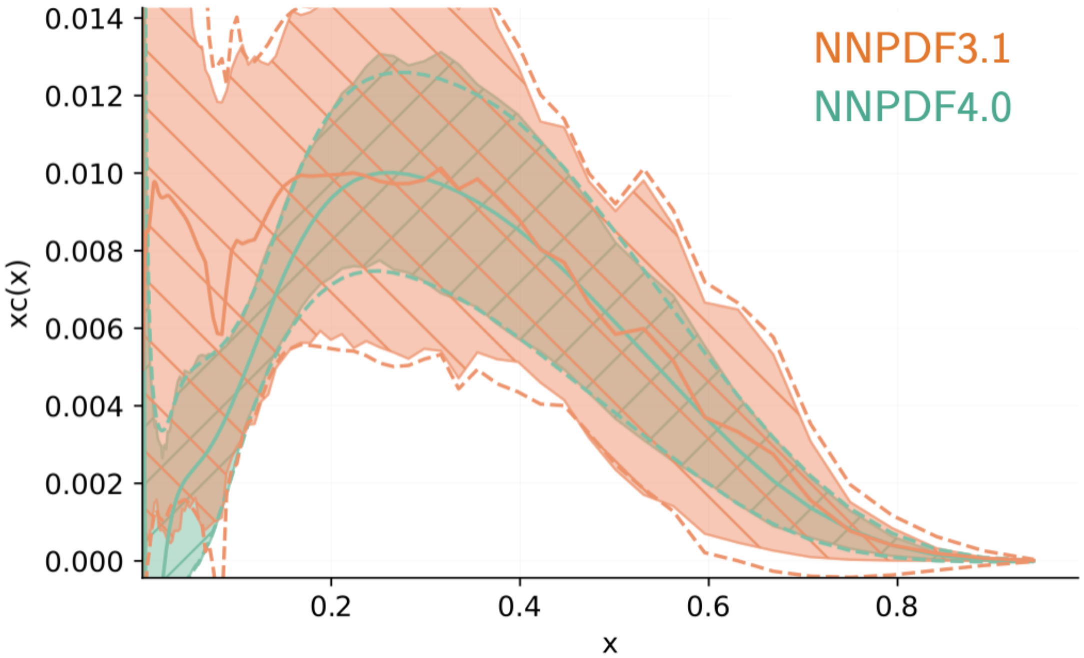
CMS TOP PAIRS DOUBLE DIFF



- EXP χ^2 LOWER $\Rightarrow$ NNPDF4.0 CENTRAL VALUE AGREES BETTER WITH DATA
- EXP AND TOTALχ^2 CLOSER $\Rightarrow$ NNPDF4.0 PDF UNCERTAINTIES SMALLER
- AGREEMENT WITH DATA OF ALL PDF SETS COMPARABLE $\Rightarrow$ ALL UNCERTAINTIES FAITHFUL

WHAT DOES ML BUY US?
CHARM IN THE PROTON
EVOLVE CHARM PDF ($N_f = 4$ SCHEME) DOWN TO $Q \sim m_c$

CHARM PDF (FITTED FROM DATA) AT Q=1.65 GeV



- IF $Q \sim m_c$ ($m_c = 1.51$ GeV), CHARM QUARK DECOUPLES (Collins, Wilczek, Zee, 1978): $\ln \frac{Q^2+m_c^2}{m_c^2} \approx \frac{m_c^2}{Q^2}$
- $N_f = 3$ ACTIVE FLAVORS IN β FUNCTION & EVOLUTION EQUATIONS
- DECOUPLING VS $\overline{\mathrm{MS}}$ $\Leftrightarrow$ DIFFERENT RENORMALIZATION & FACTORIZATION SCHEMES

MATCHING

OME CONTRIBUTING
TO THE CHARM PDF
SOLID $\Rightarrow$ HEAVY; DASHED $\Rightarrow$ LIGHT

- PDFS, α_s IN $N_f = 3$ & $N_f = 4$
  RELATED BY MATCHING CONDITIONS
- DETERMINED BY COMPUTING
  OPERATOR MATRIX ELEMENTS
  IN EITHER SCHEME AND EQUATING:
  NNLO (Buza, et al., 1998),
  N^{3}LO (Ablinger, Blümlein et al, 2009-2017)

M. Buza et al.: Charm

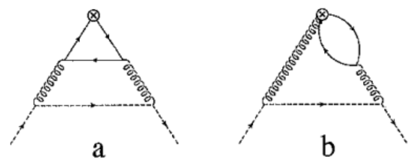


Fig. 2. $O(\alpha_s^2)$ contributions to the purely-singlet OME $A^{\rm PS}_{q'q}$. Here q and q' are represented by the *dashed* and *solid lines* respectively. In the case of $q' = H$ these graphs contribute to the heavy-quark OME $A^{\rm PS}_{Hq}$

PERTURBATIVE CHARM

- NO CHARM PDF IN $N_f = 3$ SCHEME
- IN $N_f = 4$ SCHEME, CHARM DETERMINED BY PERTURBATIVE MATCHING
  STARTING AT NNLO (TWO LOOPS) DOES NOT VANISH AT ANY SCALE (HEAVY QUARK LOOPS)

INTRINSIC CHARM

- DEFINE CHARM PDF AS OME:

$$\langle p|\bar{c}\gamma^{\mu_1} D^{\mu_2} \dots D^{\mu_n} c|p\rangle = A_c^n p^{\mu_1} \dots p^{\mu_n} - \text{traces}$$

$$A_c^n = \int_0^1 dx\, x^{n-1} c(x)$$

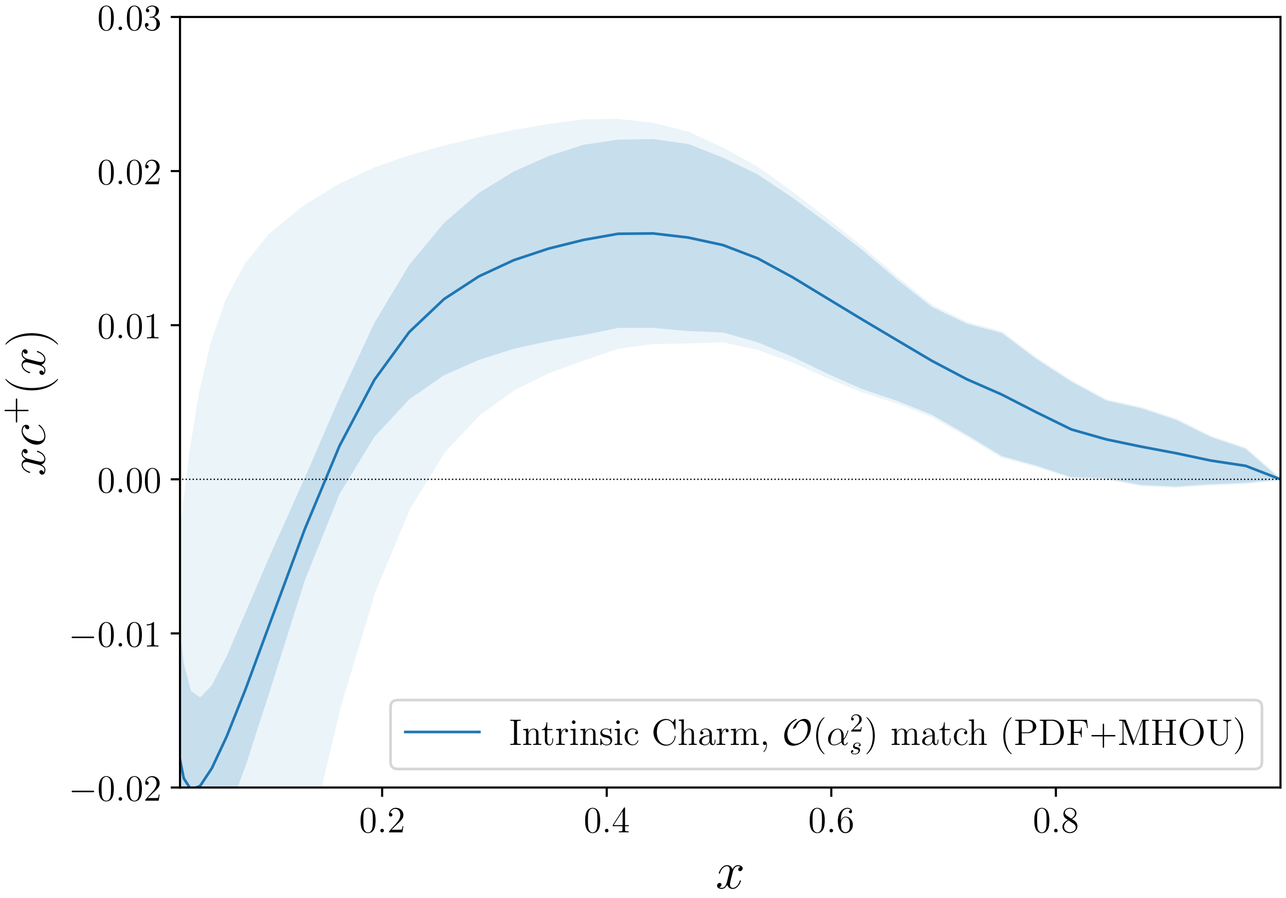
- DO NOT FACTOR CHARM MASS SINGULARITIES INTO OME
- $\Rightarrow$ CHOOSE $n_f = 3$ SCHEME
- CHARM PDF PURELY INTRINSIC, SCALE-INDEPENDENT

INTRINSIC CHARM IS CHARM IN THE $N_F = 3$ (DECOUPLING) SCHEME

INTRINSIC CHARM

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

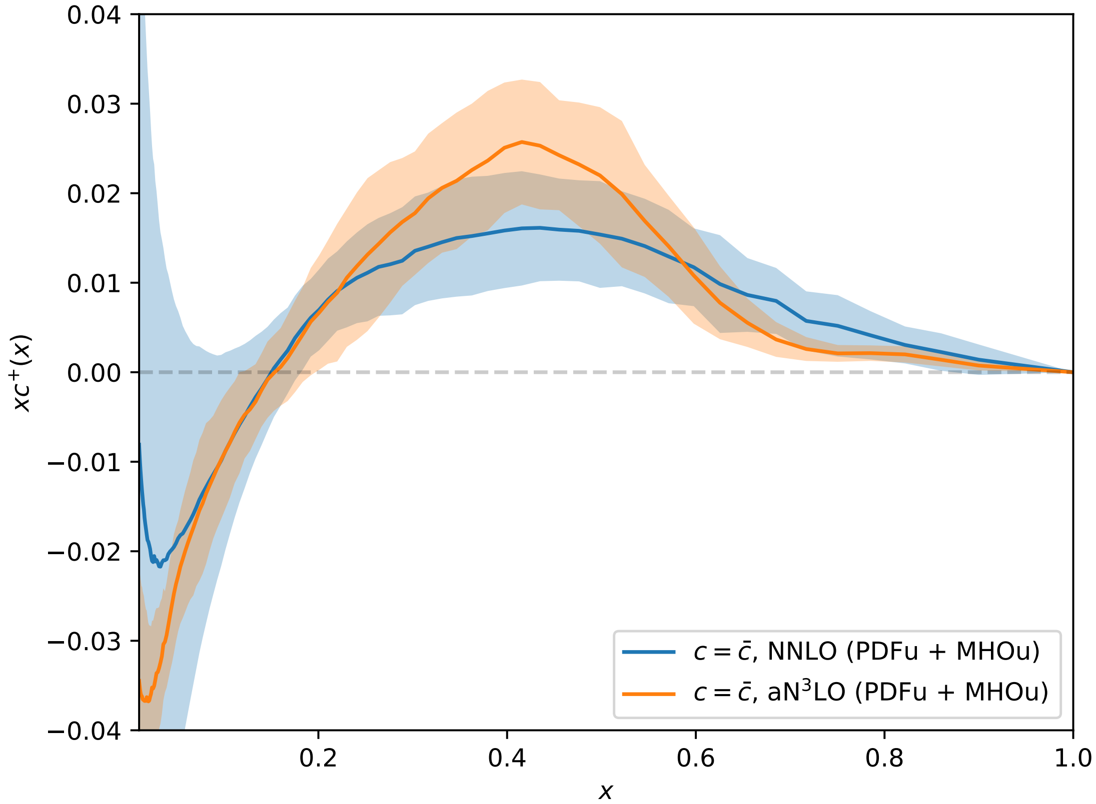
3FNS



CHARM AT AN^{3}LO

- IMPROVED N^{3}LO MATCHING (Blümlein, Ablinger et al., 2023) $\Rightarrow$ SOMEWHAT REDUCED INSTABILITY
- (APPROXIMATE) N^{3}LO PDFs $\Rightarrow$ " TRUE" MHOU
- MHOU (THEORY COVMAT FROM SCALE VARIATION) INCLUDED IN N^{3}LO RESULTS

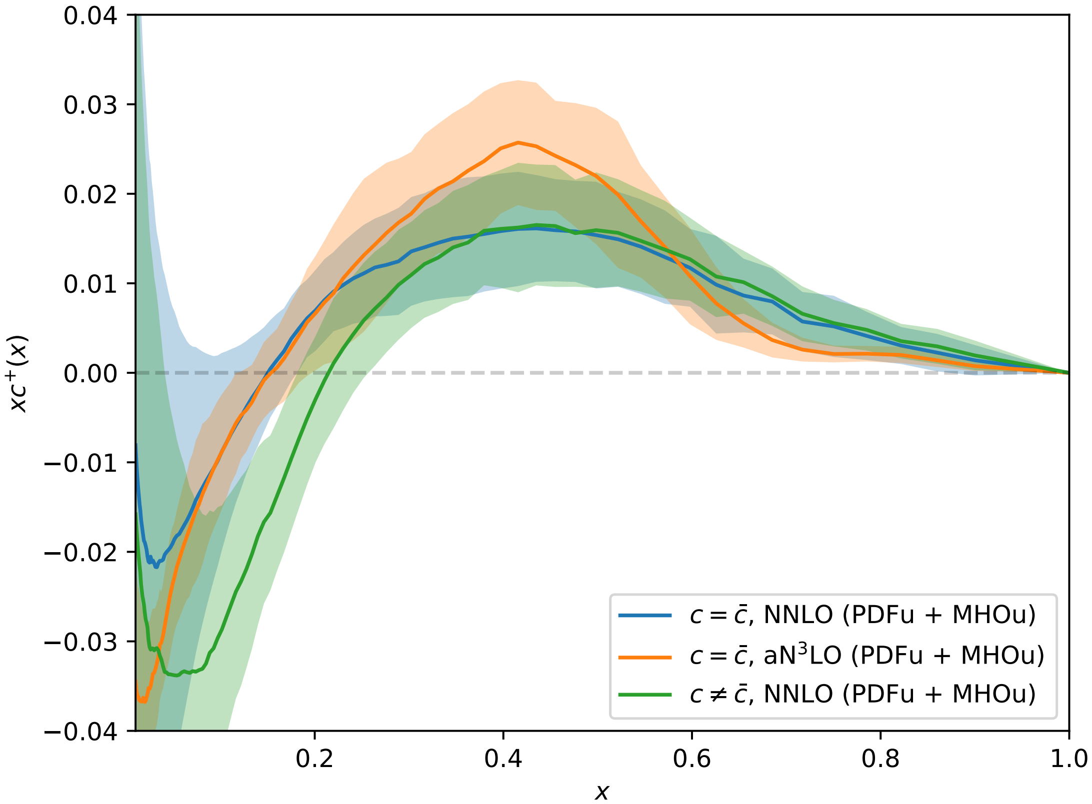
3FNS



THE VALENCE CHARM PDF

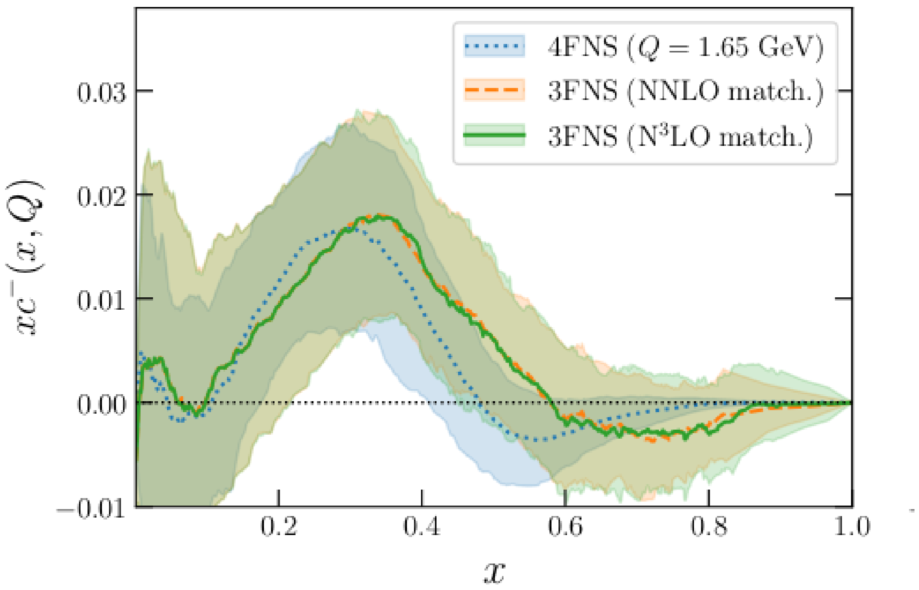
- INDEPENDENT PARAMETRIZATION FOR "SEA" $c^+ = c + \bar{c}$ AND "VALENCE" $c^- = c - \bar{c}$ PDFs
- TOTAL CHARM UNCHANGED

3FNS

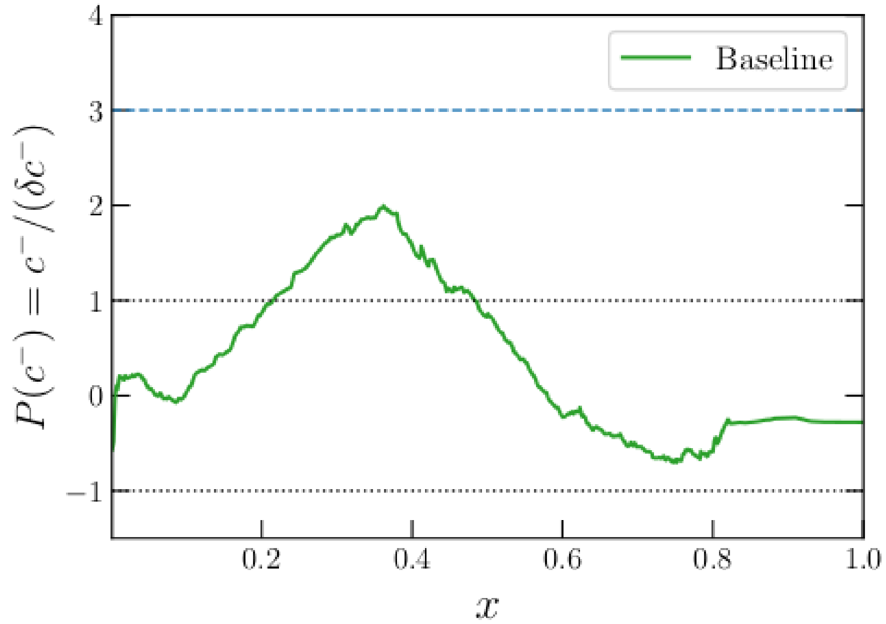


VALENCE CHARM

$c^-(x) = c(x) - \bar{c}(x)$



PULL (3FNS)



- NNLO $n_f = 4$ VALENCE PDF FROM PERTURBATIVE MATCHING VANISHES
- NONVANISHING VALENCE CHARM PDF IN VALENCE REGION $\Rightarrow$ INTRINSIC CHARM

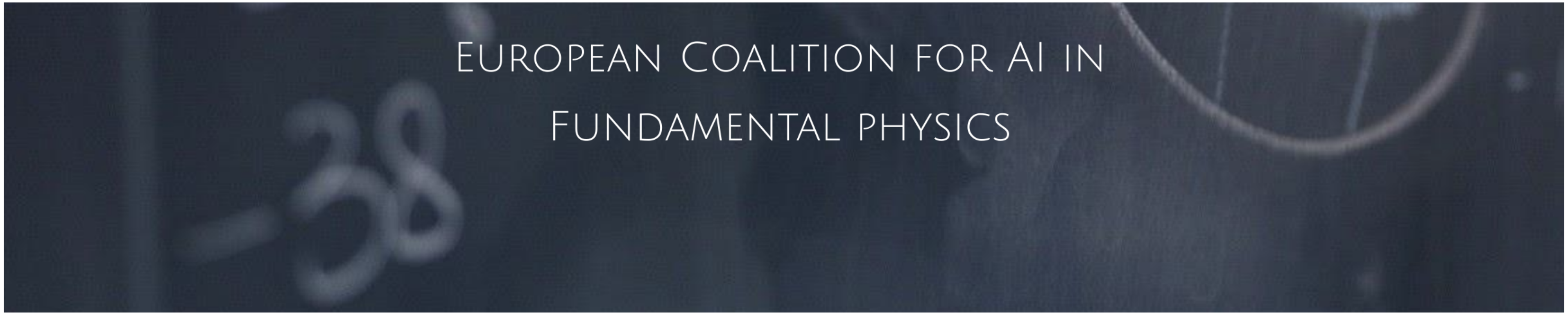
EPILOGUE
WHAT REMAINS TO BE DONE?

A TO DO LIST

- MACHINE LEARNING: XAI
  - HOW DOES THE ML MODEL RESPOND TO DATA INCONSISTENCIES?
  - AN ON-THE-FLY OVERLEARNING METRIC?
  - NEURAL NETWORKS VS. GAUSSIAN PROCESSES/BAYESIAN INFERENCE?
  - CORRELATION BETWEEN DATA FEATURES AND MODEL FEATURES?
- PDFs: PRECISION AND ACCURACY
  - AUTOMATIC K-FOLDS
  - HYPEROPT BEYOND χ^2 LOSS
  - FULL QCDxEW THEORY BEYOND K-FACTORS
  - $2 \to 2$ PROCESSES (VBS)

A COMMUNITY EFFORT

EUCAIF



JENAA
Joint ECFA-NuPECC-APPEC Activities

EuCAIF is an European initiative for advancing the use of Artificial Intelligence (AI) in Fundamental Physics. Members are working on particle physics, astroparticle physics, nuclear physics, gravitational wave physics, cosmology, theoretical physics as well as simulation and computational infrastructure.

THE AIPHY MSCA NETWORK

AIPHY - PhD Positions for Machine Learning in Particle Physics

Bohr Inst., U. Geneva (main), U. Heidelberg, ITP, INFN, Milan, LPNHE, Paris • Europe

hep-ex hep-ph hep-th cs PhD

Deadline on Jun 30, 2024

Job description:
Artificial intelligence not only pushes the boundaries of efficiency and precision in data analysis but transforms the way we think about data in scientific contexts. At the same time high energy physics is building on decades of experience in high quality data analysis combining rigorous uncertainty treatments with a fundamental understanding of data from quantum field theory.