MACHINE LEARNING AN UNKNOWN PHYSICAL LAW:

THE STRUCTURE OF THE PROTON

Stefano Forte
Università di Milano & INFN

Rome Physics Encounters

Frascati, February 20, 2020

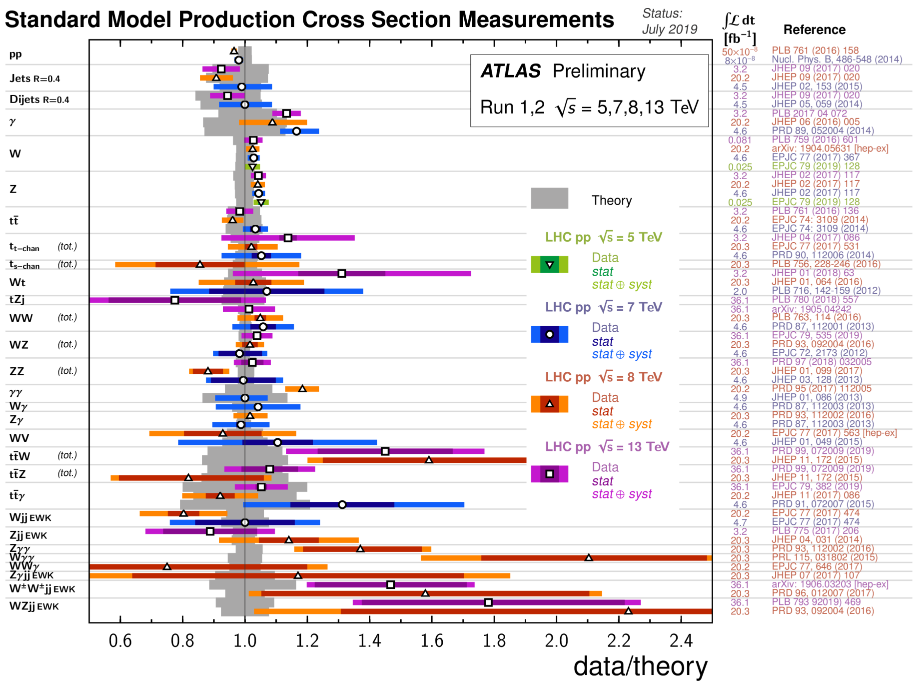
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 740006

PHYSICS AT THE LHC AS PRECISION PHYSICS

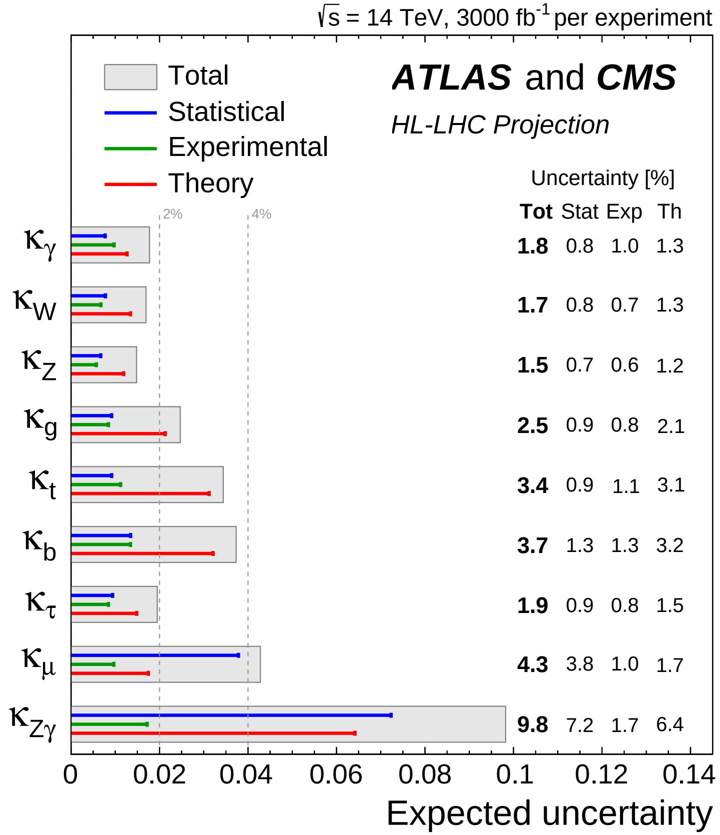
DEVIATIONS FROM SM

SM CROSS-SECTIONS TODAY:

TH. VS EXP.



HL-LHC: 2024-2040



$$\kappa_j^2 = \sigma_j / \sigma^{\rm SM}$$

- SM TESTED AT THE PERCENT LEVEL
- SEEING DEVIATIONS REQUIRES SUB-PERCENT ACCURACY

SUMMARY

PDFs: A RECAP SEQUENCE

- DETERMINING PDFs
- DISCOVERING NEW PHYSICS
- PDF UNCERTAINTIES, TOLERANCE AND ALL THAT

ARTFICIAL INTELLIGENCE

- PDFS, AI AND ML
- THE NNPDF METHODOLOGY: IDEAS AND TESTS
- THE STATE OF THE ART: ACCOMPLISHMENTS AND CHALLENGES

MACHINE LEARNING PDFs

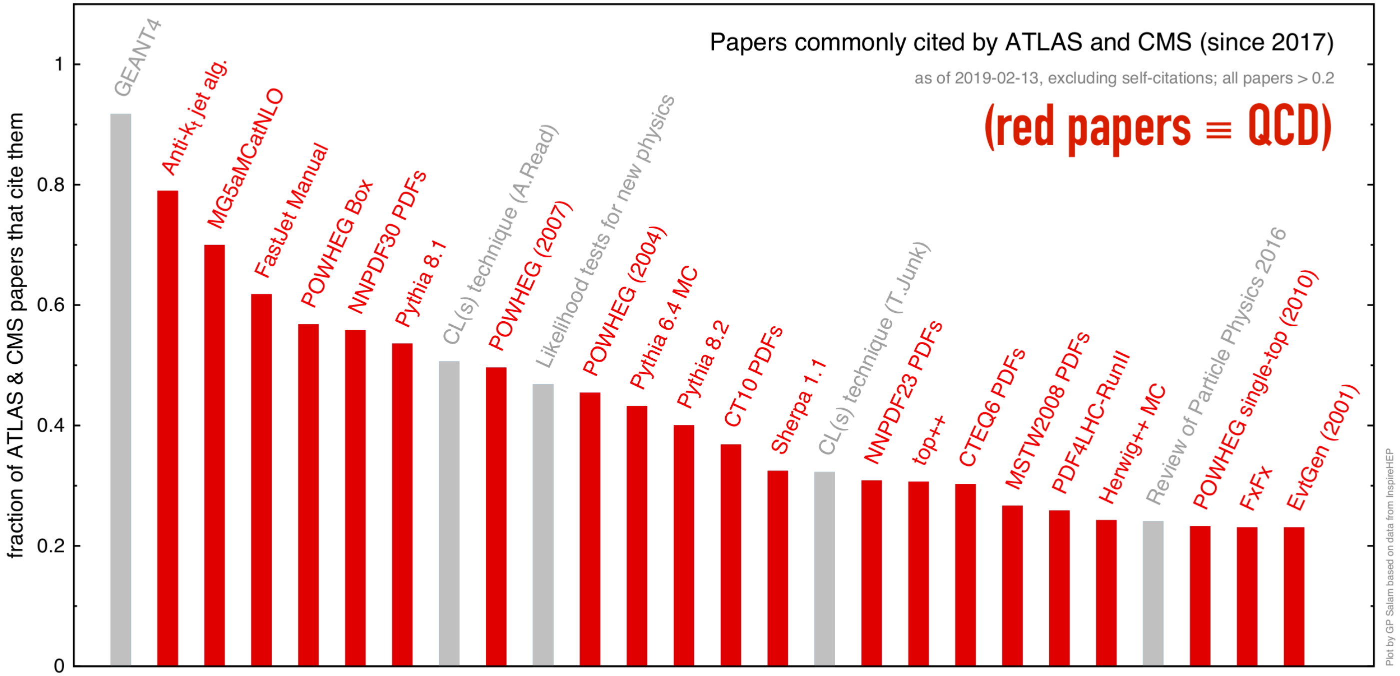
- OPTIMIZATION
- HYPEROPTIMIZATION
- INTO THE UNKNOWN

PDFS AND PRECISION PHYSICS

UNCERTAINTIES AND QCD

- THE LHC IS A PROTON COLLIDER $\Rightarrow$ ANY INTERACTION CONTAINS A STRONG INTERACTION
- QCD IS THE MAIN THEORETICAL PROBLEM
- .

PAPERS MOST CITED BY ATLAS (BY FRACTION)

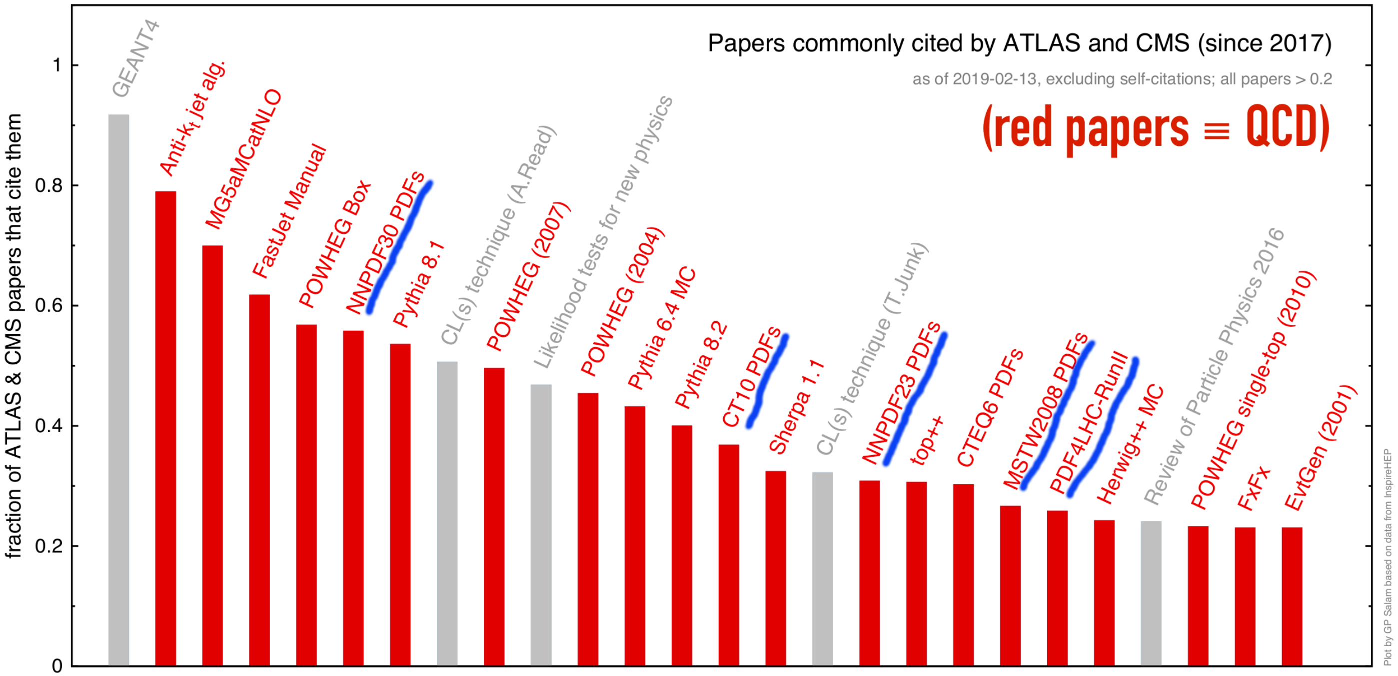


(G. Salam, 2019)

UNCERTAINTIES QCD, AND PDFs

- THE LHC IS A PROTON COLLIDER $\Rightarrow$ ANY INTERACTION CONTAINS A STRONG INTERACTION
- QCD IS THE MAIN THEORETICAL PROBLEM
- PDFs ARE THE DOMINANT ISSUE

PAPERS MOST CITED BY ATLAS (BY FRACTION)

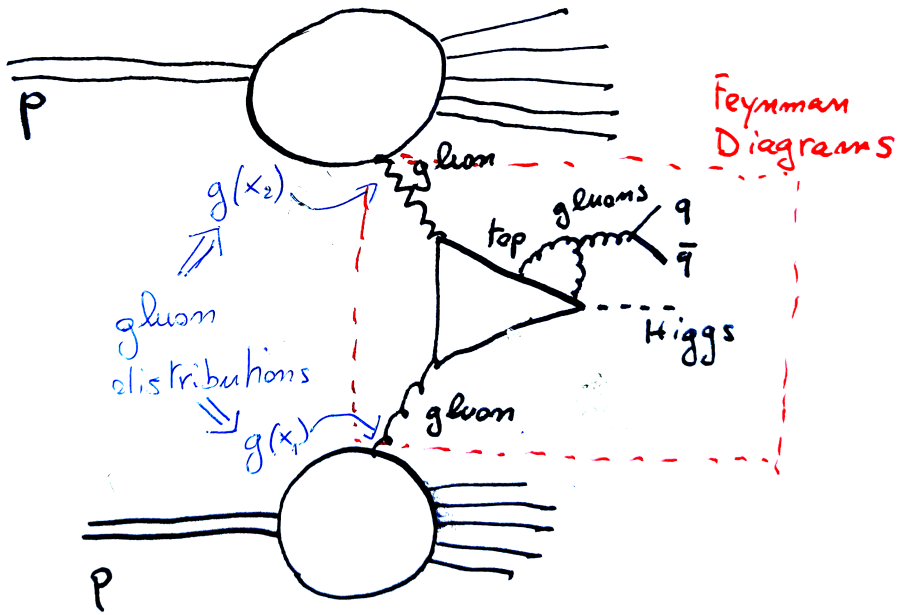


(G. Salam, 2019)

PDF papers underlined

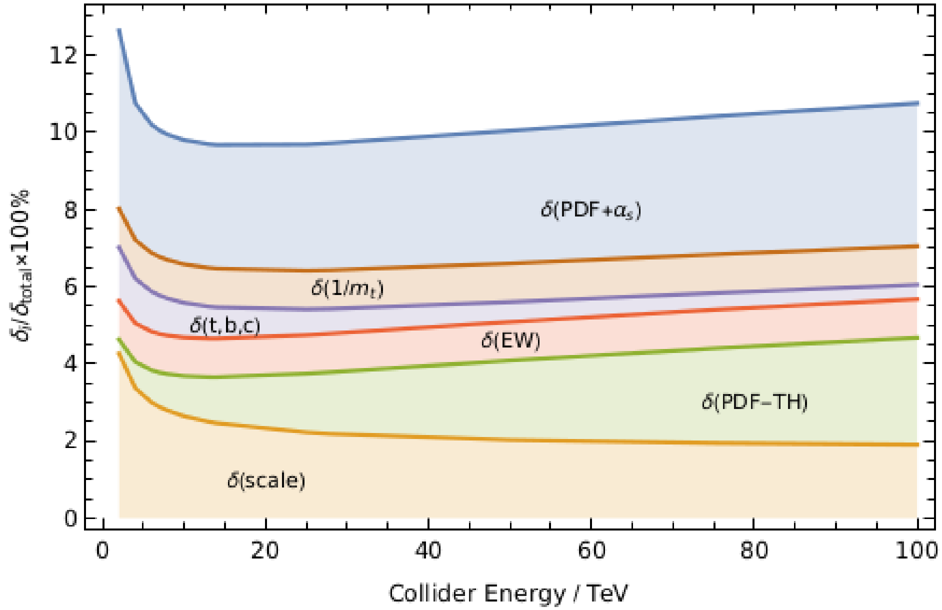
UNCERTAINTIES AND PDFs

QCD FACTORIZATION



UNCERTAINTIES:

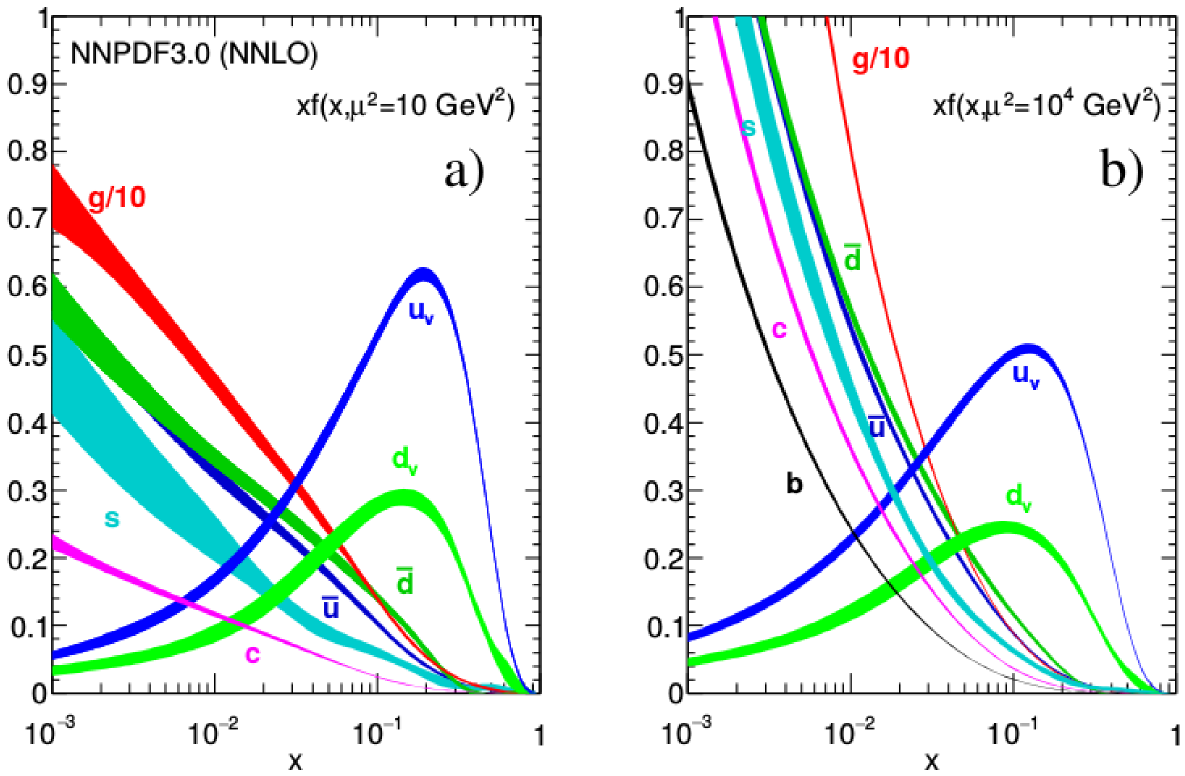
HIGGS IN GLUON FUSION



(HL-LHC Higgs WG report, 2019)

- PDF ESPRESS THE LIKELIHOOD OF A QUARK OR GLUONS (PARTONS) TO ENTER A COLLISION
- THEIR KNOWLEDGE IS A DOMINANT SOURCE OF UNCERTAINTY

A PORTRAIT OF THE PROTON
AS SEEN FROM A HIGGS BOSON



(PDG 2018)

- PARTON DISTRIBUTIONS: MOMENTUM FRACTION DISTRIBUTIONS FOR EACH TYPE OF QUARK, ANTIQUARK & THE GLUON
- EXTRACTED FROM DATA, COMPARING PDF-DEPENDENT PREDICTION & INVERTING
- MUST DETERMINE A PROBABILITY DISTRIBUTION OF FUNCTIONS FROM A DISCRETE SET OF DATA

HOW DID WE GET HERE?

DISCOVERY AT A HADRON COLLIDER AND PDFs

THE DISCOVERY OF THE W (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%

THEORETICAL PREDICTION

TABLE 2
Values (in nb) of the total cross sections for $W^{\pm}$ and Z^0 production

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

- 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

DISCOVERY AT A HADRON COLLIDER AND PDFs

THE DISCOVERY OF THE W (1984)

PDFs IN 1984

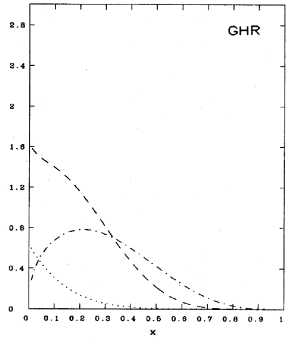


FIG. 25. Parton distributions of Glück, Hoffmann, and Reya (1982), at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and q_v (dotted line).

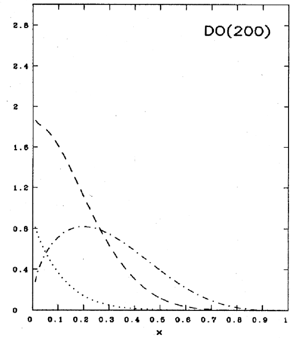


FIG. 27. "Soft-gluon" ($\Lambda=200$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

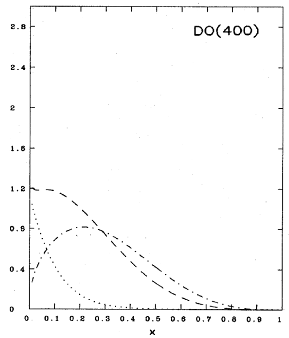


FIG. 26. "Hard-gluon" ($\Lambda=400$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

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

GHR VS DUKE-OWENS

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

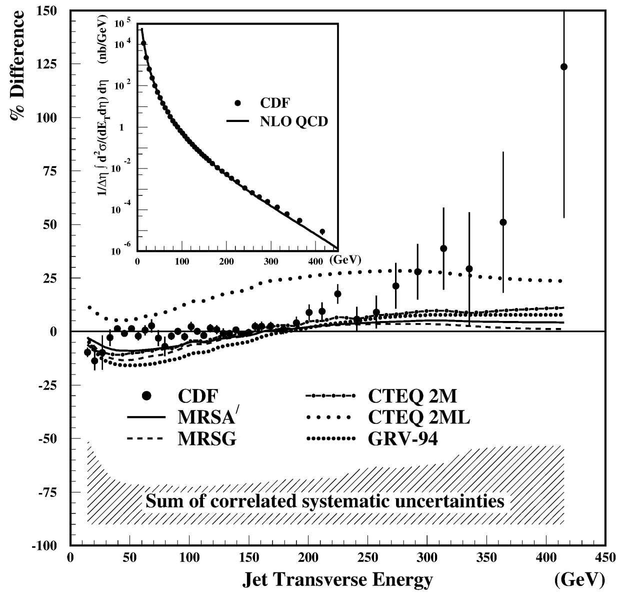
$\sqrt{S}$ (GeV)	$W^+ + W^-$ GHR	$W^+ + W^-$ DO1	$W^+ + W^-$ DO2	Z^0 GHR	Z^0 DO1	Z^0 DO2	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ GHR	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ DO1	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ DO2
540	4.2	4.3	4.1	1.3	1.3	1.2	3.1	3.4	3.5
700	6.2	6.3	6.1	2.0	1.9	1.8	3.1	3.3	3.4
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

- 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

DISCOVERY AT A HADRON COLLIDER AND PDFs
THE DISCOVERY OF QUARK COMPOSITENESS (1995)

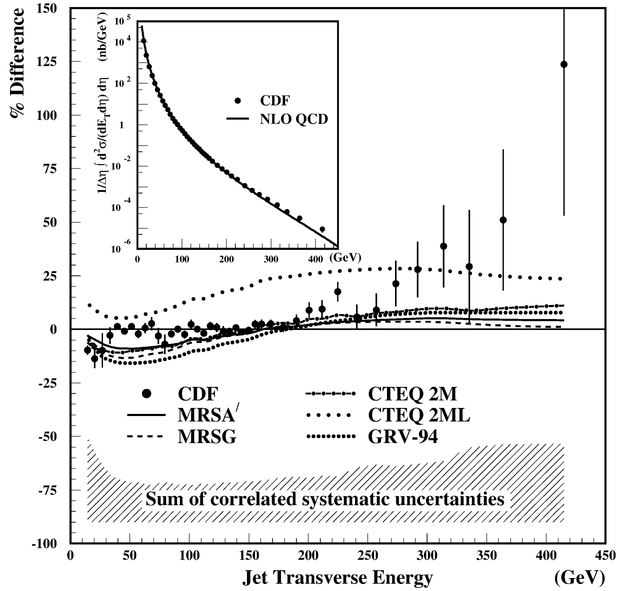
- DISCREPANCY BETWEEN QCD CALCULATION AND CDF JET DATA (1995)
- EVIDENCE FOR QUARK COMPOSITENESS
- .



Discovery at a hadron collider and PDFs

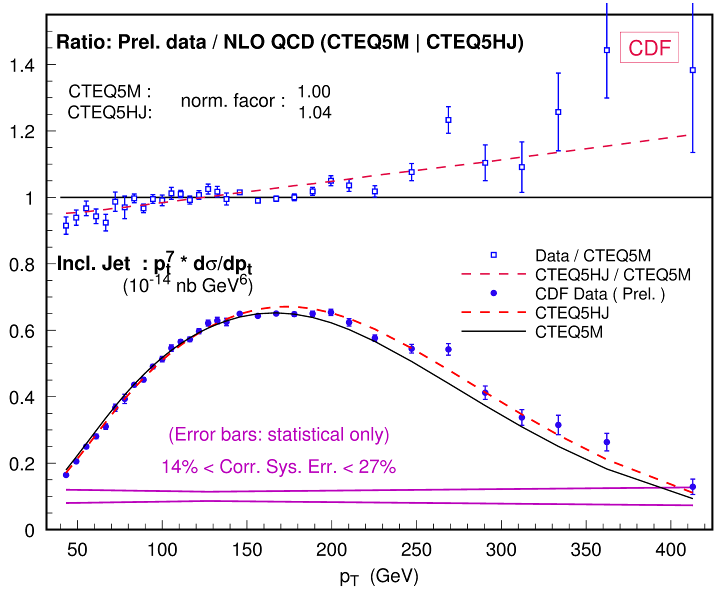
A better determination of the gluon PDF (1995)

- 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 $x \gtrsim 0.1$

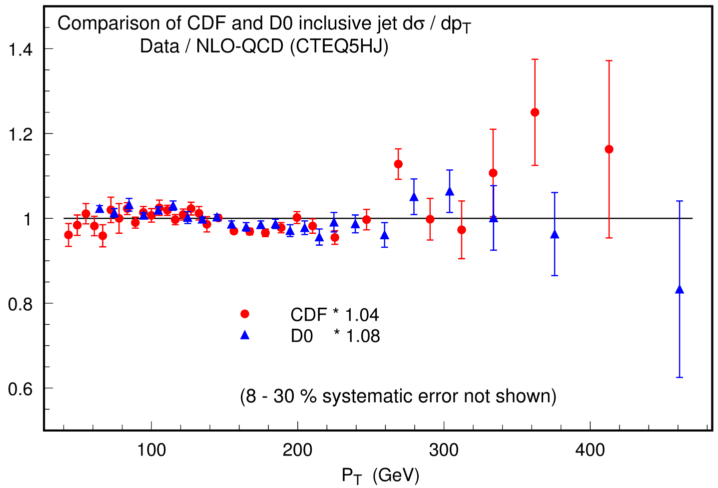


Discrepancy removed if jet data included in the fit

New CTEQ fit (1996)



Final CTEQ fit (1998)



WHAT'S THE PROBLEM ~ 2000

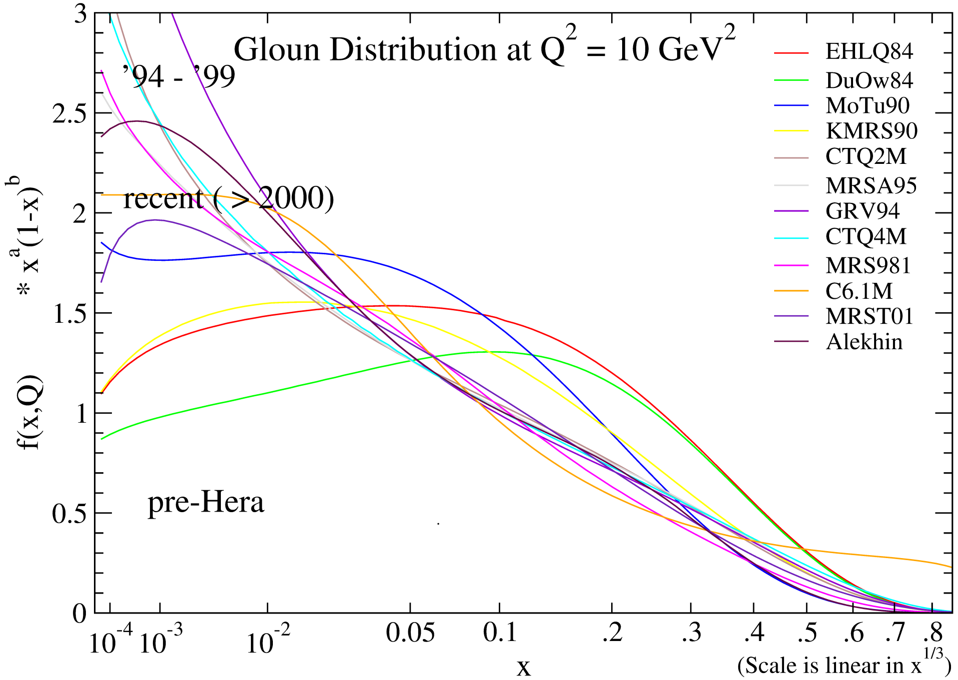
PDFs DETERMINED FITTING A MODEL-INSPIRED FUNCTIONAL FORM

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?

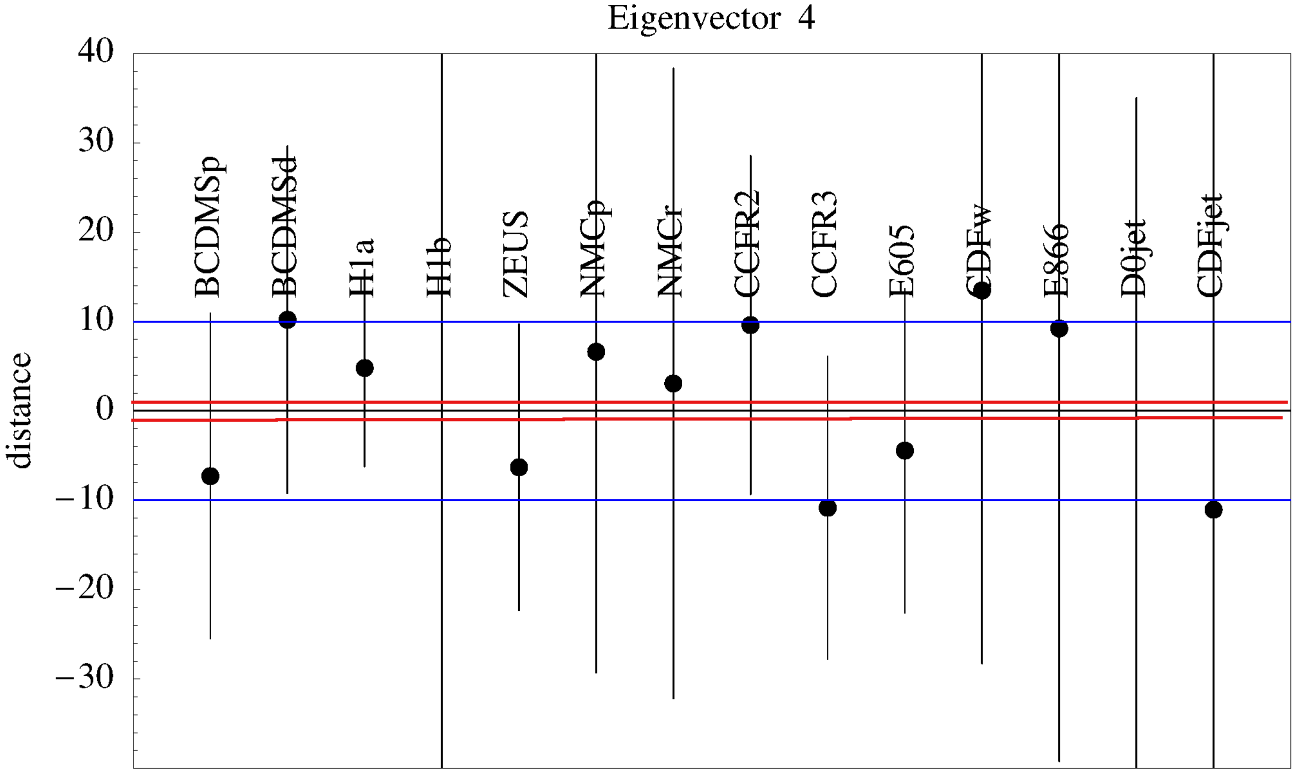
HISTORICAL COMPILATION OF GLUON PDFs



W.K.Tung, DIS 2004

FIRST PDFs WITH UNCERTAINTIES (2002)
"TOLERANCE"

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

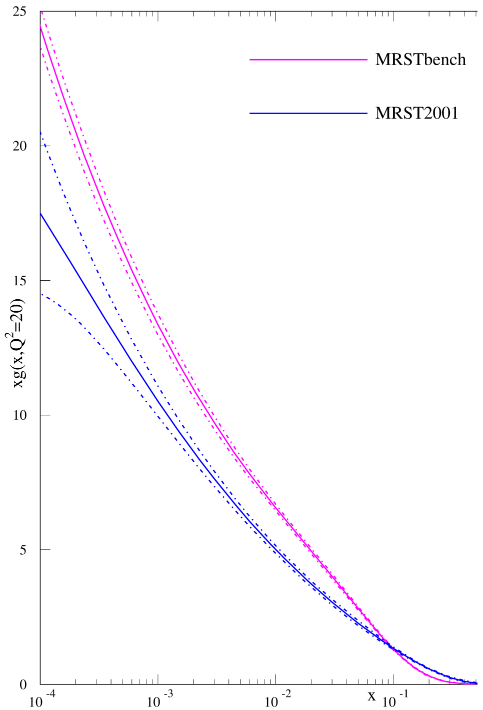


- SPREAD OF BEST-FIT FROM DIFFERENT DATA HUGE W.R. TO TEXTBOOK UNCERTAINTIES
- PDF UNCERTAINTIES RESCALED BY "TOLERANCE" $T \sim 10$

THE HERA-LHC BENCHMARK (2005)

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

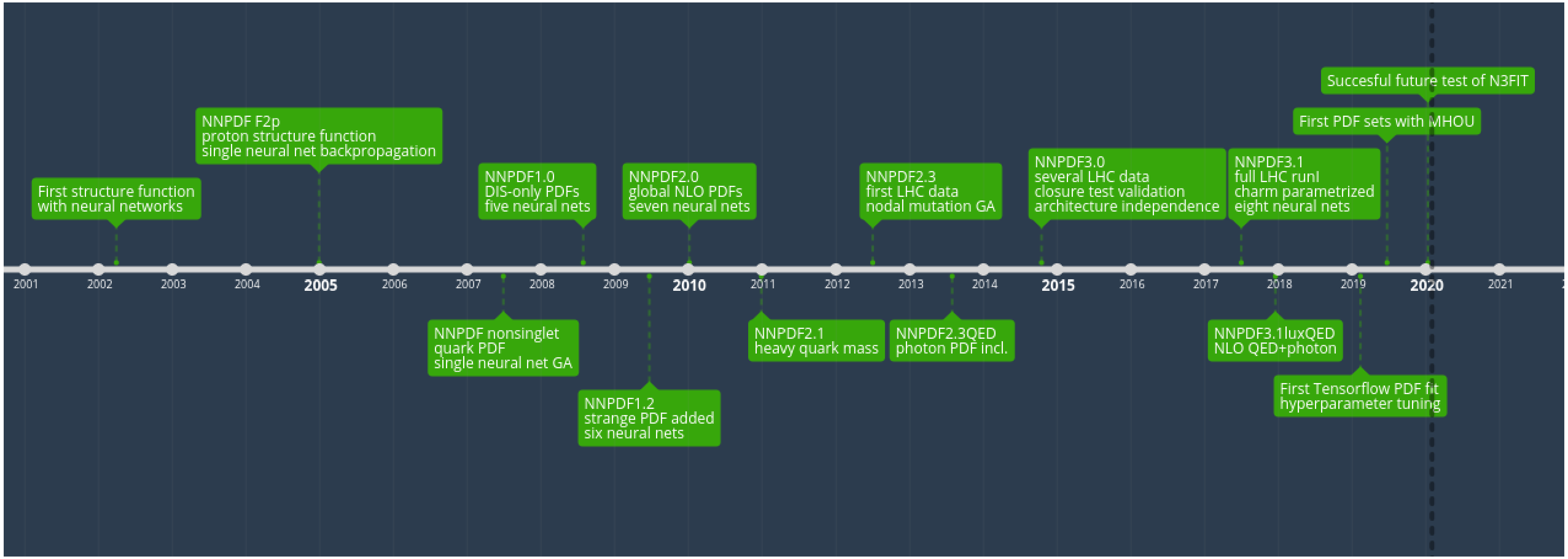
BENCHMARK VS DEFAULT GLUON



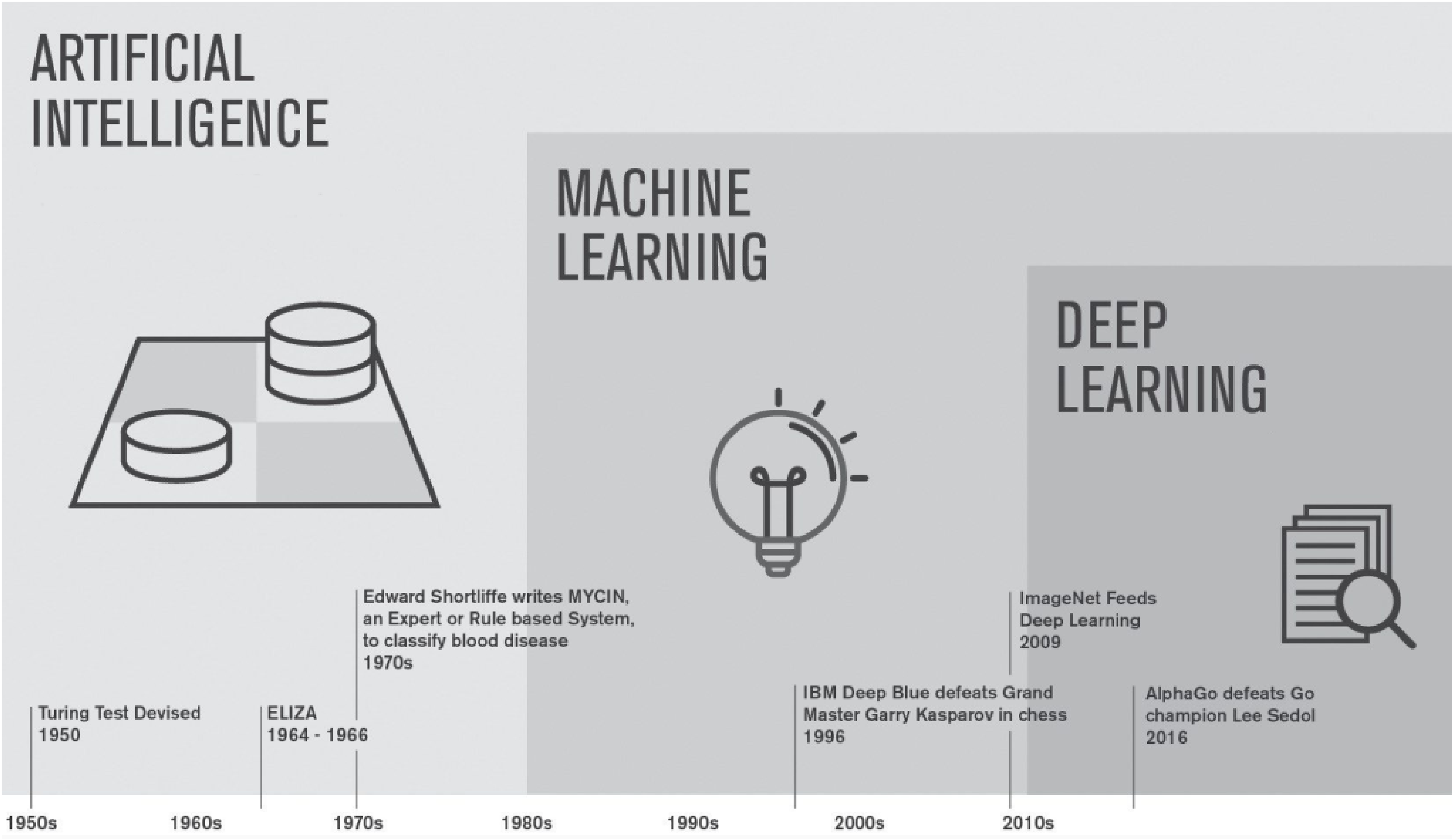
"...the partons extracted using a very limited data set are completely incompatible, even allowing for the uncertainties, with those obtained from a global fit with an identical treatment of errors...The comparison illustrates the problems in determining the true uncertainty on parton distributions."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

PDFS AND AI: NNPDF

PROTON STRUCTURE AS AN AI PROBLEM: NNPDF



FROM AI TO ML



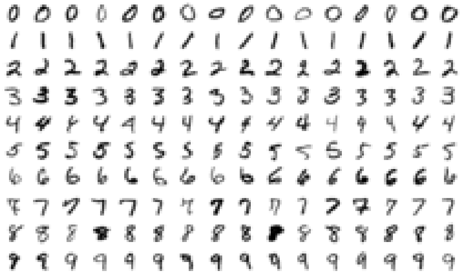
ARTIFICIAL INTELLIGENCE
MACHINE LEARNING
DEEP LEARNING
Edward Shortliffe writes MYCIN, an Expert or Rule based System, to classify blood disease 1970s
ImageNet Feeds Deep Learning 2009
Turing Test Devised 1950
ELIZA 1964 - 1966
IBM Deep Blue defeats Grand Master Garry Kasparov in chess 1996
AlphaGo defeats Go champion Lee Sedol 2016
1950s
1960s
1970s
1980s
1990s
2000s
2010s

SHIFTING OF PARADIGMS

"KNOWLEDGE BASED" AI

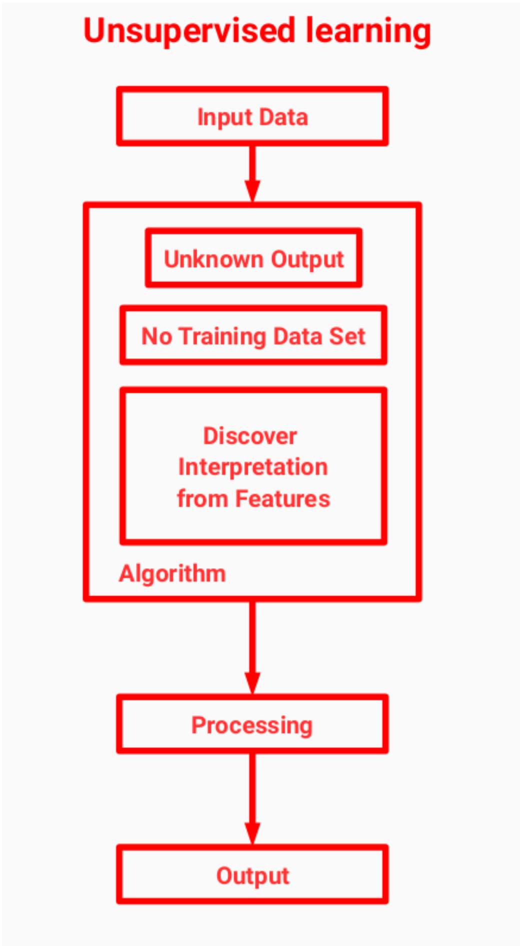
- LEARN AND IMPLEMENT A SET OF RULES
- GOOD FOR CHESS, BAD FOR REAL LIFE

MACHINE LEARNING

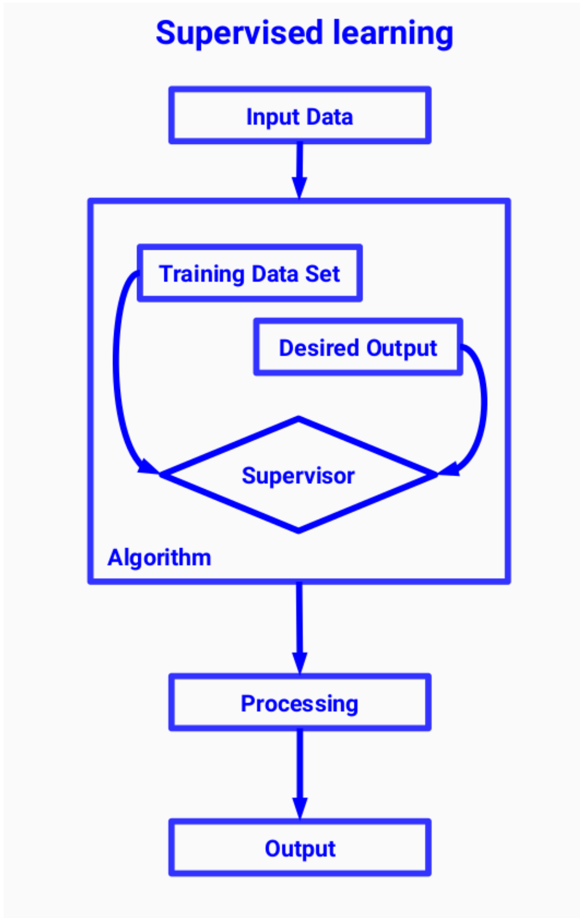


- "INTUITIVE" REPRESENTATION
- THE AI AGENT BUILID UP ITS OWN KNOWLEDGE

MACHINE LEARNING ALGORITHMS



EXTRACT AND OPTIMIZE
DATA FEATURES



OPTIMIZE A PROPERTY
LEARNING FROM DATA

LEARN FROM DATA
THE LEARNING STRATEGY

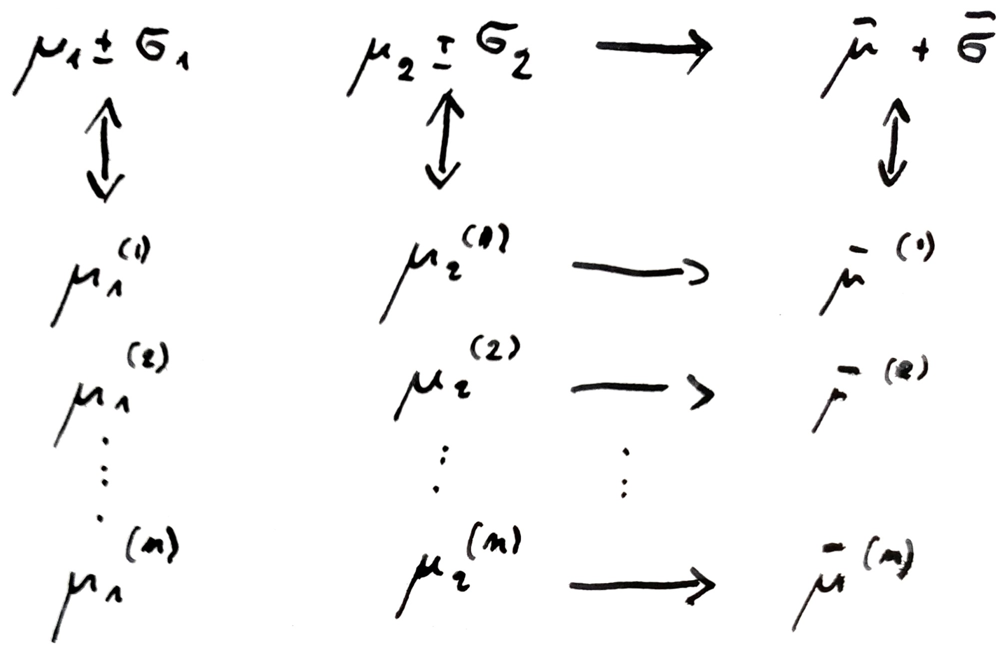
THE NNPDF APPROACH

COMBINING DATA BY MONTE CARLO

TWO MEASUREMENTS: $\mu_1 \pm \sigma_1; \qquad \mu_2 \pm \sigma_2$

COMBINATION: $\bar{\mu} \pm \bar{\sigma}; \quad \bar{\mu} = \frac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}; \; \bar{\sigma}^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$

MONTE CARLO REPRESENTATION

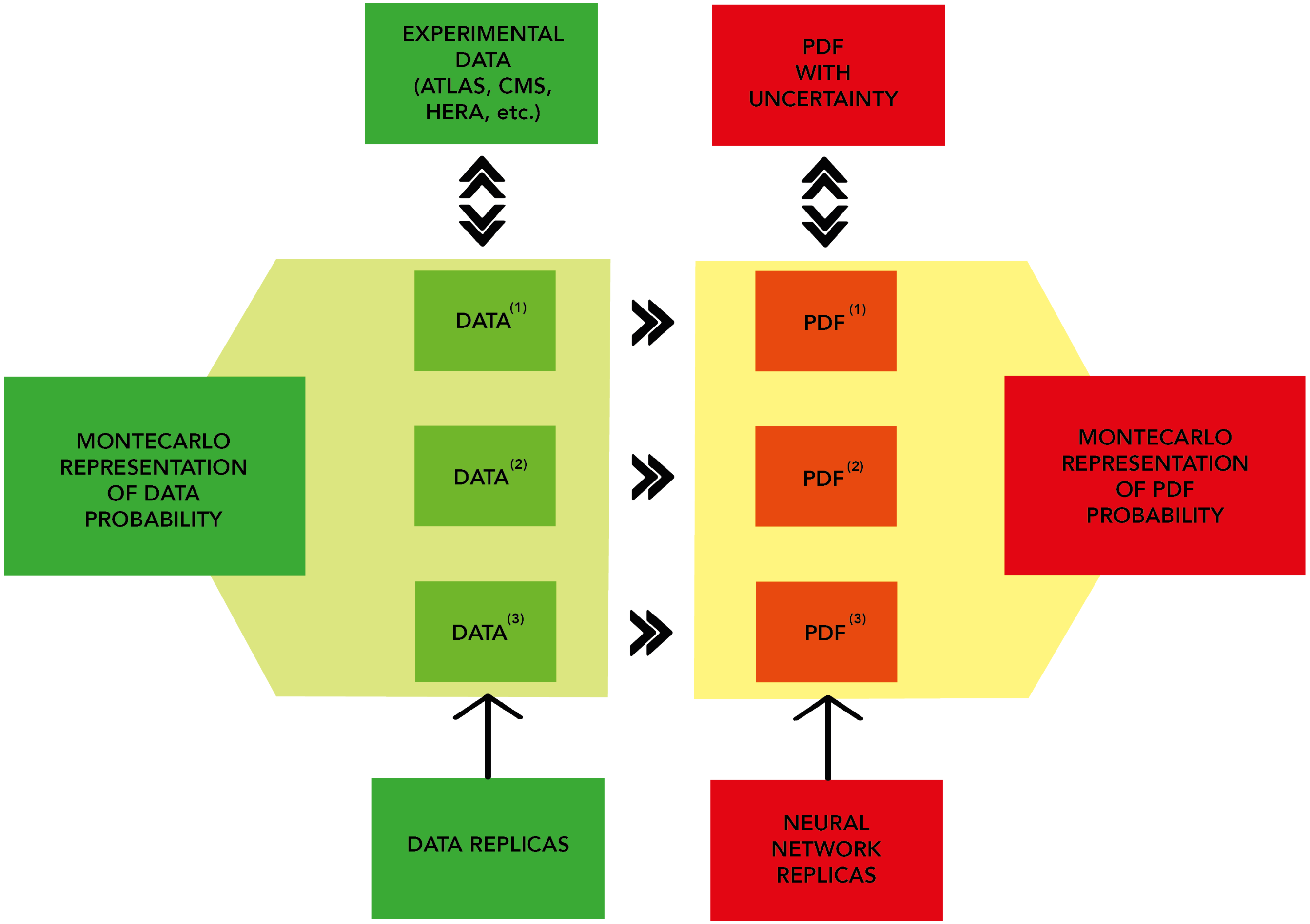


$\mu^{(i)} \Leftrightarrow$ REPLICA SAMPLE $\Leftrightarrow$ REPRESENTATION OF PROBABILITY DISTRIBUTION

NEED ONLY TO KNOW HOW TO COMBINE CENTRAL VALUES

AI FOR PDFS: THE NNPDF APPROACH
THE FUNCTIONAL MONTE CARLO

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

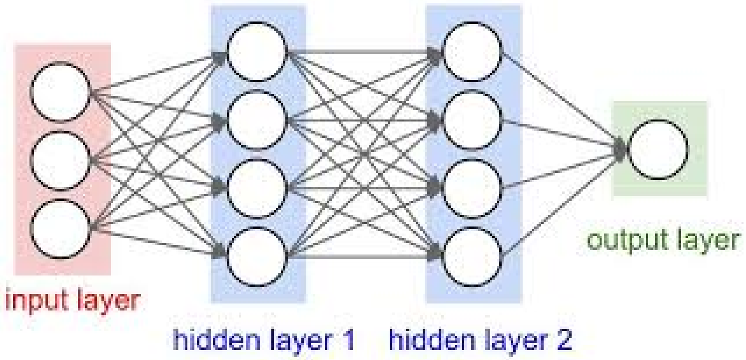


FINAL PDF SET: $f_i^{(a)}(x,\mu)$;
i =up, antiup, down, antidown, strange, antistrange, charm, gluon; $j = 1, 2, \ldots N_{\rm rep}$

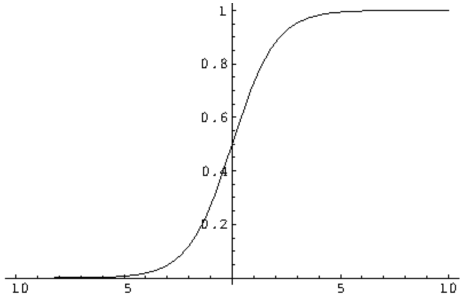
ARTIFICIAL INTELLIGENCE

NEURAL NETWORKS

ARCHITECTURE



ACTIVATION FUNCTION



PARAMETERS

- WEIGHTS ω_{ij}
- THRESHOLDS θ_i

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

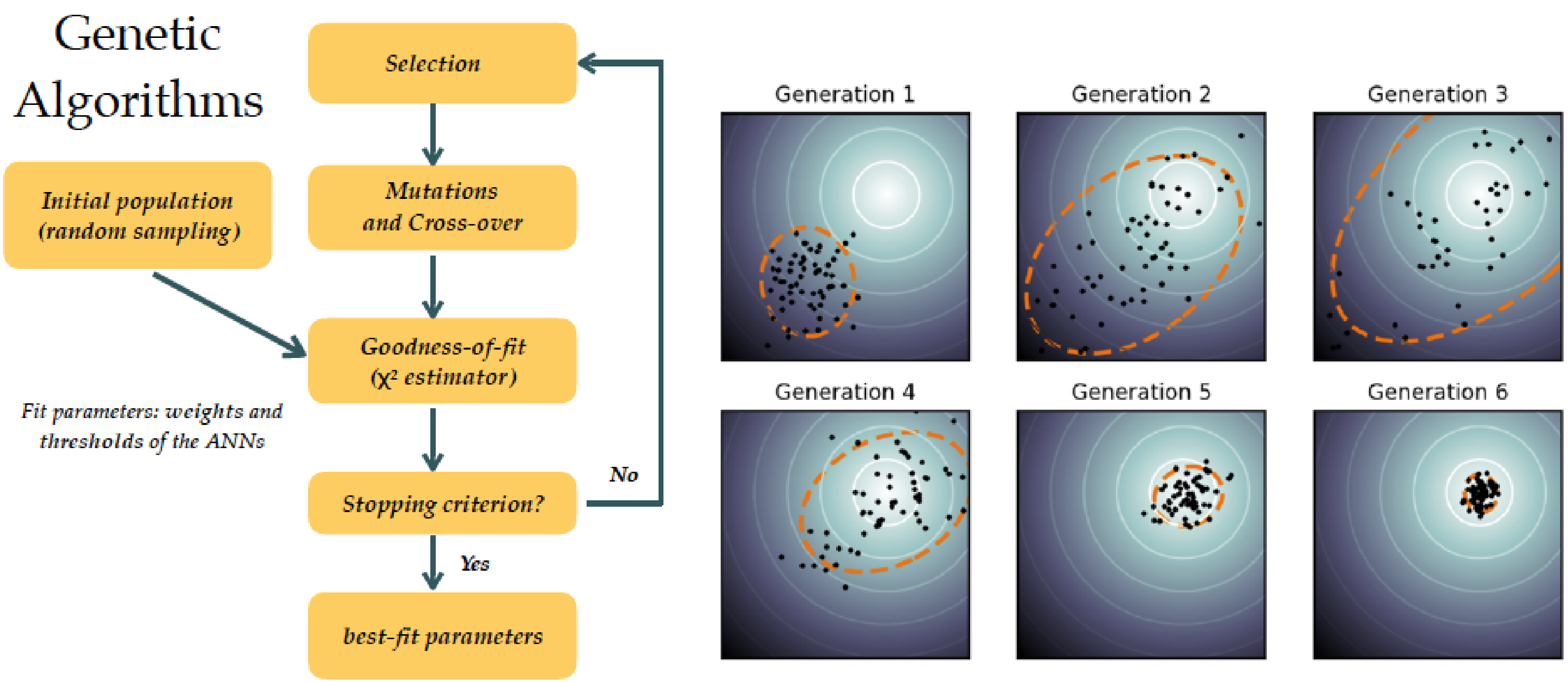
SIMPLEST EXAMPLE
1-2-1

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

NNPDF: $2-5-3-1$ NN FOR EACH PDF: $37\times 8 = 296$ PARAMETERS

GENETIC ALGORITHMS

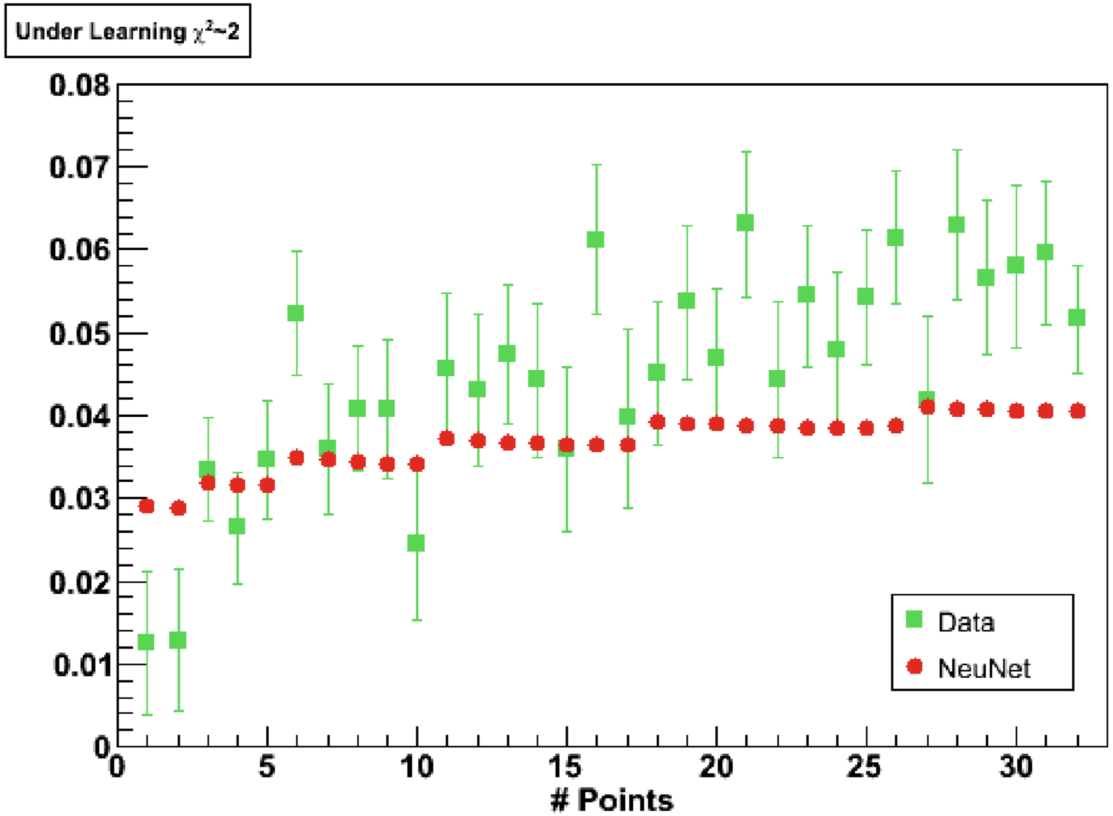
- BASIC IDEA: RANDOM MUTATION OF THE NN PARAMETER
- SELECTION OF THE FITTEST



NEURAL LEARNING

- COMPLEXITY INCREASES AS THE FITTING PROCEEDS
- UNTIL LEARNING NOISE
- WHEN SHOULD ONE STOP?

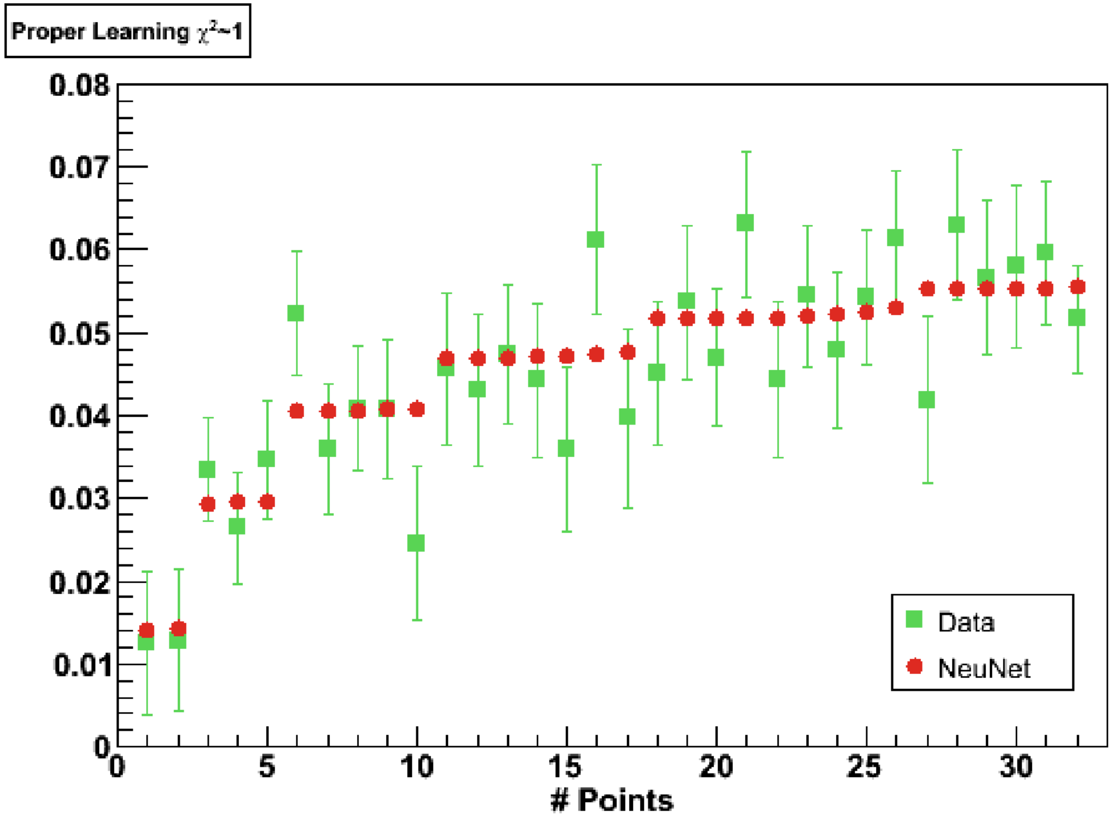
UNDERLEARNING



NEURAL LEARNING

- COMPLEXITY INCREASES AS THE FITTING PROCEEDS
- UNTIL LEARNING NOISE
- WHEN SHOULD ONE STOP?

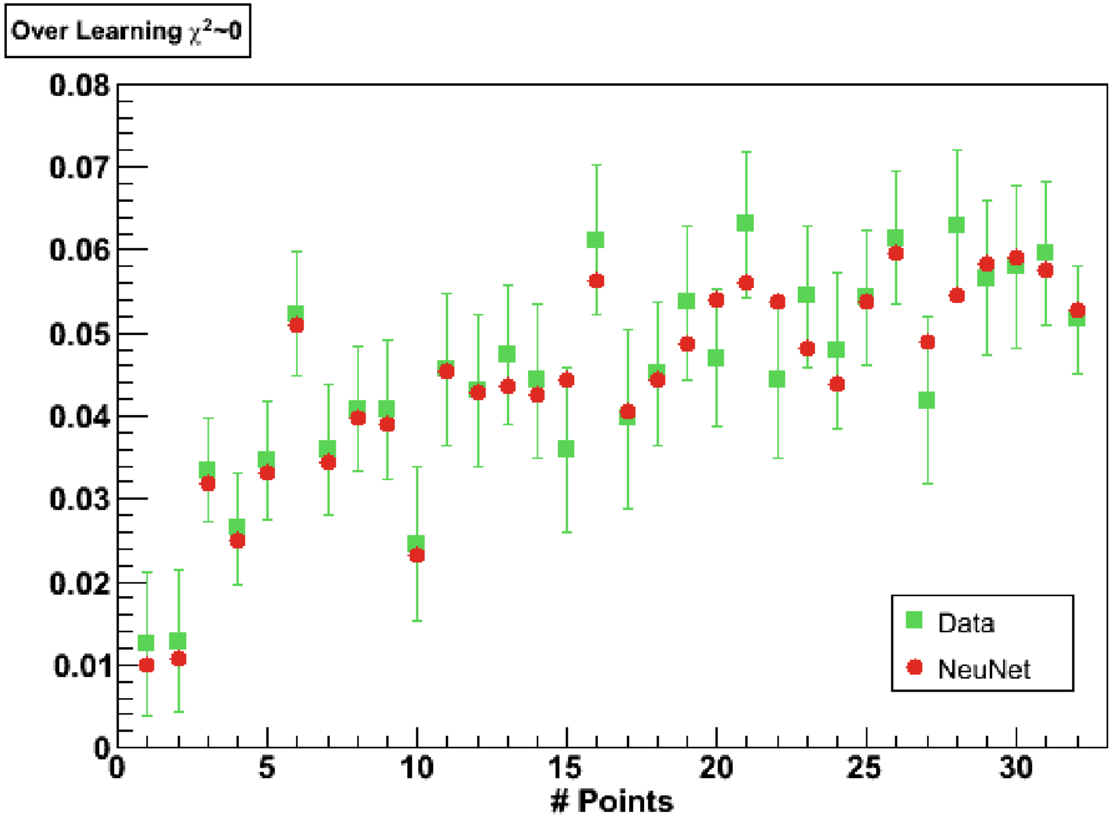
PROPER LEARNING



NEURAL LEARNING

- COMPLEXITY INCREASES AS THE FITTING PROCEEDS
- UNTIL LEARNING NOISE
- WHEN SHOULD ONE STOP?

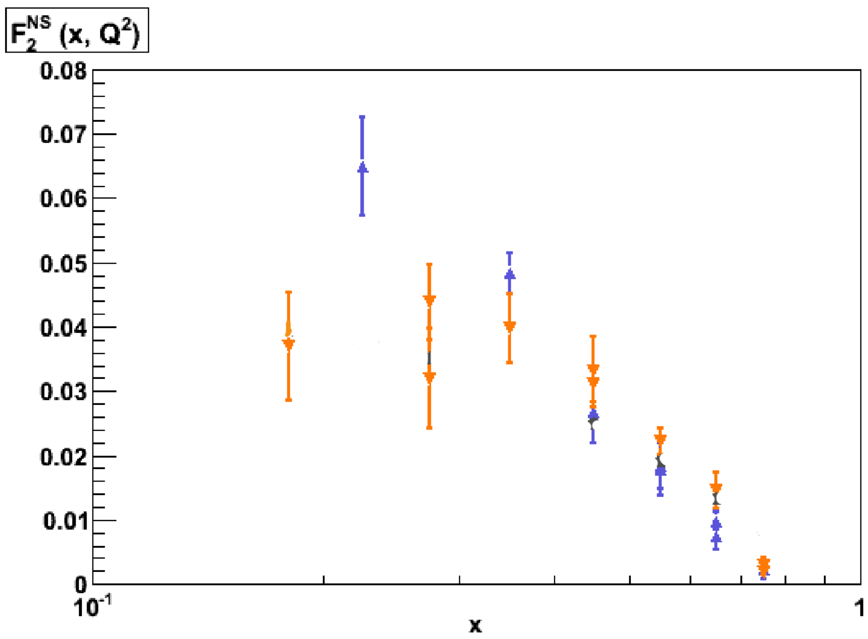
OVERLEARNING



OPTIMAL FIT: CROSS-VALIDATION

GENETIC MINIMIZATION:
AT EACH GENERATION, χ^2 EITHER UNCHANGED OR DECREASING

- DIVIDE THE DATA IN TWO SETS: TRAINING AND VALIDATION
- MINIMIZE THE χ^2 OF THE DATA IN THE TRAINING SET
- AT EACH ITERATION, COMPUTE THE χ^2 FOR THE DATA IN THE VALIDATION SET (NOT USED FOR FITTING)
- WHEN THE VALIDATION χ^2 STOPS DECREASING, STOP THE FIT

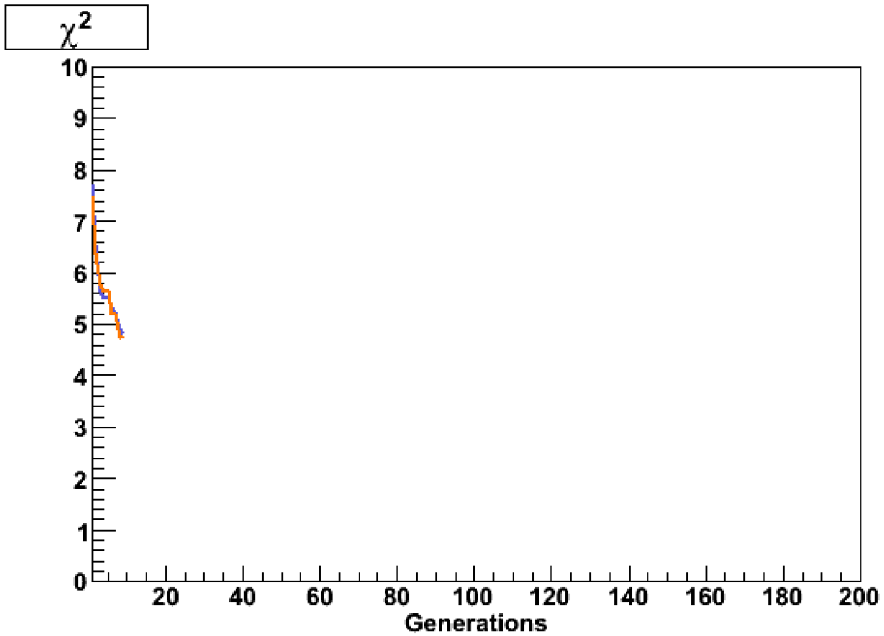


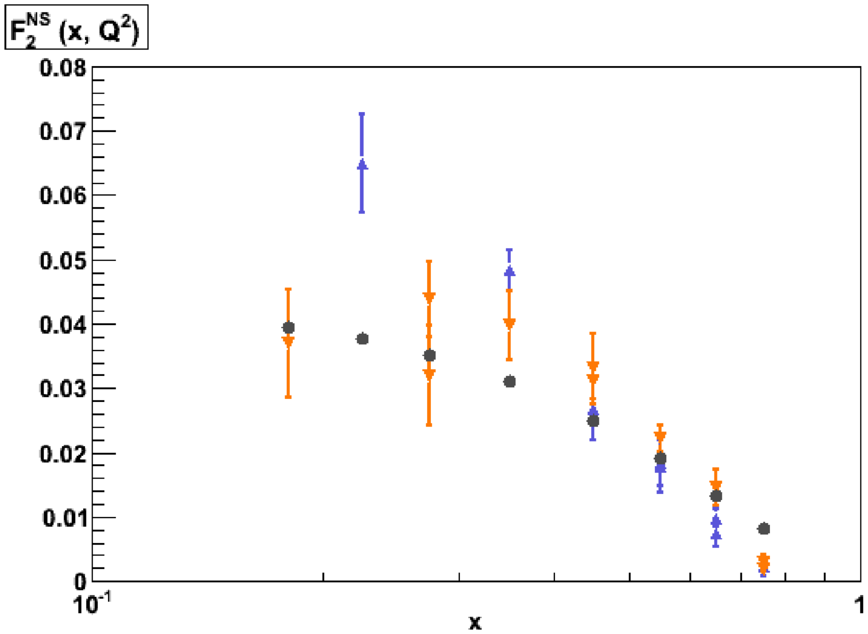
OPTIMAL FIT: CROSS-VALIDATION

GENETIC MINIMIZATION:
AT EACH GENERATION, χ^2 EITHER UNCHANGED OR DECREASING

- DIVIDE THE DATA IN TWO SETS: TRAINING AND VALIDATION
- MINIMIZE THE χ^2 OF THE DATA IN THE TRAINING SET
- AT EACH ITERATION, COMPUTE THE χ^2 FOR THE DATA IN THE VALIDATION SET (NOT USED FOR FITTING)
- WHEN THE VALIDATION χ^2 STOPS DECREASING, STOP THE FIT

GO!



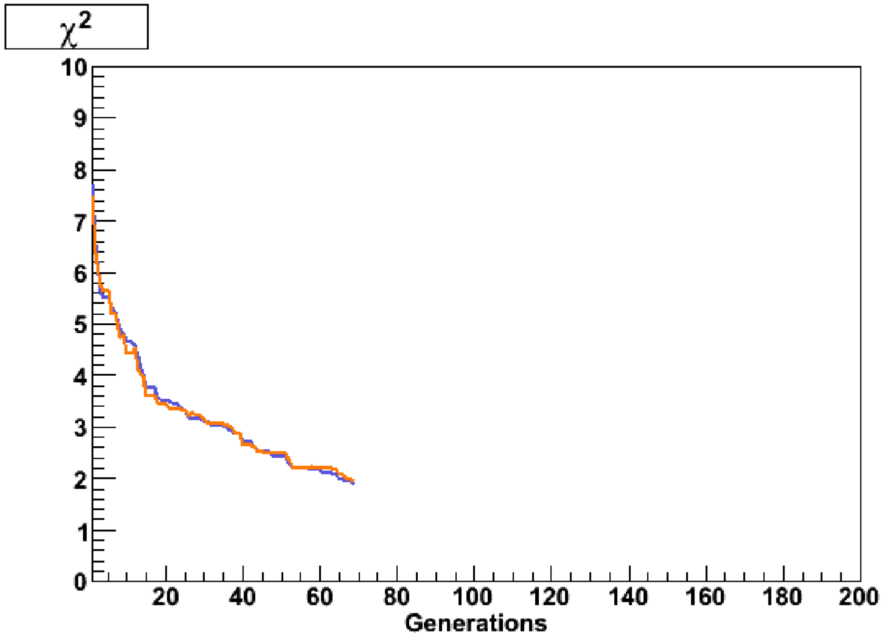


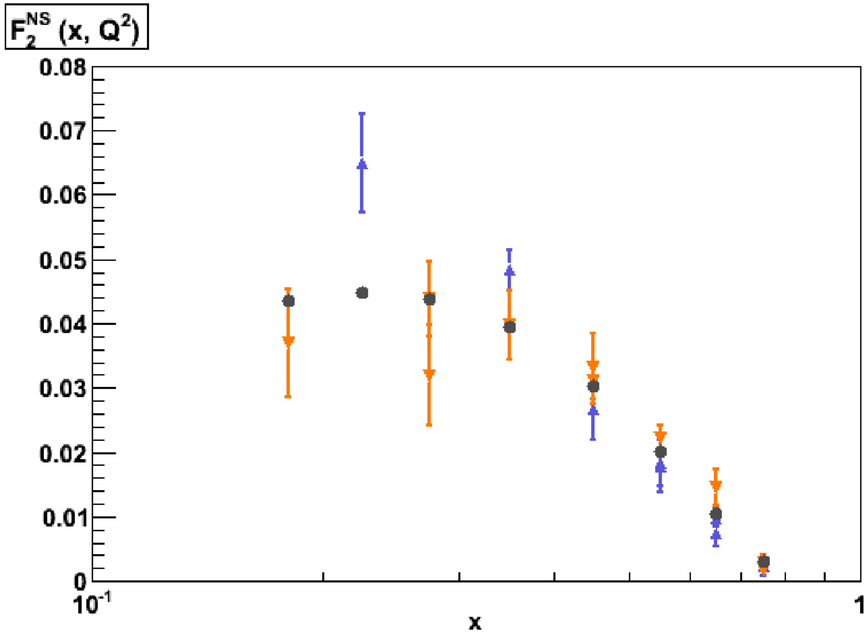
OPTIMAL FIT: CROSS-VALIDATION

GENETIC MINIMIZATION:
AT EACH GENERATION, χ^2 EITHER UNCHANGED OR DECREASING

- DIVIDE THE DATA IN TWO SETS: TRAINING AND VALIDATION
- MINIMIZE THE χ^2 OF THE DATA IN THE TRAINING SET
- AT EACH ITERATION, COMPUTE THE χ^2 FOR THE DATA IN THE VALIDATION SET (NOT USED FOR FITTING)
- WHEN THE VALIDATION χ^2 STOPS DECREASING, STOP THE FIT

STOP!



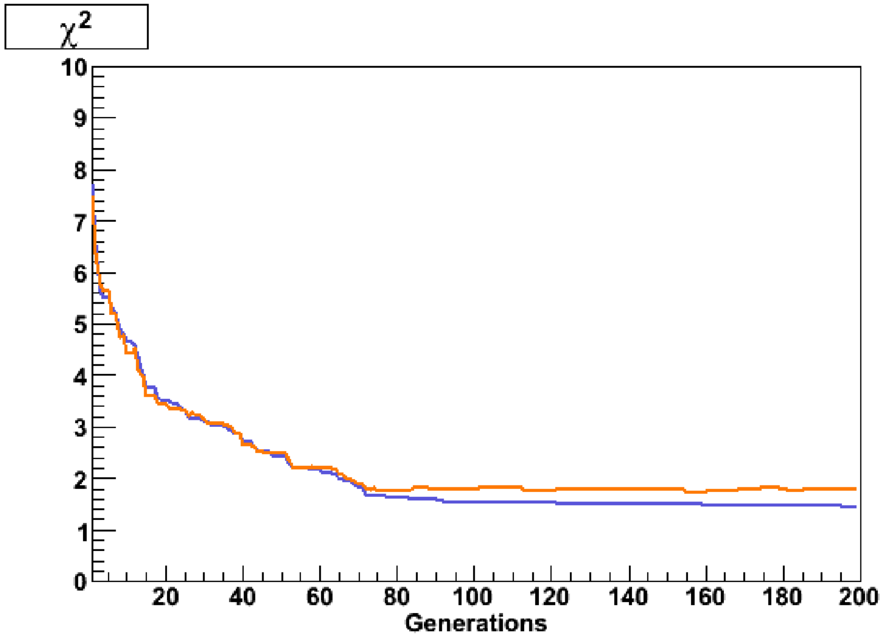


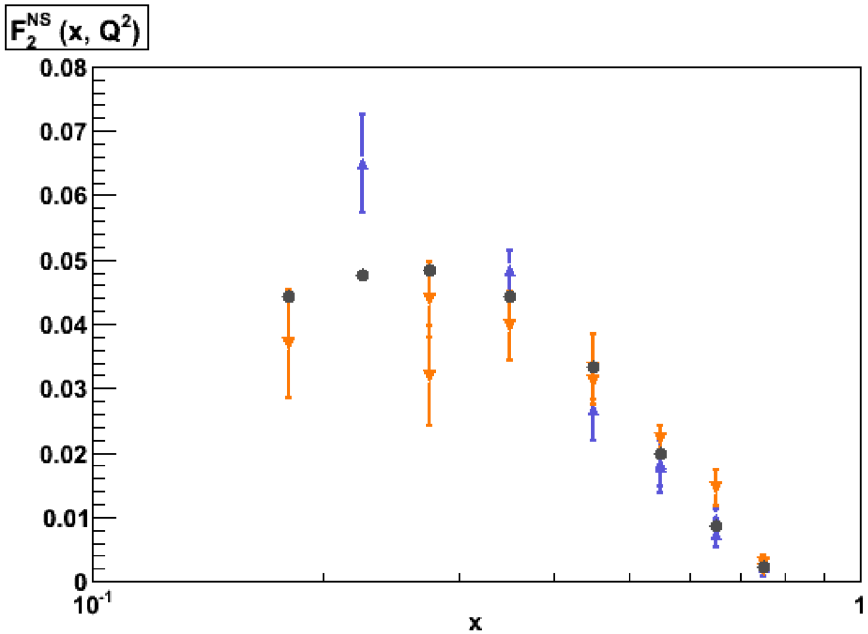
OPTIMAL FIT: CROSS-VALIDATION

GENETIC MINIMIZATION:
AT EACH GENERATION, χ^2 EITHER UNCHANGED OR DECREASING

- DIVIDE THE DATA IN TWO SETS: TRAINING AND VALIDATION
- MINIMIZE THE χ^2 OF THE DATA IN THE TRAINING SET
- AT EACH ITERATION, COMPUTE THE χ^2 FOR THE DATA IN THE VALIDATION SET (NOT USED FOR FITTING)
- WHEN THE VALIDATION χ^2 STOPS DECREASING, STOP THE FIT

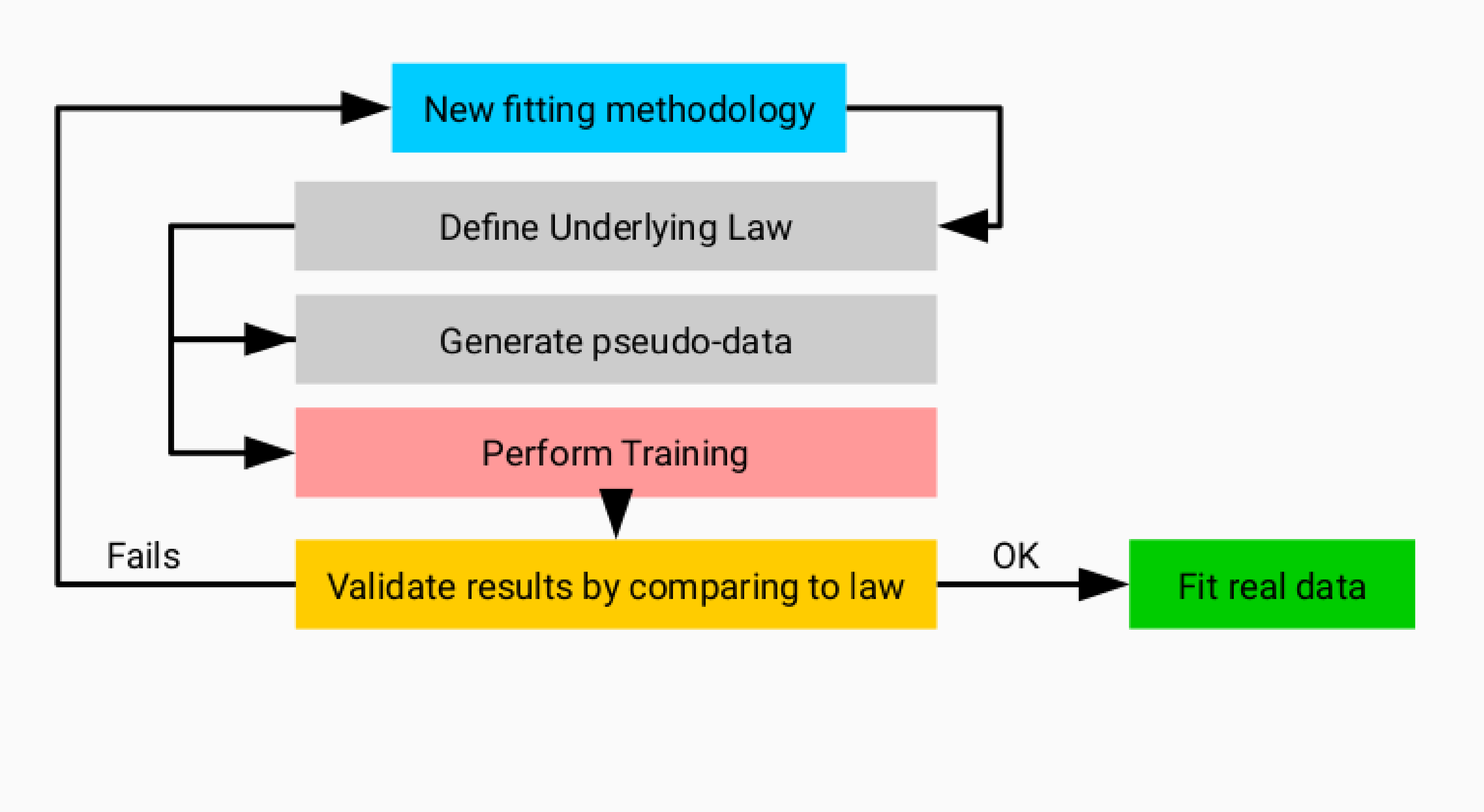
TOO LATE!





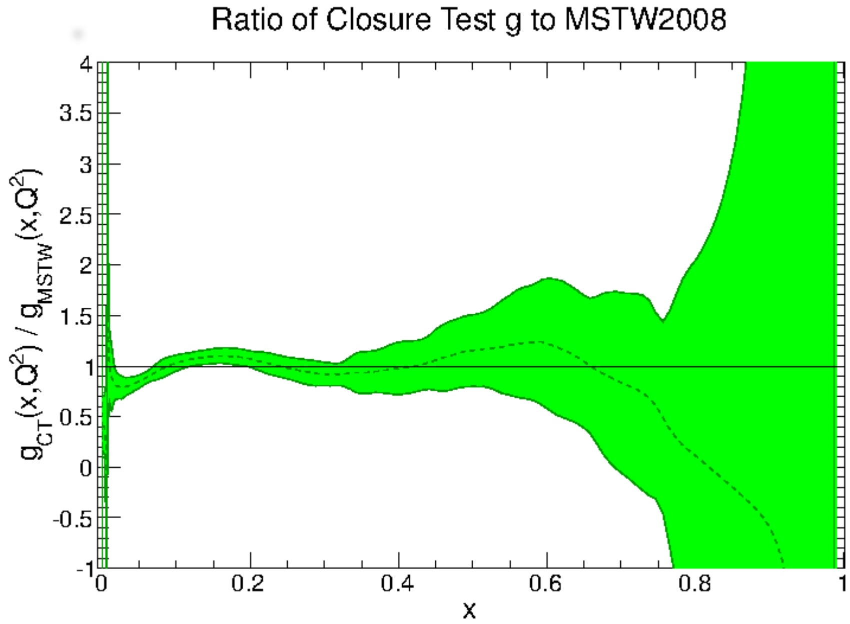
HOW DO WE KNOW THAT WE GOT THE RIGHT ANSWER?

CLOSURE TEST

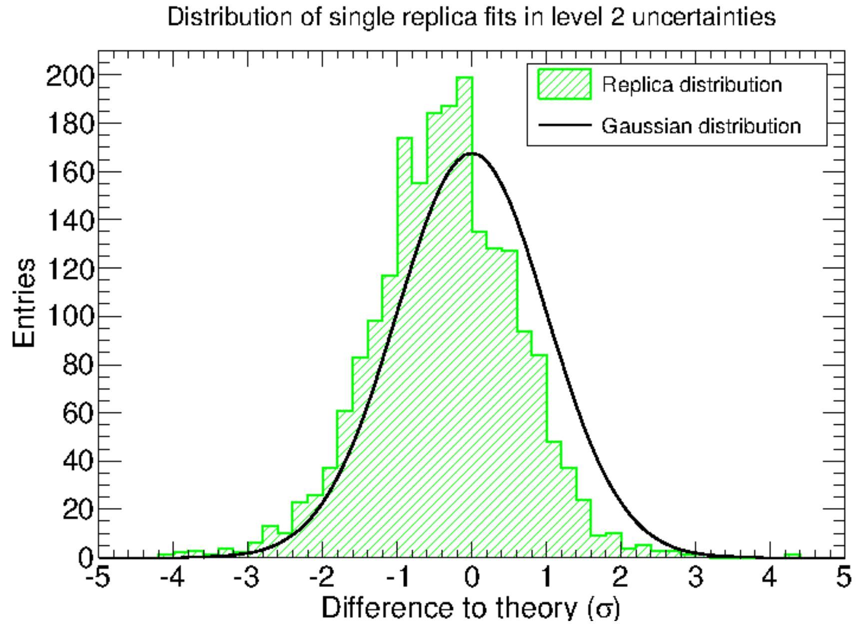


FIRST CLOSURE TEST (NNPDF3.0; 2014)

THE GLUON: RESULT/"TRUTH"



NORMALIZED DISTRIBUTION OF DEVIATIONS

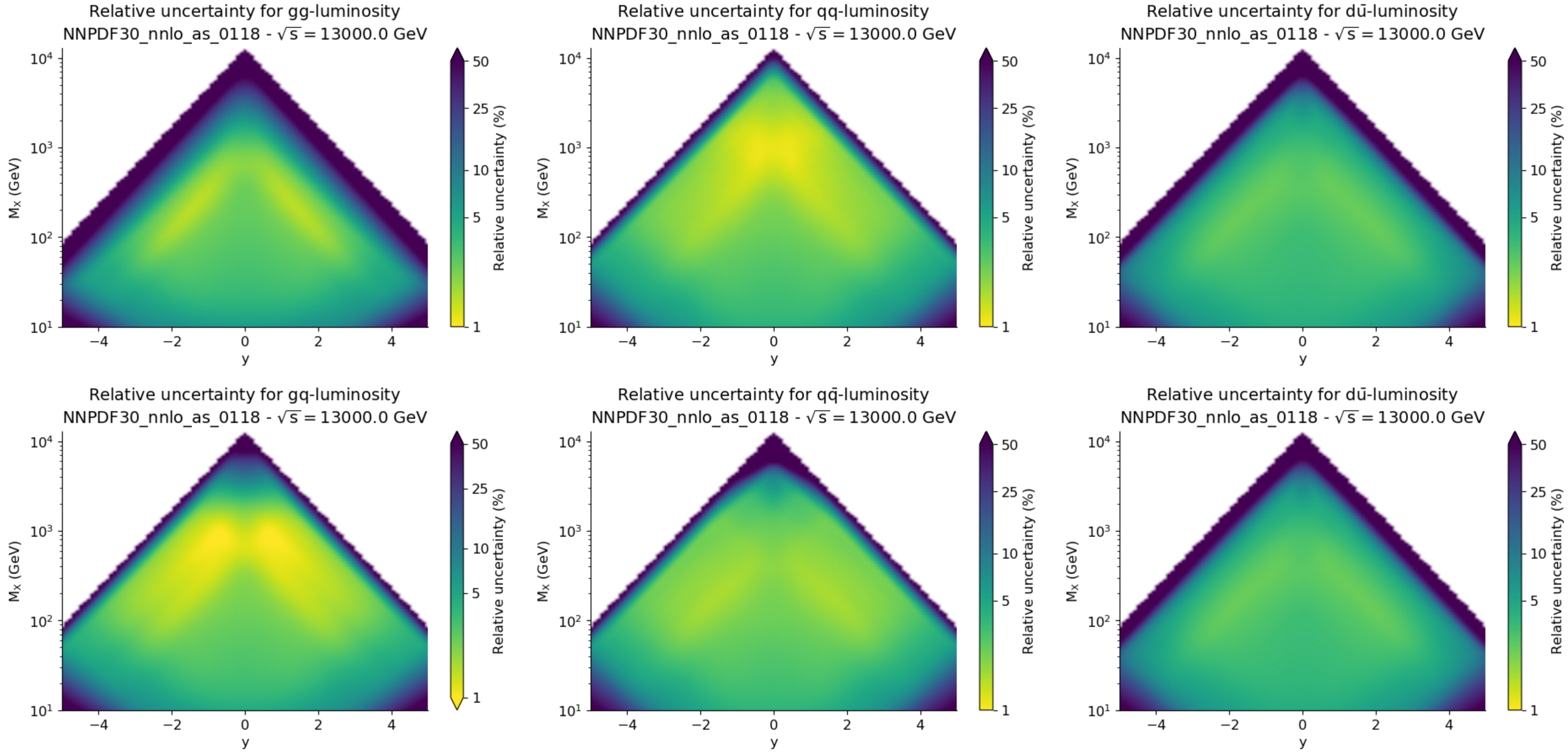


1 σ: 70% (should be 68%)

- THE METHODOLOGY IS FAITHFUL

THE STATE OF THE ART: PRECISION
PDF4LHC PDFs (2014) NNPDF3.0 NNLO

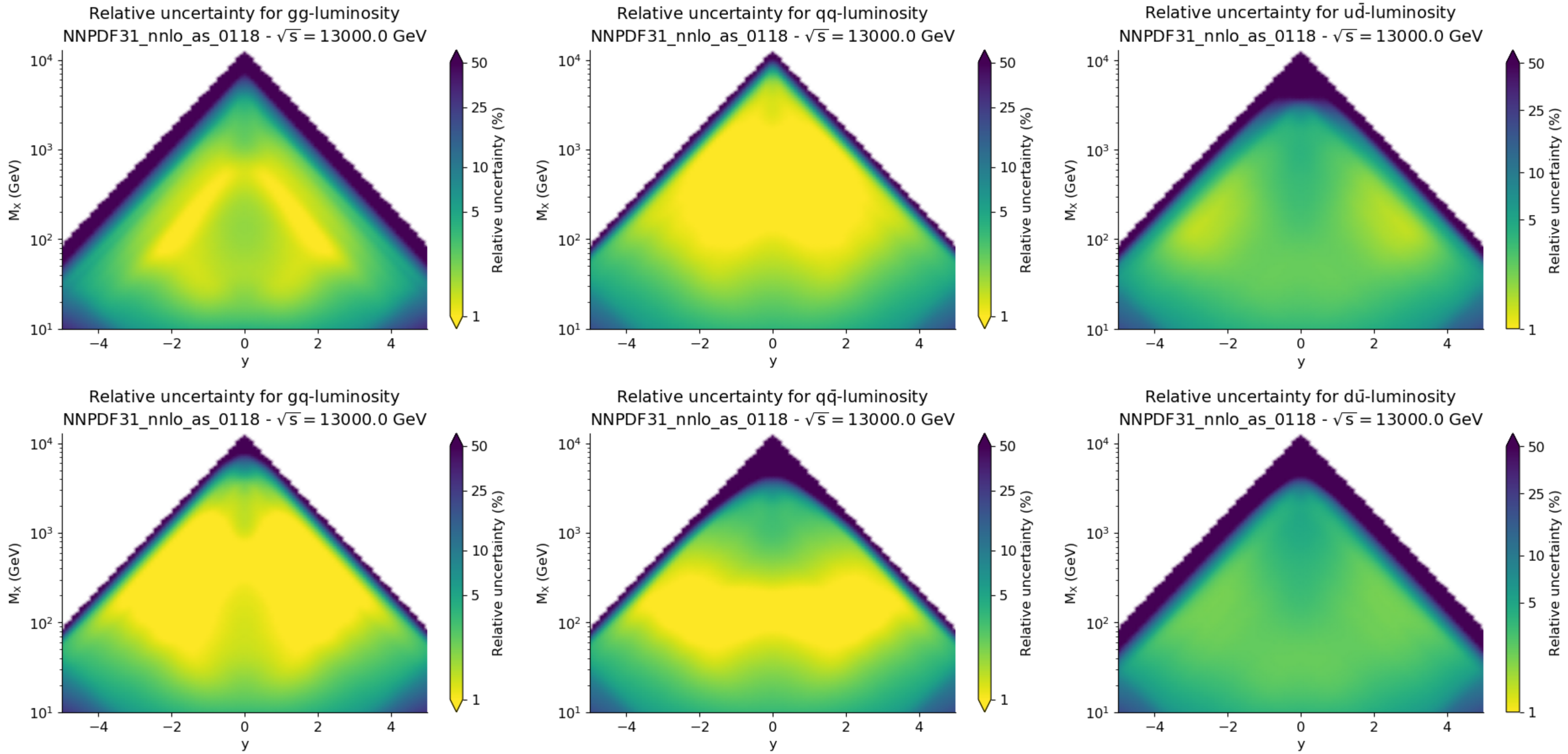
GLUON SINGLET FLAVORS



- GLUON BETTER KNOWN AT SMALL x, VALENCE QUARKS AT LARGE x, SEA QUARKS IN BETWEEN
- TYPICAL UNCERTAINTIES IN DATA REGION $\sim 3-5\%$
- SWEET SPOT: VALENCE Q - G; DOWN TO 1%
- UP BETTER KNOWN THAN DOWN; FLAVOR SINGLET BETTER THAN INDIVIDUAL FLAVORS

THE STATE OF THE ART: PRECISION
CURRENT PDFs (2017) NNPDF3.1 NNLO

GLUON | SINGLET | FLAVORS

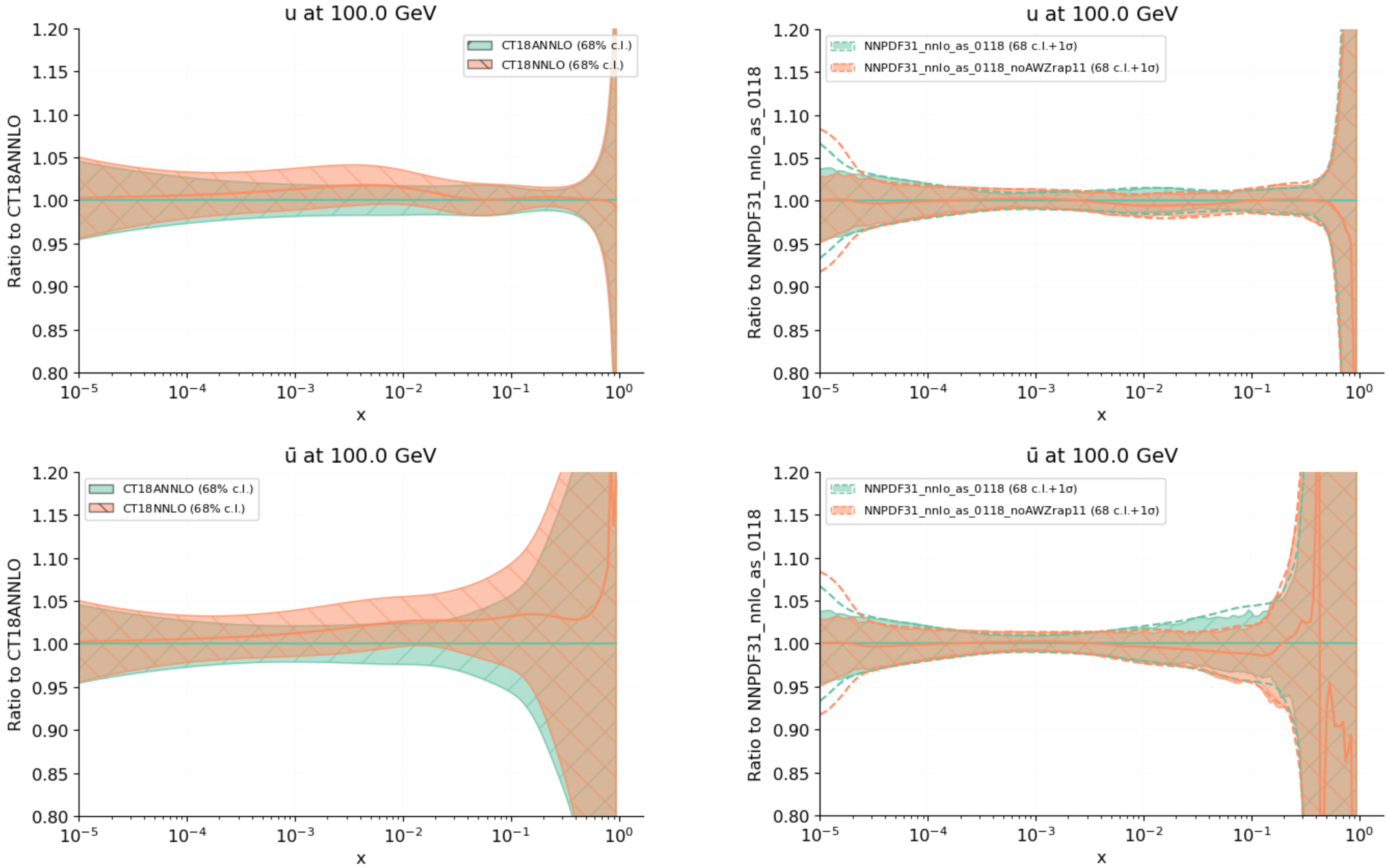


- GLUON BETTER KNOWN AT SMALL x, VALENCE QUARKS AT LARGE x, SEA QUARKS IN BETWEEN
- TYPICAL UNCERTAINTIES IN DATA REGION $\sim 1-3\%$
- SWEET SPOT: VALENCE Q - G; 1% OR BELOW
- UP BETTER KNOWN THAN DOWN; FLAVOR SINGLET BETTER THAN INDIVIDUAL FLAVORS

THE STATE OF THE ART: CONSISTENCY

IMPACT OF ATLAS W/Z 7TeV DATA

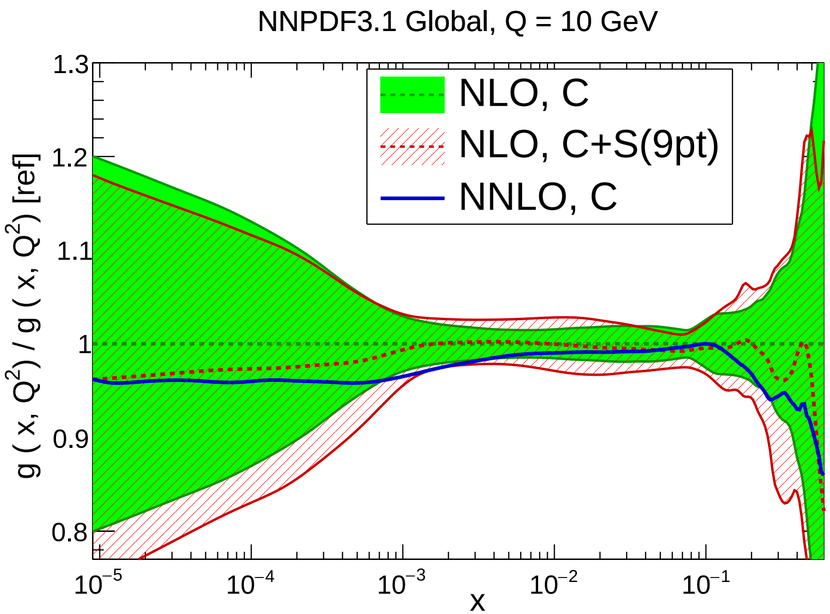
CT18 NNPDF3.1



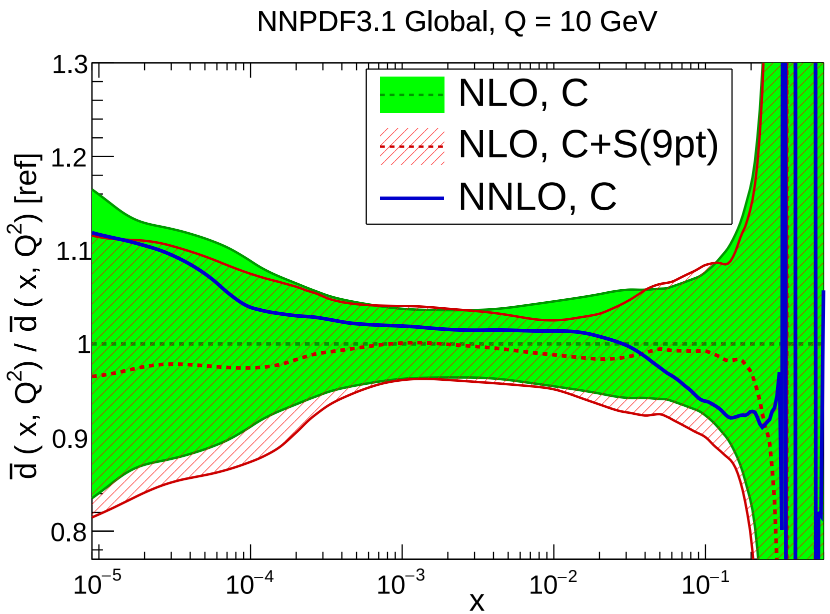
- CT18: PDF SETS RELEASED WITH/WITHOUT ATLAS W/Z DATA INCLUDED
- NNPDF3.1: CONSISTENCY OF ALL DATASETS INCLUDED

THE STATE OF THE ART: ACCURACY
PDFs WITH THEORY UNCERTAINTIES

GLUON



ANTIDOWN



	C	$C + S^{(9\mathrm{pt})}$
χ^2	1.139	1.109
ϕ	0.314	0.415

- FIT QUALITY χ^2 IMPROVES
- RELATIVE ERROR ϕ ON PREDICTION MILDLY INCREASED
- CENTRAL VALUE MOVES TOWARDS KNOWN NNLO

EQUALLY PRECISE BUT MORE ACCURATE RESULT!

THE STATE OF THE ART:

QUESTIONS

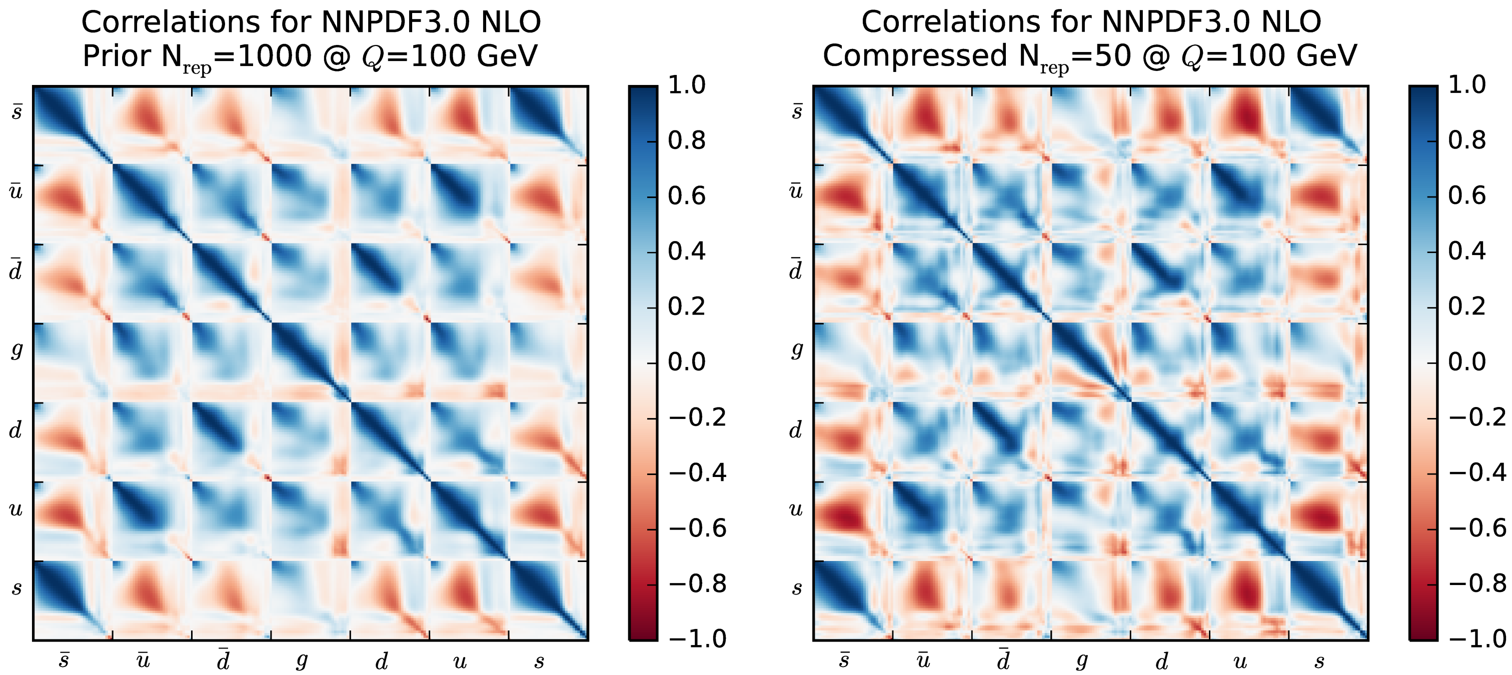
- DO WE REALLY NEED 1000 REPLICAS? OR 100? $\Rightarrow$ EFFICIENCY
- ARE 1000 REPLICAS ENOUGH? OR 10000? $\Rightarrow$ ACCURACY
- PDF UNCERTAINTIES ARE FAITHFUL, BUT ARE THEY OPTIMAL?
  $\Rightarrow$ PRECISION

PDFS FROM AI TO ML

ML: UNSUPERVISED LEARNING
OPTIMIZATION I

- HOW TO MAXIMIZE ACCURACY?
- LARGE (PRIOR) REPLICA SET
- GENETIC SELECTION ⇒ OPTIMIZATION OF STATISTICAL INDICATORS (KULLBACK-LEIBLER DIVERGENCE)
- 50 OPTIMIZES REPLICAS ⇔ 1000 STARTING REPLICAS

CORRELATION MATRIX



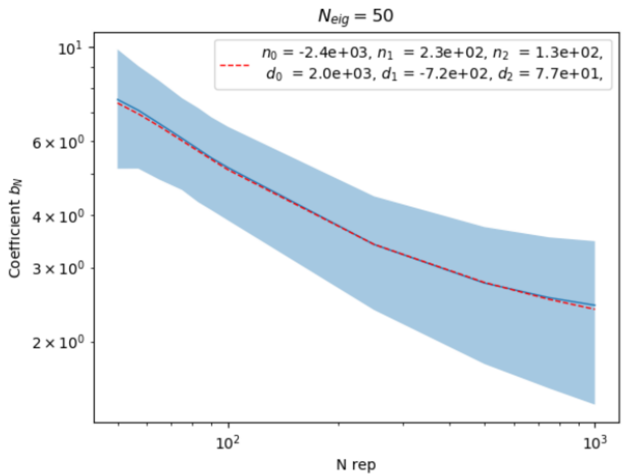
ML: SUPERVISED LEARNING

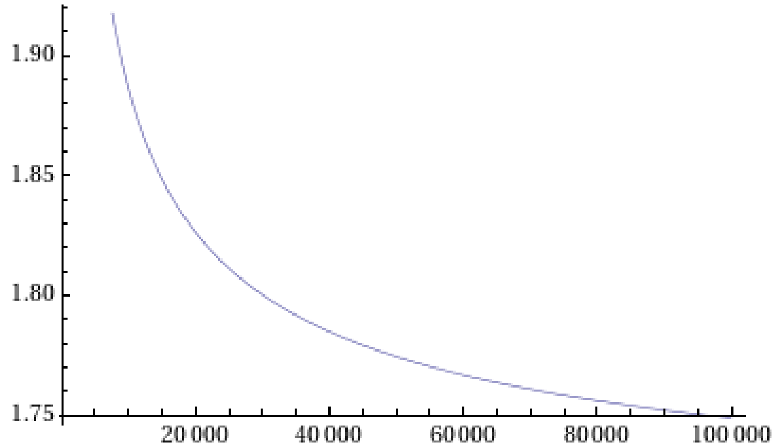
OPTIMIZATION II

HOW MANY PDF REPLICAS DO WE NEED?

FINITE-SIZE EFFECTS

ONE-σ $\Delta\chi^2$ VS NUMBER OF REPLICAS





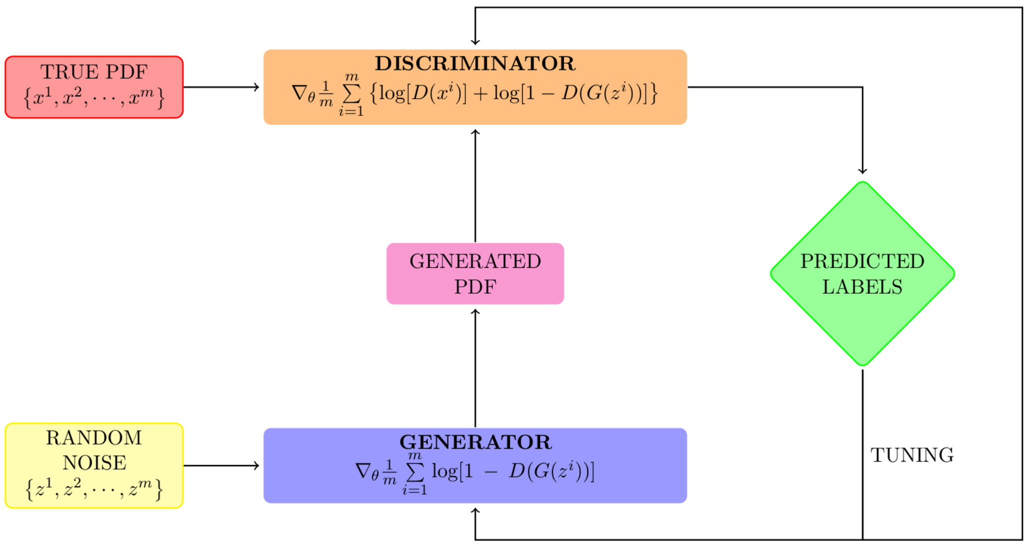
- SIGNIFICANT DEPENDENCE ON NUMBER OF REPLICAS
- ASYMPTOTIC "TOLERANCE" $T = 1.3 \pm 0.3$; $\Delta\chi^2 = 1.7 \pm 0.7$
- FOR $N_{\rm rep} = 100$, $T = 2.3$, EVEN FOR $N_{\rm rep} = 1000$, $T = 1.6$

DO WE HAVE TO FIT 10000 REPLICAS? DO WE HAVE TO USE 10000 REPLICAS?

ML: SUPERVISED LEARNING
OPTIMIZATION II

- CAN WE REDUCE THE NUMBER OF COMPRESSED REPLICAS WITHOUT LOSS OF INFORMATION? SOLUTION FOR USER
- CAN WE INCREASE THE NUMBER OF REPLICAS WITHOUT REFITTING? SOLUTION FOR PDF FITTER

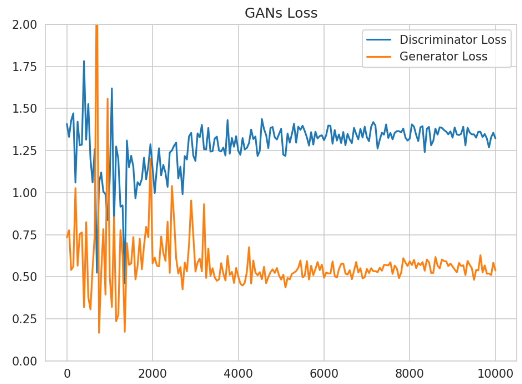
GENERATIVE ADVERSARIAL NETWORKS



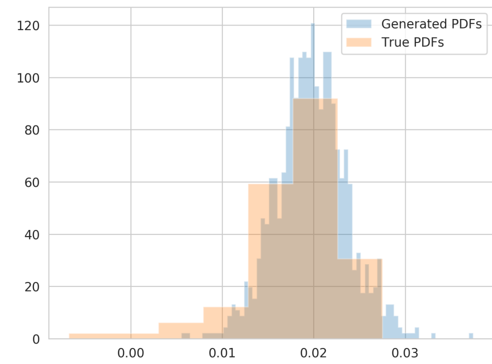
- TRAIN A NETWORK TO SIMULATE THE TRUE DISTRIBUTION (GENERATOR)
- TRAIN A NETWORK TO DISCRIMINATE TRUTH FROM SIMULATION (DISCRIMINATOR)
- TRAIN THE GENERATOR TO TRICK THE DISCRIMINATOR

SOLVING THE PROBLEM....
GAN REPLICA GENERATION

GAN TRAINING



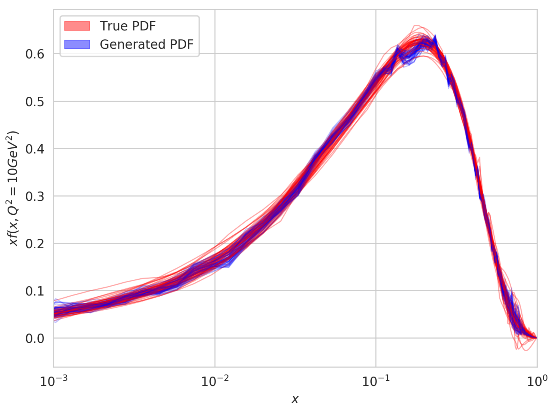
UP VALENCE AT FIXED x



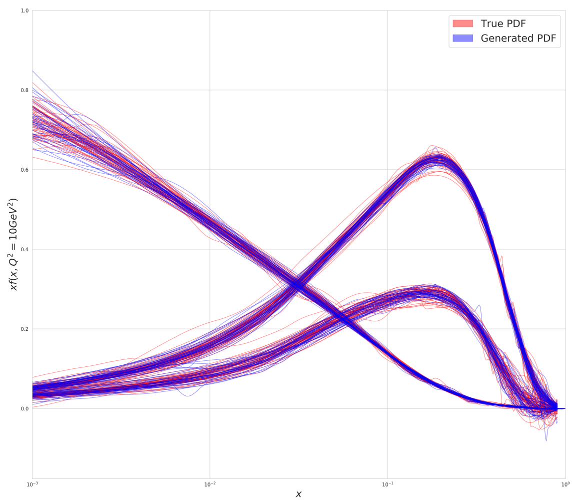
- 1D GAN: REPRODUCE THE INFORMATION IN THE UNDERLYING REPLICA SET, BUT NO GAIN (WIGGLY REPLICAS)
  $\Rightarrow$ REDUCE THE NUMBER OF COMPRESSED REPLICA WITH FIXED NUMBER OF FITTED REPLICAS W/O INFORMATION LOSS
- 2D GAN: COMBINE CORRELATED INFORMATION FROM UNDERLYING REPLICA SET INFERRING THE TRUE UNDERLYING DISTTRIBUTION
  $\Rightarrow$ REDUCE THE NUMBER OF INPUT REPLICAS W/O INFORMATION LOSS

N3PDF
Machine Learning • PDFs • QCD

ONE-DIMENSIONAL



TWO-DIMENSIONAL

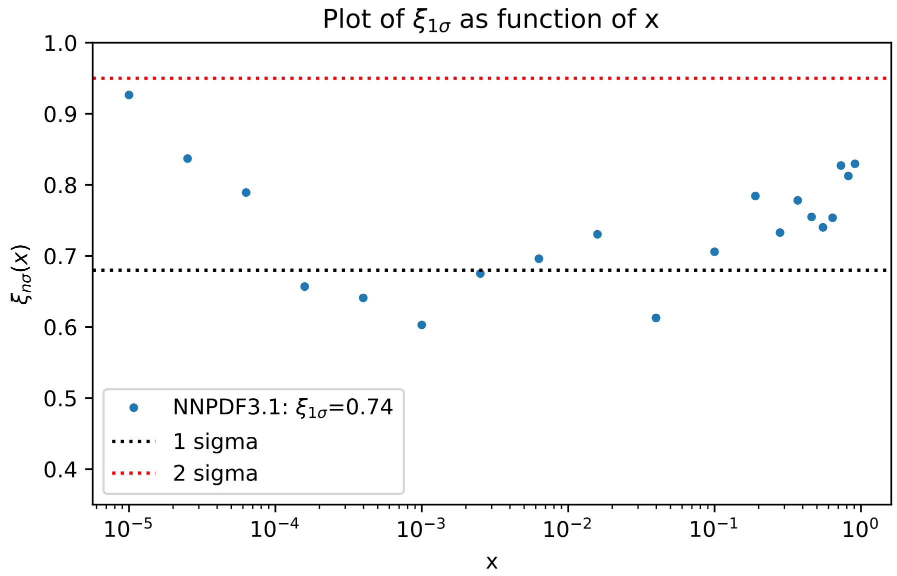


CLOSURE TEST: A CLOSER LOOK (NNPDF3.1)

ONE σ: ACTUAL/PREDICTED
FOR DATA, BY EXPERIMENT

experiment	NNPDF3.1 ratio
NMC	0.882828
SLAC	0.767063
BCDMS	0.730569
CHORUS	0.698907
NTVDMN	0.991090
HERACOMB	0.847359
HERAF2CHARM	1.867597
F2BOTTOM	1.124157
DYE886	0.655955
DYE605	0.585725
CDF	0.961652
D0	0.881199
ATLAS	0.904127
CMS	1.090241
LHCb	1.092194
Total	0.842168

ONE σ VALUE
FOR PDFS, VS x



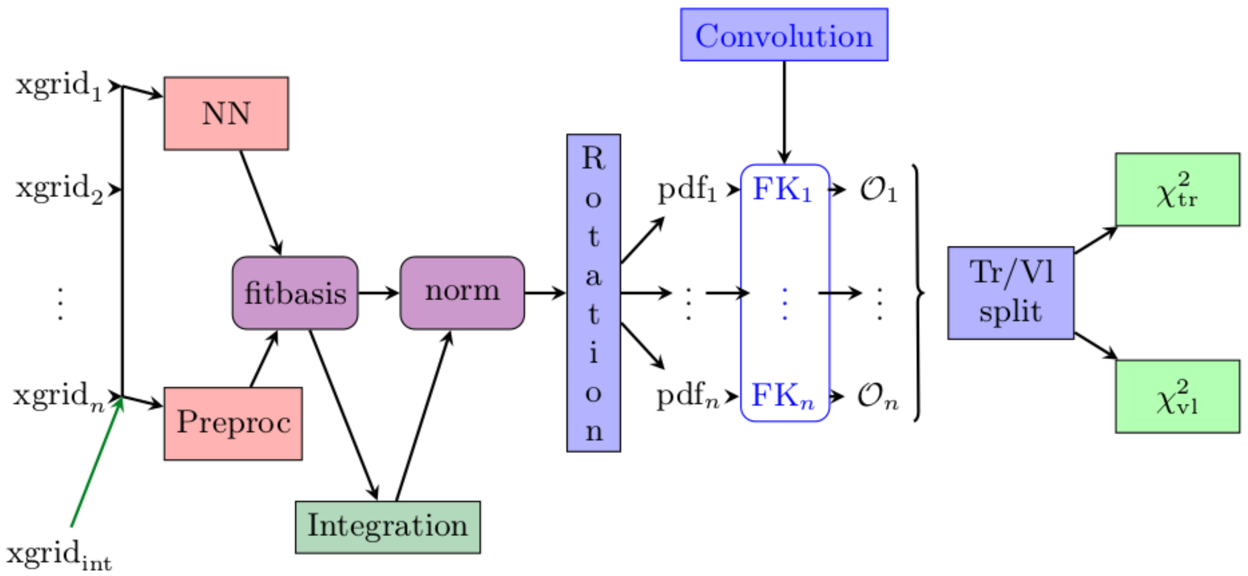
- UNCERTAINTIES OVERESTIMATED
- 1 σ>68% AT VERY SMALL AND VERY LARGE x;
  1 σ <68% AT INTERMEDIATE x

CAN WE DO BETTER?

THE N3FIT PROJECT

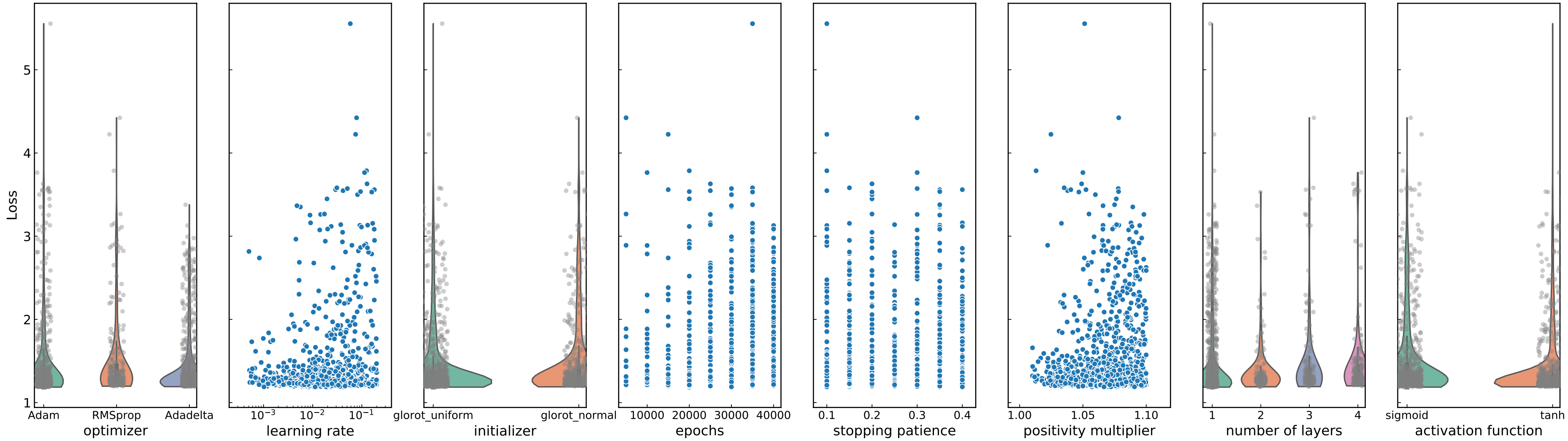
HOW DO WE KNOW THAT THE METHODOLOGY IS THE BEST?
"ACCUMULATED WISDOM" INEFFICIENT AND SLOW

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



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

FITTING THE METHODOLOGY
HYPEROPTIMIZATION SCANS



HYPEROPT PARAMETERS

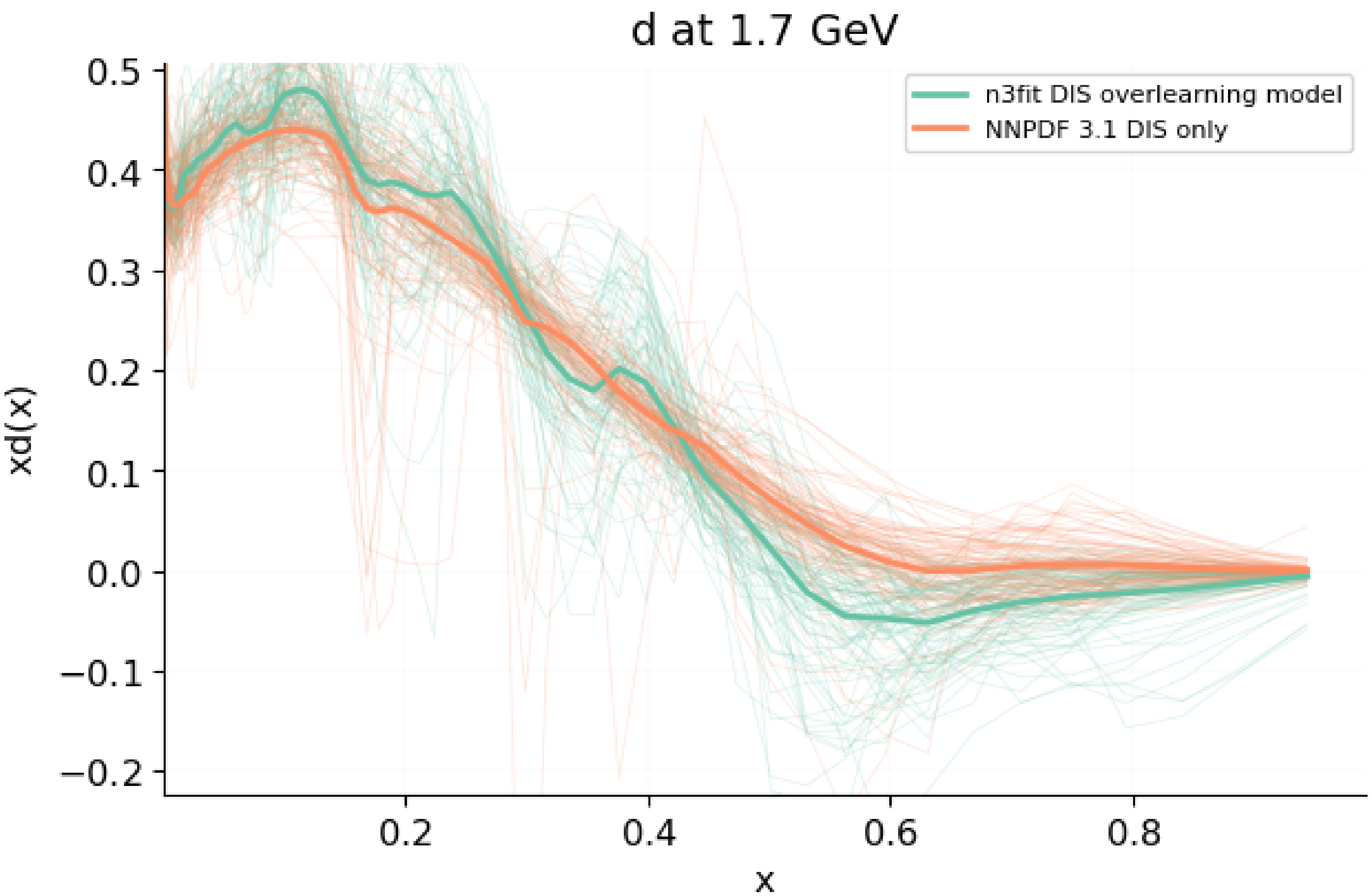
NEURAL NETWORK	FIT OPTIONS
NUMBER OF LAYERS (*)	OPTIMIZER (*)
SIZE OF EACH LAYER	INITIAL LEARNING RATE (*)
DROPOUT	MAXIMUM NUMBER OF EPOCHS (*)
ACTIVATION FUNCTIONS (*)	STOPPING PATIENCE (*)
INITIALIZATION FUNCTIONS (*)	POSITIVITY MULTIPLIER (*)

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

FITTING THE METHODOLOGY

THE OVERFITTING PROBLEM

DOWN QUARK: HYPEROPTIMIZED VS. STANDARD



- NNPDF3.1: 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?

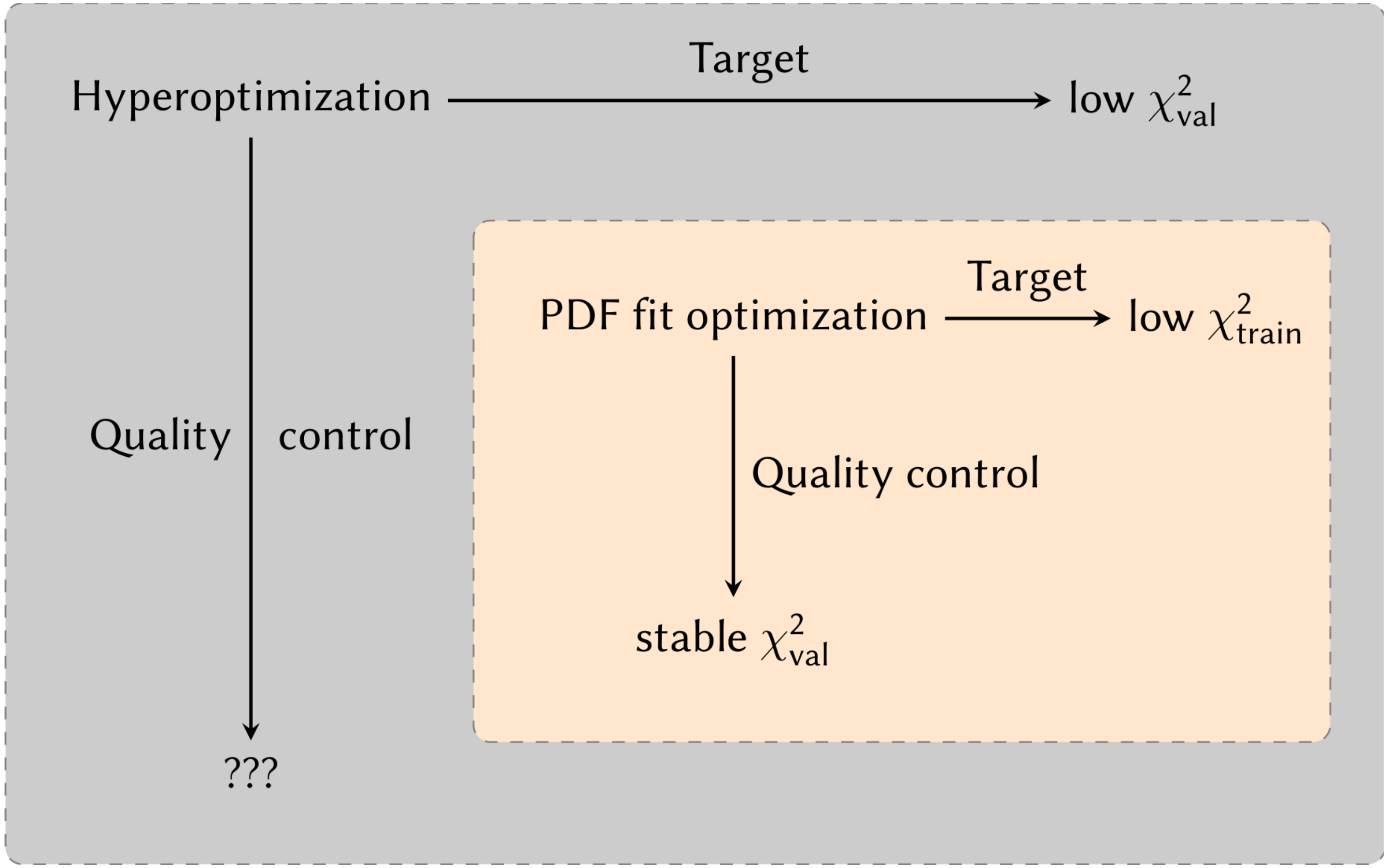
OPTIMIZATION

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

CROSS-VALIDATION SELECTS THE OPTIMAL MINIMUM

WHAT HAPPENED?

HYPEROPTIMIZATION

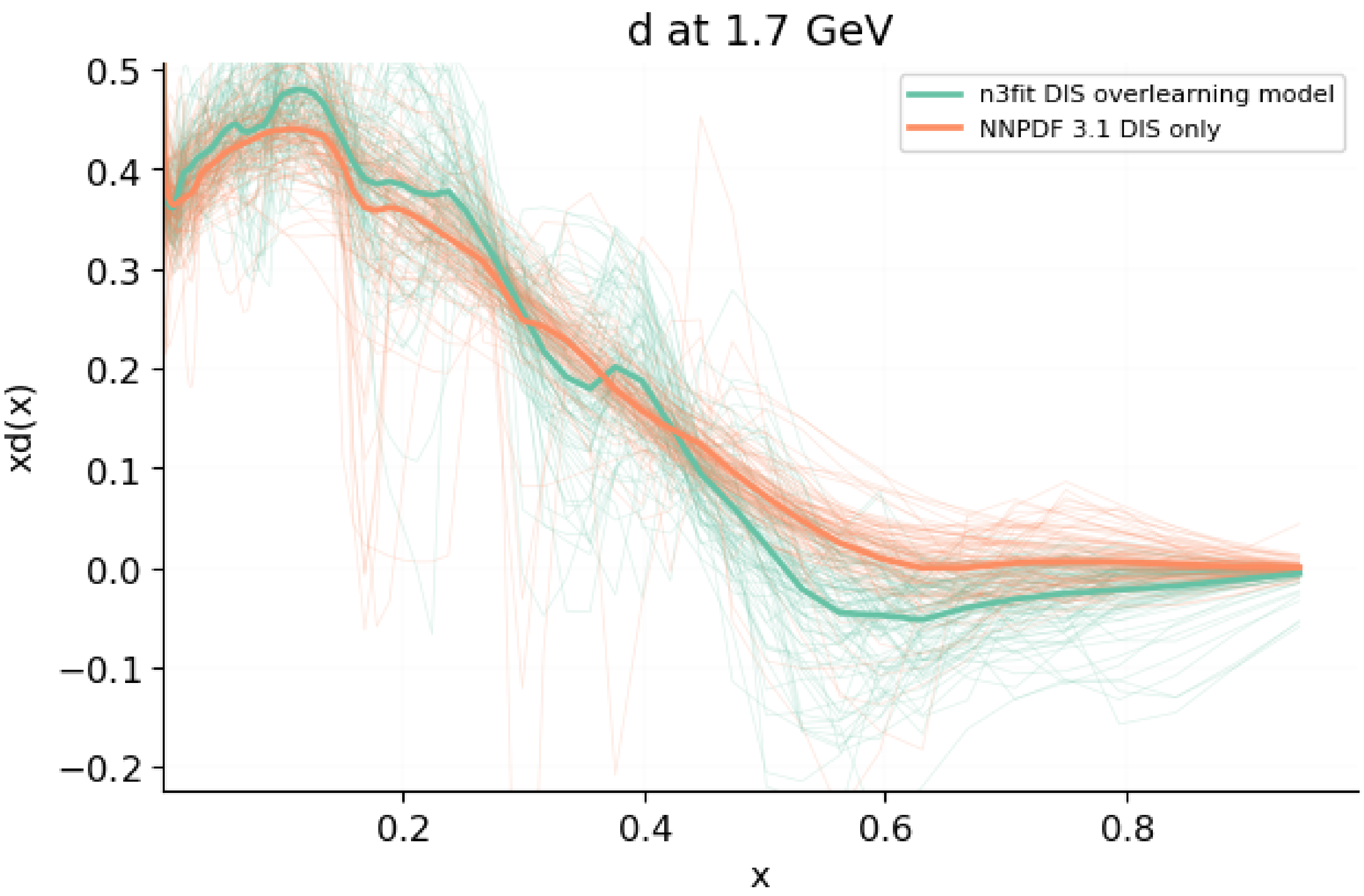


WE ARE MISSING A SELECTION CRITERION

FITTING THE METHODOLOGY

THE OVERFITTING PROBLEM

DOWN QUARK: HYPEROPTIMIZED VS. STANDARD

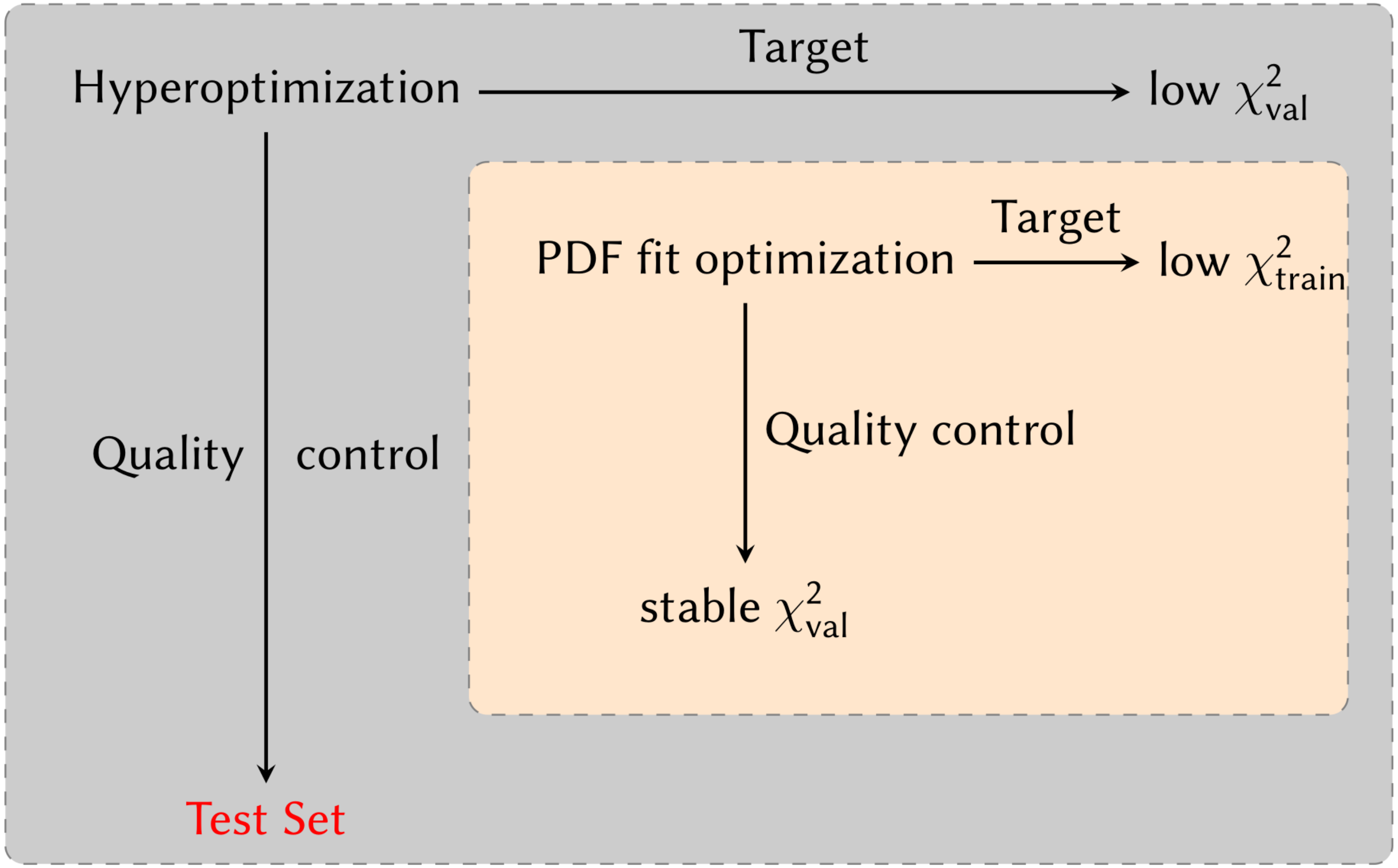


- NNPDF3.1: WIGGLES: FINITE SIZE $\Rightarrow$ WILL GO AWAY AS $N_{\rm rep}$ GROWS
- N3FIT: WIGGLY PDFS $\Leftrightarrow$ OVERFITTING $\Rightarrow$ WILL NOT GO AWAY ($\chi^2_{\rm train} \ll \chi^2_{\rm valid}$!!)
- CORRELATIONS BETWEEN TRAINING AND VALIDATION DATA

MACHINE LEARNING

THE SOLUTION

TUNED HYPEROPTIMIZATION



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

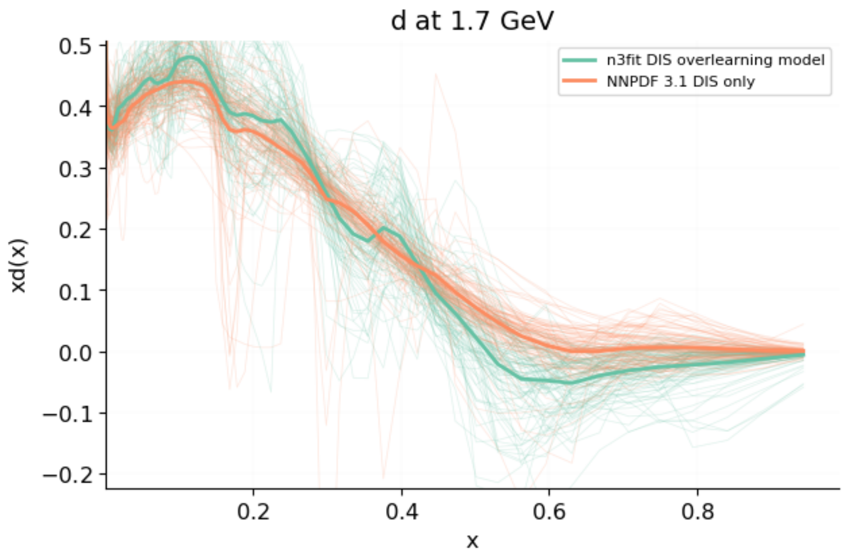
THE TEST SET METHOD

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

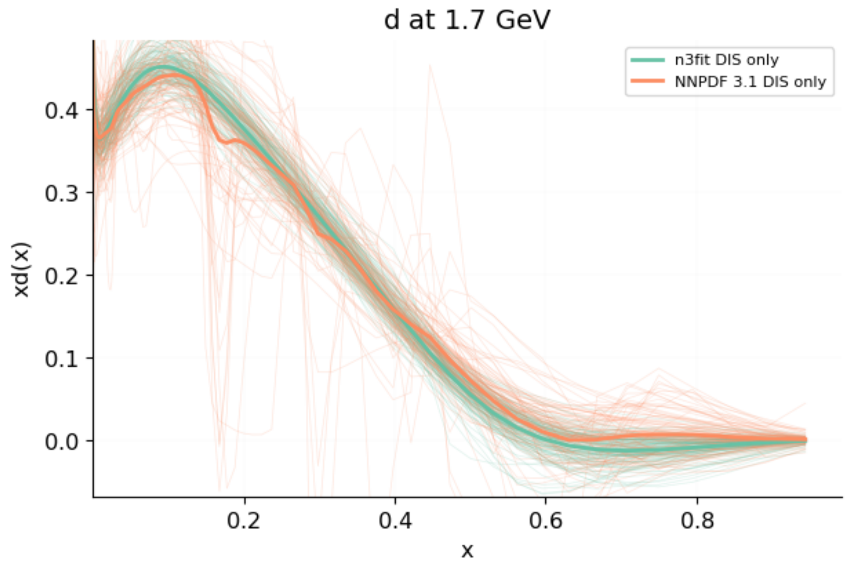
OPTIMIZED PDFs

DOWN QUARK

N3 OVERFIT VS NNPDF3.1



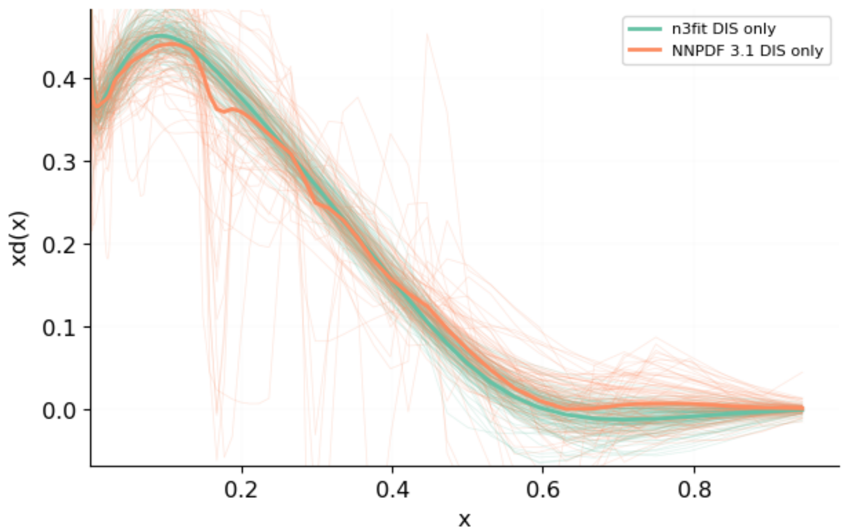
N3FIT VS NNPDF3.1



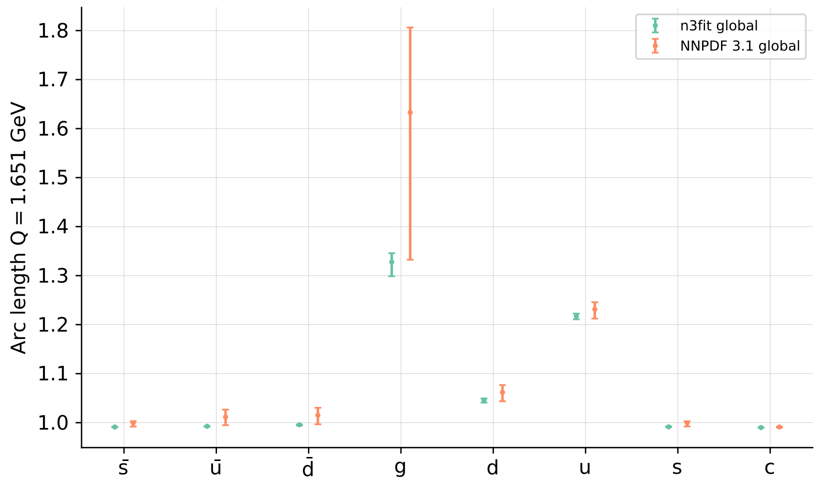
THE TEST SET METHOD

N3FIT vs NNPDF3.1

down PDF



ARCLENGTHS



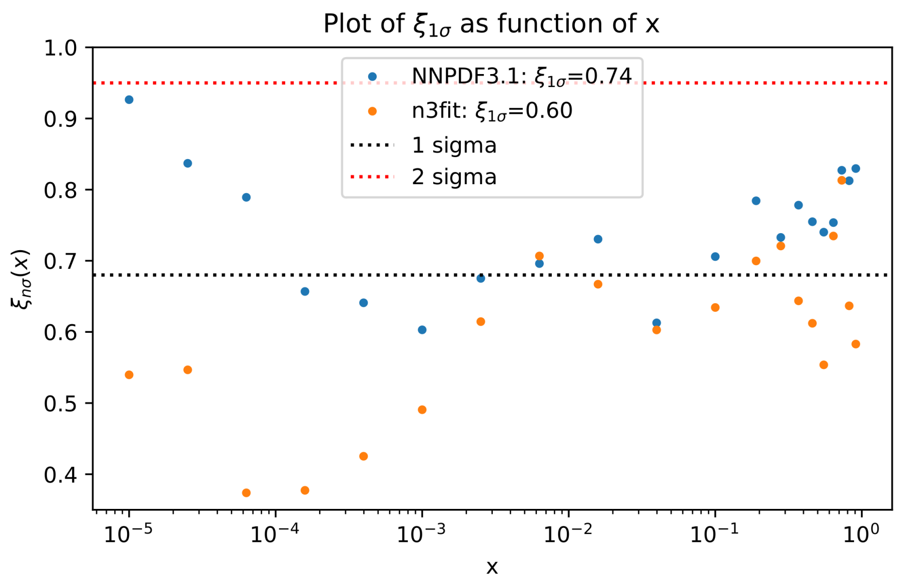
- NO OVERFITTING
- COMPARED TO NNPDF3.1
  - MUCH GREATER STABILITY $\Rightarrow$ FEWER REPLICAS FOR EQUAL ACCURACY
  - UNCERTAINTIES SOMEWHAT REDUCED

CLOSURE TESTS AGAIN

ONE σ: ACTUAL/PREDICTED
FOR DATA, BY EXPERIMENT

experiment	NNPDF3.1 ratio	n3fit ratio
NMC	0.882828	0.843427
SLAC	0.767063	0.690118
BCDMS	0.730569	0.770704
CHORUS	0.698907	0.734656
NTVDMN	0.991090	0.797017
HERACOMB	0.847359	1.326333
HERAF2CHARM	1.867597	3.566076
F2BOTTOM	1.124157	1.532634
DYE886	0.655955	0.857915
DYE605	0.585725	0.870151
CDF	0.961652	0.779424
D0	0.881199	1.015202
ATLAS	0.904127	1.132229
CMS	1.090241	1.017136
LHCb	1.092194	0.993525
Total	0.842168	0.940737

ONE σ VALUE
FOR PDFS, VS x



- UNCERTAINTIES WELL ESTIMATED ON AVERAGE;
  BUT SIZABLE FLUCTUATIONS
- ONE σ PERFECT IN DATA REGION;
  BUT UNDERESTIMATED IN EXTRAPOLATION

BEYOND THE STATE OF THE ART:

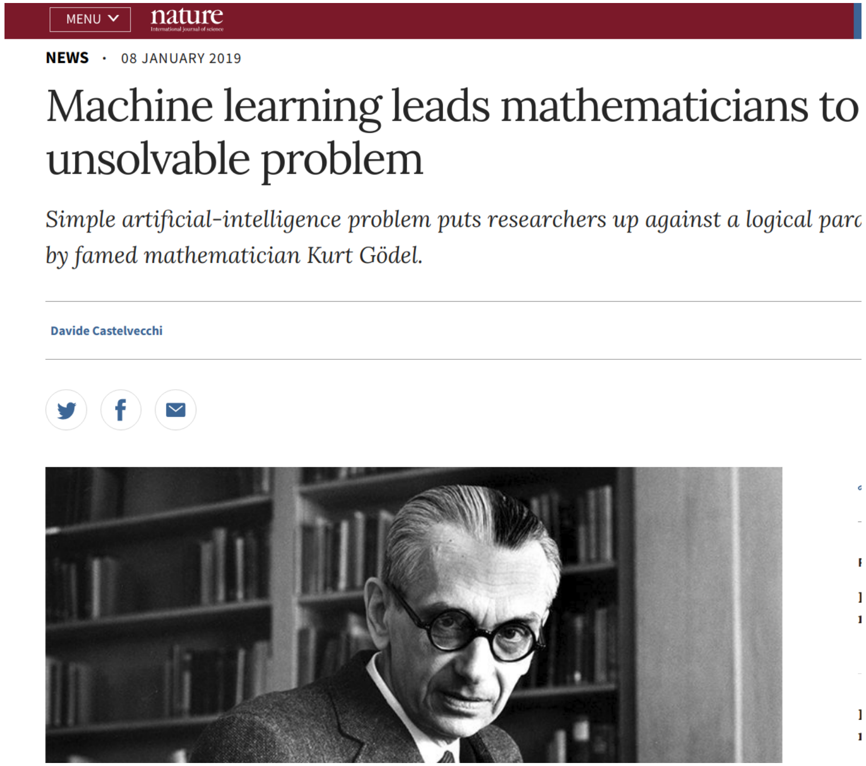
DREAMS

- WHAT IS THE UNCERTAINTY WHERE THERE IS NO DATA?
- WHAT IS THE UNCERTAINTY WHERE THERE IS NO THEORY?

ML THE UNKNOWN

WHAT IS "PROPER LEARNING"?

FORECASTING AN UNKNOWN TRUTH $\Rightarrow$ WHAT IS "OPTIMAL"?


MENU nature

NEWS · 08 JANUARY 2019

Machine learning leads mathematicians to unsolvable problem

Simple artificial-intelligence problem puts researchers up against a logical parc by famed mathematician Kurt Gödel.

Davide Castelvecchi

SOME POSSIBLE ANSWERS/CRITERIA

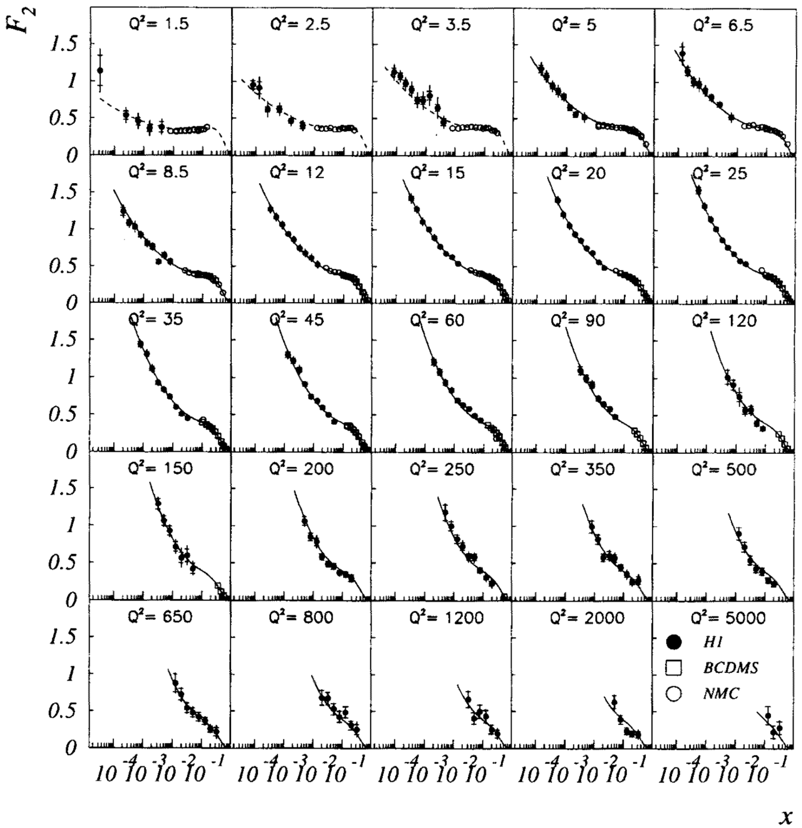
- PASS A CLOSURE TEST
- REPRODUCE THE EXPECTED STATISTICAL PROPERTIES: ONE $\sigma \Leftrightarrow \Delta\chi^2 = 1$
- SATISFY THEORETICAL PREJUDICE?

PASS A "FUTURE TEST":
GENERALIZE TO CURRENT DATA BASED ON PAST DATA

THE “FUTURE TEST”

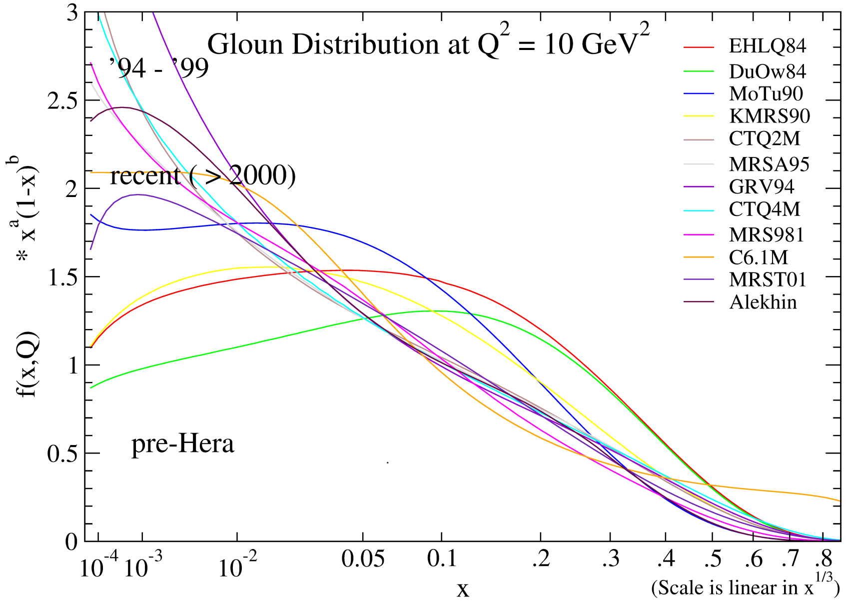
1995: THE RISE OF STRUCTURE FUNCTIONS AT HERA

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 PREJUDICE

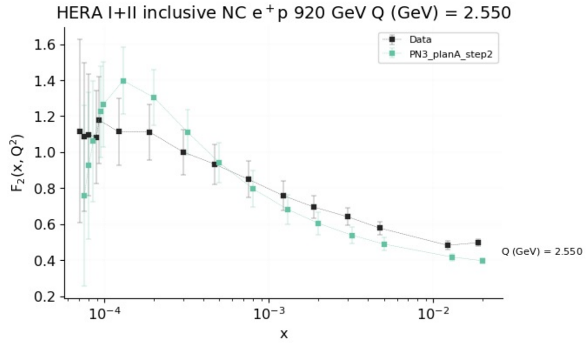
COULD WE HAVE PREDICTED IT?

THE N3FIT FUTURE TEST

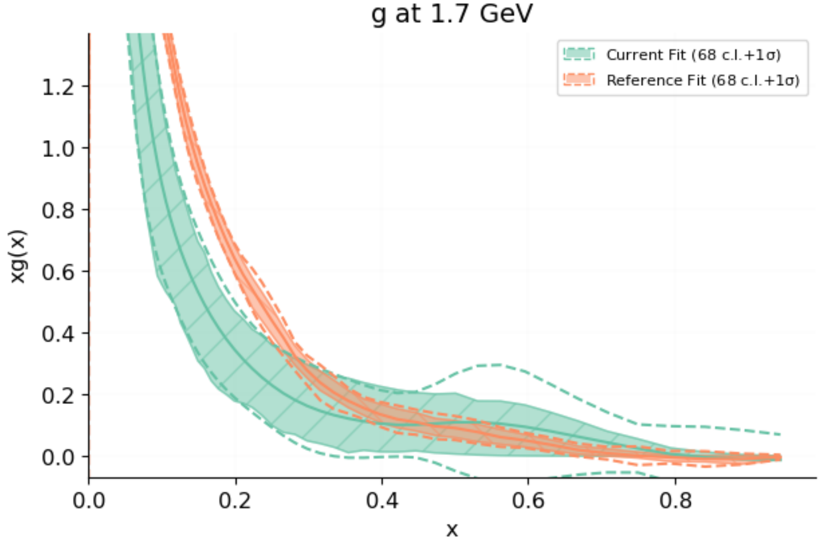
ONLY PRE-HERA DATA USED

PREDICTION COMPARED TO DATA

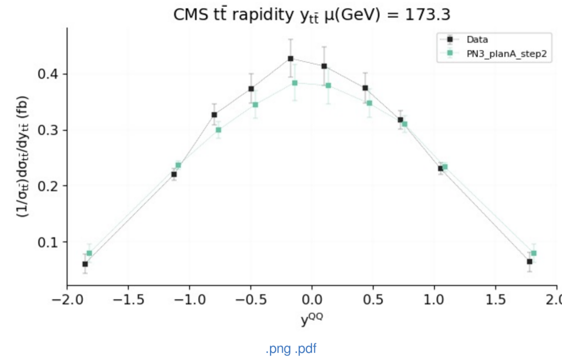
HERA F_2



PREDICTED VS TRUE GLUON



CMS TOP



- N3FIT METHDOLOGY APPLIED AND HYPEROPTIMIZED TO PRE-HERA DATASET
- RESULTS WITH PDF UNCERTAINTY COMPARED TO FUTURE DATA
- χ^2/dat=1.1 ON FULL PREDICTED CURRENT DATASET (ABOUT 200 DATAPOINTS)

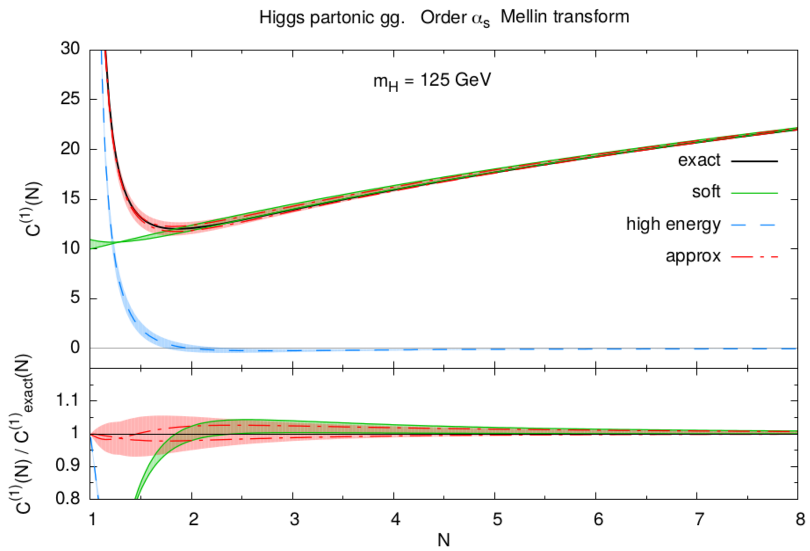
SUCCESS!

HOWEVER.... PREPROCESSING $\Rightarrow$ TUNED METHODOLOGY

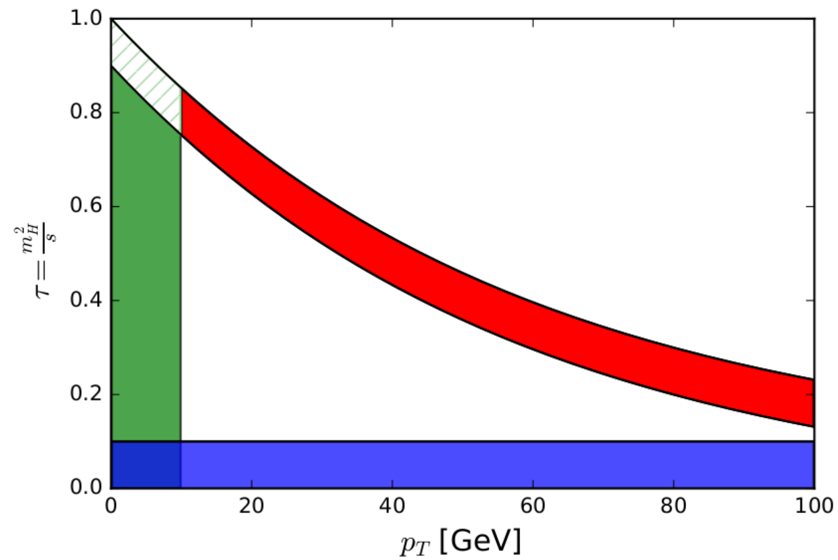
- GAUSSIAN PROCESSES? (KERNEL METHODS)
- REINFORCEMENT LEARNING

MISSING HIGHER ORDERS FROM RESUMMATION

N-SPACE GGHIGGS: APPROX VS. EXACT



(τ, p_T) RESUMMATION REGIONS



- THEORY UNCERTAINTIES $\Leftrightarrow$ APPROXIMATE NEXT ORDER
- RESUMMATION $\Rightarrow$ SINGULARITIES
- MATCHING THROUGH LSTM? (RECURRENT NN)

THE WORK OF MANY PEOPLE

NNPDF collaboration and N^{3}PDF team meeting,
Varenna, Italy, September 2019

"Io stimo più il trovare un vero, benché di cosa leggiera, che il disputar lungamente delle massime questioni senza verità nissuna"

"I am more interested in uncovering a fact, however trifling, than to dispute at length about profound questions devoid of any truth"

Galileo Galilei, letter to Tommaso Campanella

EXTRAS

CONTEMPORARY PDF TIMELINE (ONLY PUBLISHED GLOBAL)

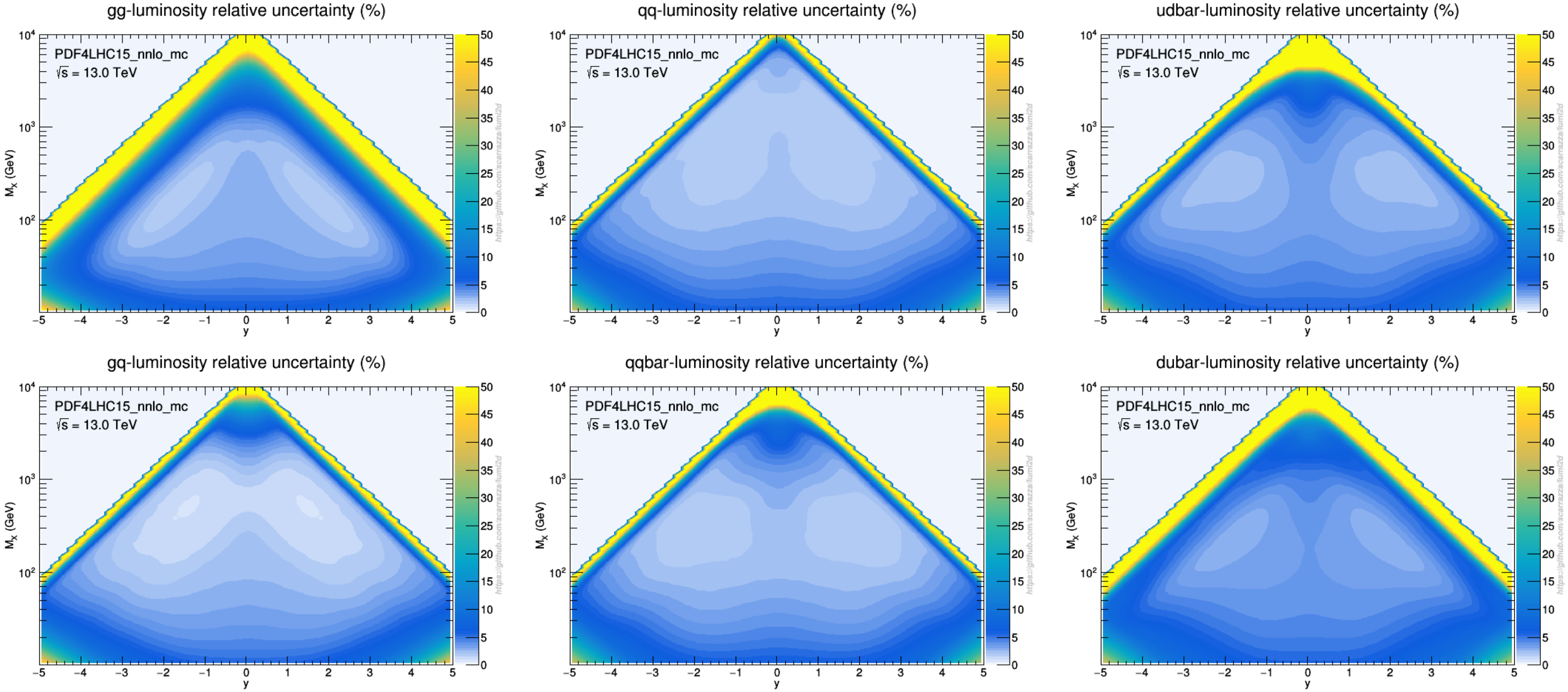
SET / MONTH	2008 CTEQ6.6 (02)	2008 NNPDF1.0 (08)	2009 MSTW (01)	2009 ABKM09 (08)	2010 NNPDF2.0 (02)	2010 CT10 (NLO) (07)	2011 NNPDF2.1 (NNLO) (07)	2012 ABM11 (02)	2012 NNPDF2.3 (07)	2013 CT10 (NNLO) (02)	2013 ABM12 (10)	2014 NNPDF3.0 (10)	2014 MMHT (12)	2015 CT14 (06)	2017 ABMP16 (01)	2017 NNPDF3.1 (06)	2019 CT18 (12)
F. T. DIS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
ZEUS+H1-HI	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
COMB. HI	✗	✗	✗	✗	✔	✗	✔	✗	✔	✗	✔	✔	✗	✗	✔	✔	✔
ZEUS+H1-HII	✗	✗	✗	✗	✗	✗	some	✗	✗	some	✗	✔	✗	✗	✔	✔	✔
HERA JETS	✗	✗	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✗
F. T. DY	✔	✗	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
TEV W+Z	✔	✗	✔	✗	✔	✔	✔	✗	✔	✔	✗	✔	✔	✔	✗	✔	✔
LHC W+Z	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	some	✔	✔	✔	some	✔	✔
TEV JETS	✔	✗	✔	✗	✔	✔	✗	✔	✔	✔	✗	✔	✔	✔	✗	✔	✔
LHC JETS	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✔	✔	✔	✗	✔	✔
TOP TOTAL	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✔	✗	✗	✔	✔	✔
SINGLE TOP TOTAL	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗
TOP DIFFERENTIAL	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✔
W p_T	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✗	✗
W+C	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✗	✗
Z p_T	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✔

THEORY PROGRESS:

- MSTW, ABKM: all NNLO; NNPDF NNLO since 07/11 (2.1), CT since 02/13 (CT10); NNPDF THRESHOLD RESUMMATION (3.0RESUM, 07/15), SMALL x RESUMMATION (3.1SX, 10/17)
- MSTW, CT, NNPDF all GM-VFN; NNPDF since 01/11 (2.1); ABM FFN+ZM-VFN since 01/17 (ABMP16)
- NNPDF FITTED CHARM since 05/16 (NNPDF3IC)
- PHOTON PDF: (mrst2004qed), NNPDF2.3QED (08/13), NNPDF3.0QED (06/16), NNPDF3.1LUXQED (12/17)

PDF4LHC15: PDF UNCERTAINTIES (NNLO)

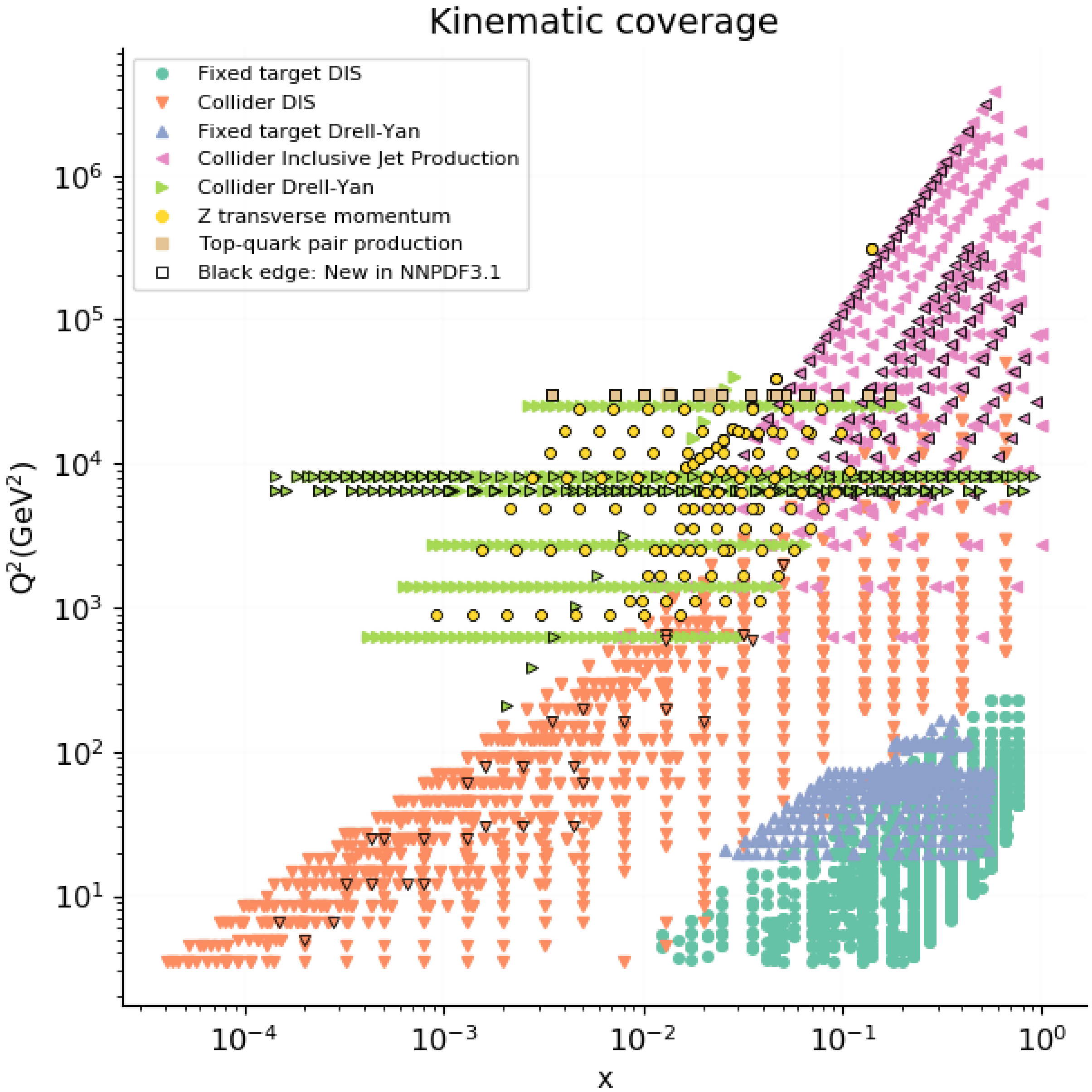
GLUON SINGLET FLAVORS



- GLUON BETTER KNOWN AT SMALL x, VALENCE QUARKS AT LARGE x, SEA QUARKS IN BETWEEN
- TYPICAL UNCERTAINTIES IN DATA REGION $\sim 3 - 5\%$
- SWEET SPOT: VALENCE Q - G; DOWN TO 1%
- UP BETTER KNOWN THAN DOWN; FLAVOR SINGLET BETTER THAN INDIVIDUAL FLAVORS
- NO QUALITATIVE DIFFERENCE BETWEEN NLO AND NNLO

DATASET WIDENING

NNPDF3.0 VS NNPDF3.1



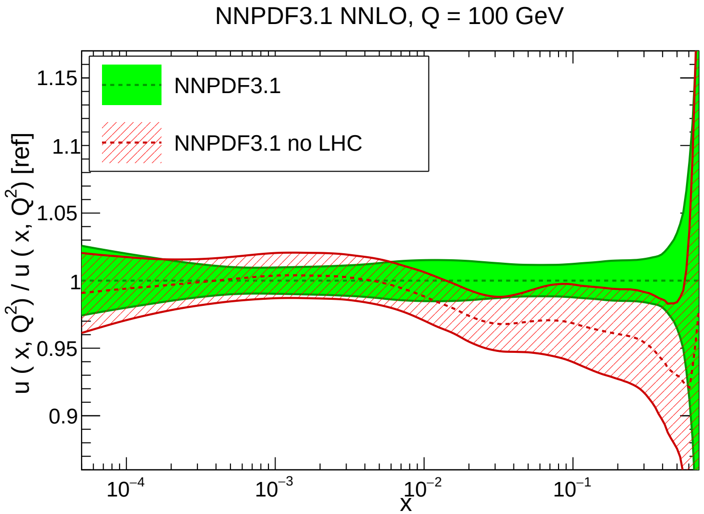
NEW DATA: (BLACK EDGE)

- HERA COMBINED F_2^b
- D0 W LEPTON ASYMMETRY
- ATLAS W, Z 2011, HIGH & LOW MASS DY 2011; CMS $W^{\pm}$ RAPIDITY 8TEV LHCB W, Z 7TEV & 8TEV
- ATLAS 7TEV JETS 2011, CMS 2.76TEV JETS
- ATLAS & CMS TOP DIFFERENTIAL RAPIDITY
- ATLAS Z p_T DIFFERENTIAL RAPIDITY & INVARIANT MASS 8TEV, CMS Z p_T DIFFERENTIAL RAPIDITY 8TEV

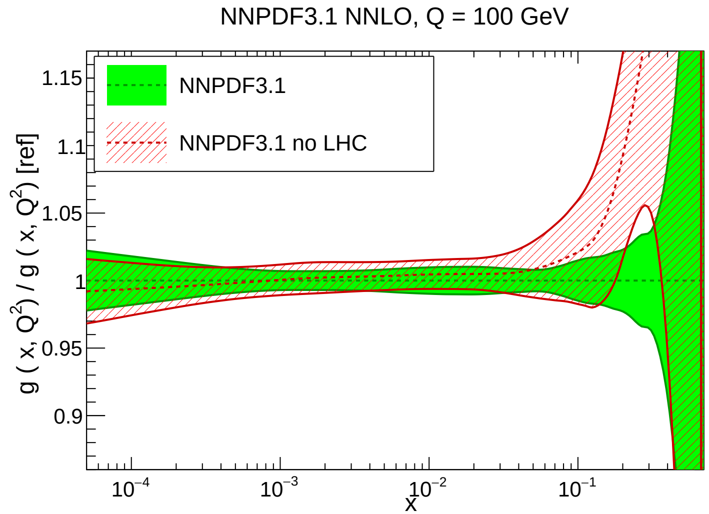
THE IMPACT OF LHC DATA

NEXT-GENERATION PDFS LARGELY DETERMINED BY LHC DATA: A FIRST!

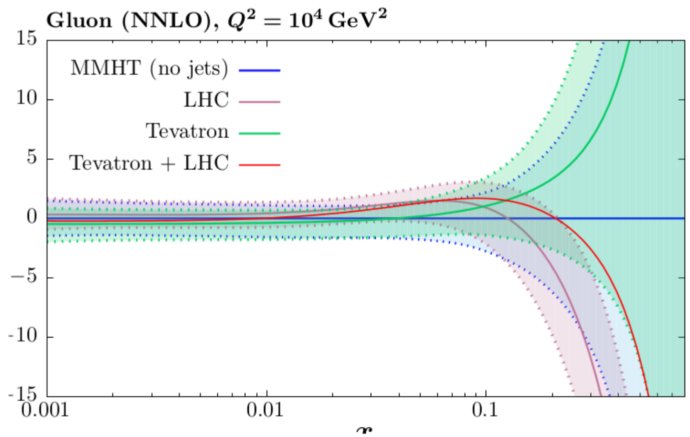
NNPDF3.1 up



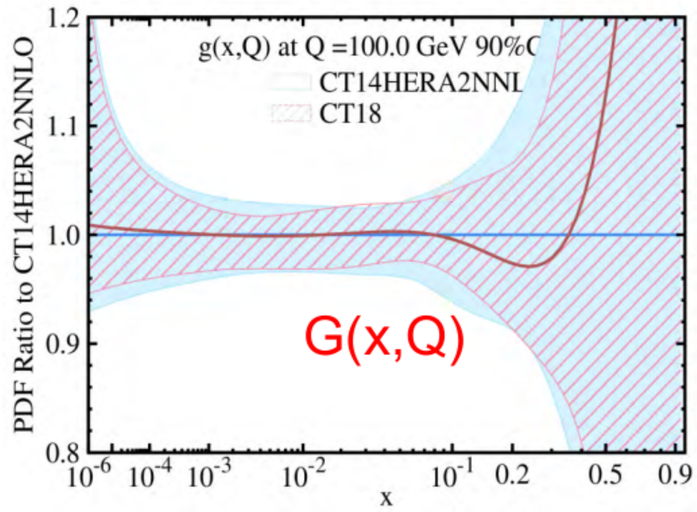
NNPDF3.1 glue



'MMHT' 19 glue (prelim., unpublished)



CT18 glue (preliminary, unpublished)



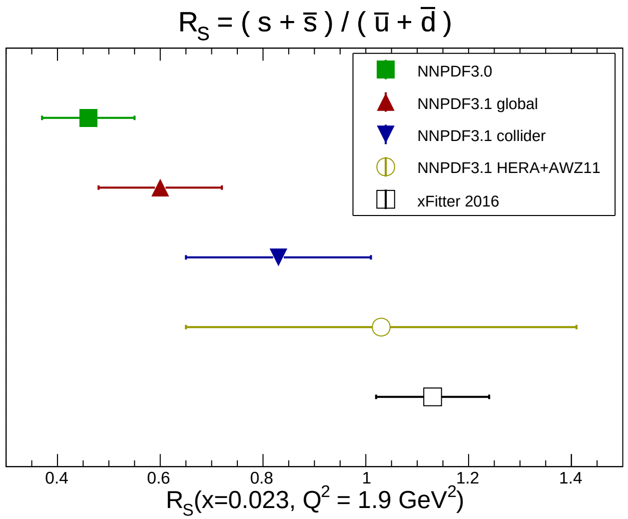
- SIGNIFICANT UNCERTAINTY REDUCTION
- MANY PDFS CHANGE BY MORE THAN ONE SIGMA
- BOTH FLAVOR SEPARATION & GLUON SIGNIFICANTLY AFFECTED

DATA VS. THEORY/METHODOLOGY
THE STRANGE PDF: DIS VS. W PRODUCTION

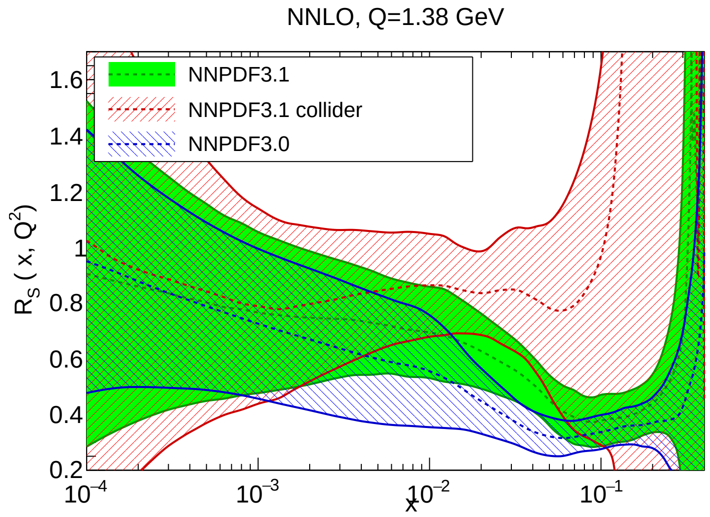
- STRANGE PDF CONTROLLED BY NEUTRINO DIS CHARM PRODUCTION + W PRODUCTION
- DIS DATA FAVOR "SUPPRESSED STRANGE" $\Rightarrow$ SMALL $R_s \equiv \frac{s+\bar{s}}{\bar{u}+\bar{d}}$
- ATLAS FAVORS ENHANCED STRANGENESS
- ATLAS IMPACT EXAGGERATED IN XFITTER ANALYSIS
- EVERYTHING CONSISTENT WITHIN UNCERTAINTIES IN GLOBAL FIT

THE STRANGENESS SUPPRESSION

XFITTER VS HERA+ATLAS VS. DIS ONLY VS ATLAS ONLY VS ALL



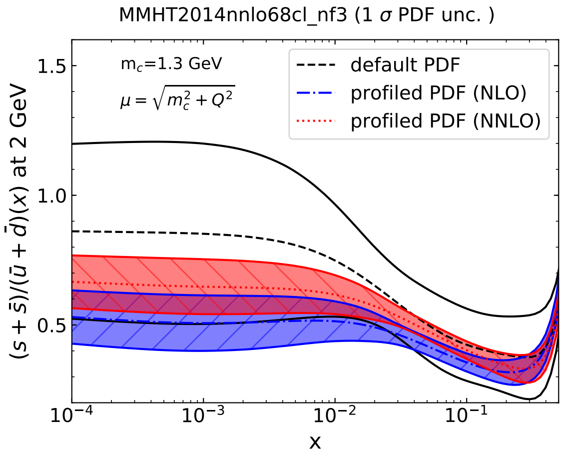
DIS ONLY VS ATLAS ONLY VS ALL



DATA VS. THEORY/METHODOLOGY
THE STRANGE PDF: DIS VS. W PRODUCTION

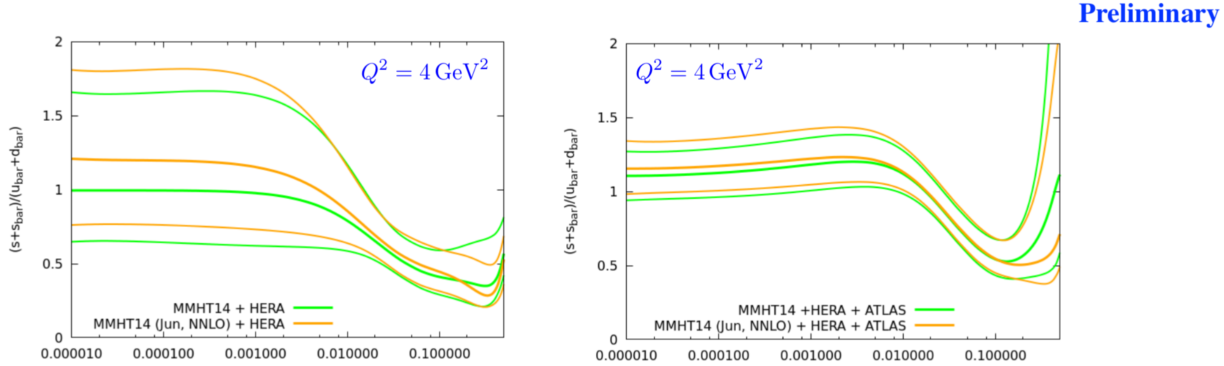
- MASSIVE CORRECTIONS TO CHARGED CURRENT DIS HITERTO INCLUDED TO NLO MASSLESS TO NNLO
- Gao, 2018 $\Rightarrow$ NNLO COMPUTED
- STRANGENESS ENHANCED BY NNLO CORRECTIONS

HERAPDF +NLO CC DIS VS NNLO CC DIS



(Gao, 2108)

MMHT WITH NLO VS NNLO CC DIS



(Harland-Lang, Thorne, prelim.)

LESSONS:

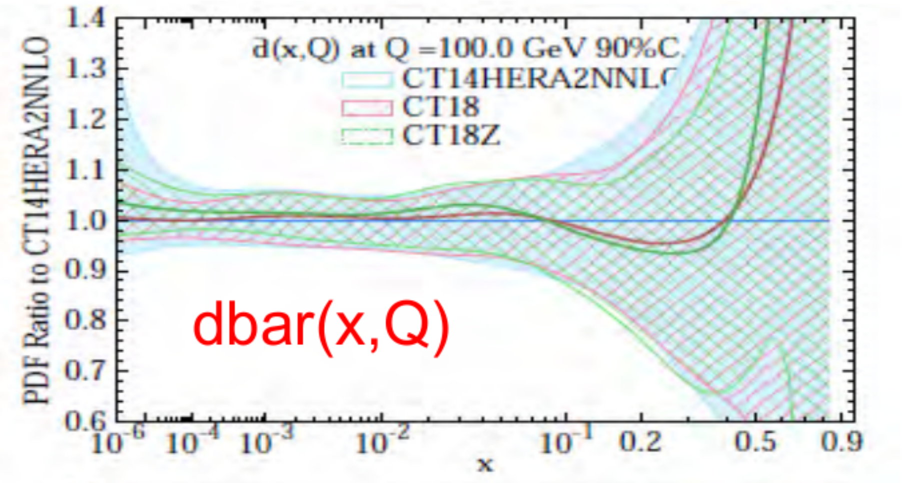
- BEWARE OF XFITTER HERA+X FITS
- IN A GLOBAL FIT DIFFERENT DATA ALWAYS PULL IN DIFFERENT DIRECTIONS!
- TENSIONS CAN BE RESOLVED BY BETTER THEORY

DATA VS. THEORY/METHODOLOGY
THE CHARM MASS AND TREATMENT
CT18 $\to$ CT18Z

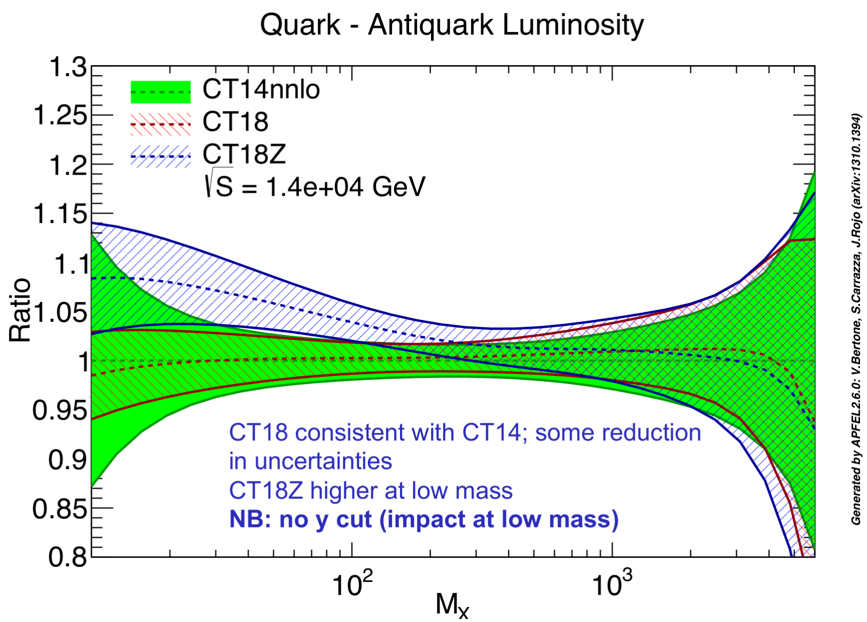
- ATLAS W AND Z 7TEV RAPIDITY INCLUDED
- CHARM MASS INCREASED
- x-DEPENDENT FACTORIZATION SCALE

CT18 VS. CT18Z (preliminary, unpublished)

DBAR PDF



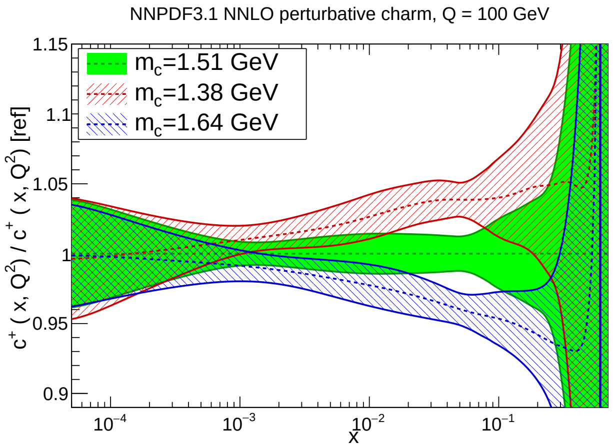
QQBAR LUMI



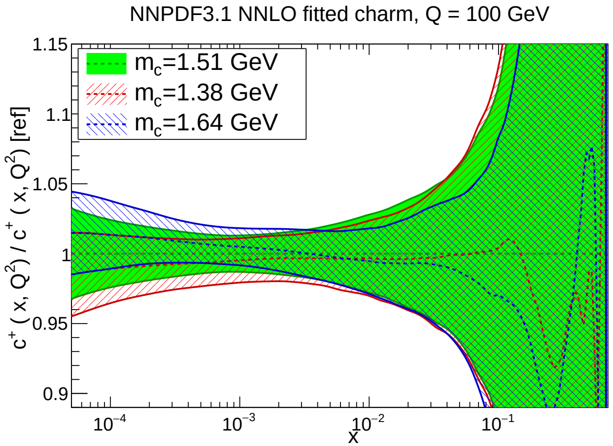
DATA VS. THEORY/METHODOLOGY
THE CHARM MASS AND TREATMENT
CHARM FROM DATA

- CHARM SHOULD NOT DEPEND STRONGLY ON CHARM MASS

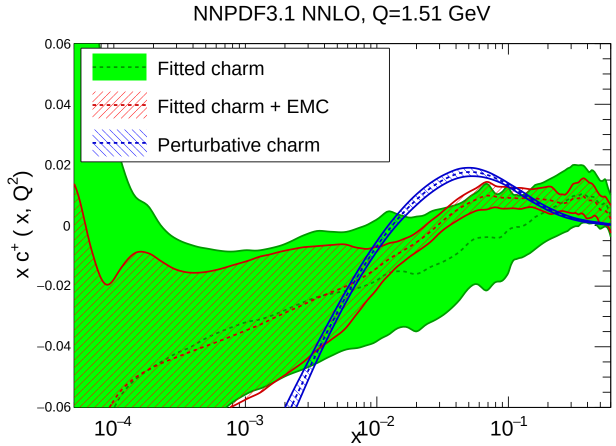
PERTURBATIVE CHARM VS m_c



FITTED CHARM VS m_c



FITTED VS. PERTURBATIVE CHARM



- ITS SHAPE SHOULD NOT BE DETERMINED BY FIRST-ORDER MATCHING (NO HIGHER NONTRIVIAL ORDERS KNOWN)
- MIGHT EVEN HAVE A NONPERTURBATIVE COMPONENT

FITTED VS. PERTURBATIVE:
SUPPRESSED AT MEDIUM-SMALL x,
ENHANCED AT VERY SMALL, VERY LARGE x

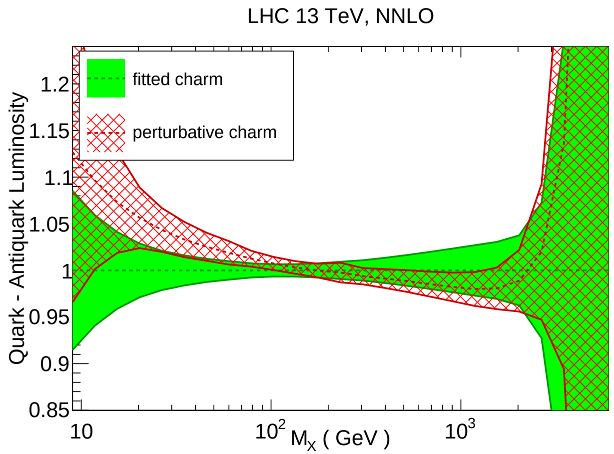
DATA VS. THEORY/METHODOLOGY

THE CHARM MASS AND TREATMENT

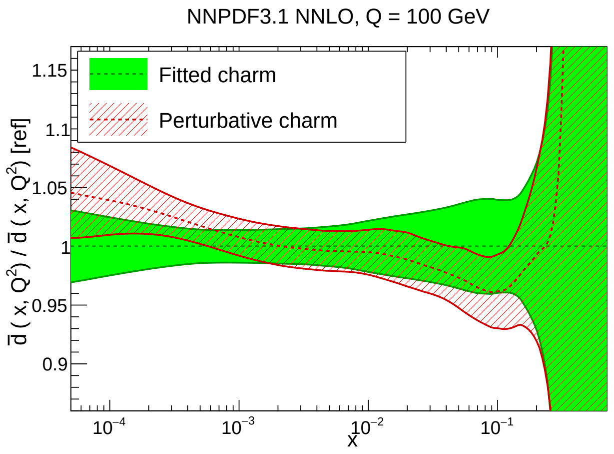
CHARM FROM DATA IMPACT ON LIGHT QUARK PDFS

FITTED VS. PERTURBATIVE CHARM

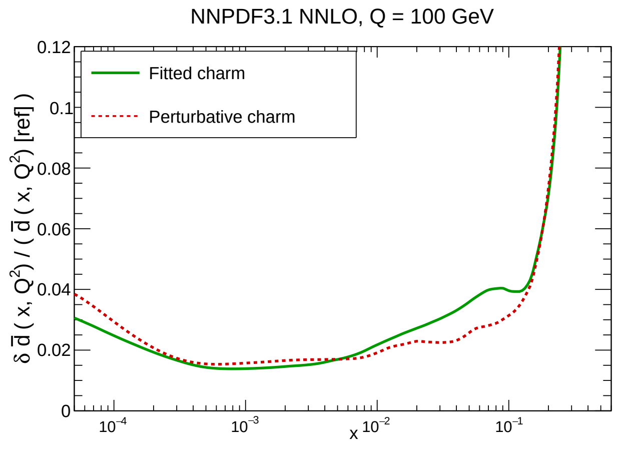
QQBAR LUMI



ANTIDOWN PDF



ANTIDOWN PDF UNCERTAINTY



- QUARK LUMI AFFECTED BECAUSE OF CHARM SUPPRESSION AT MEDIUM-x
- FLAVOR DECOMPOSITION ALTERED
- UNCERTAINTIES ON LIGHT QUARKS NOT SIGNIFICANTLY INCREASED
- AGREEMENT OF 13TeV W,Z PREDICTED CROSS-SECTIONS IMPROVES!

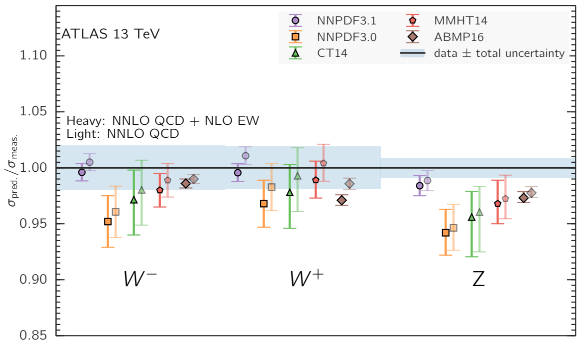
DATA VS. THEORY/METHODOLOGY

THE CHARM MASS AND TREATMENT

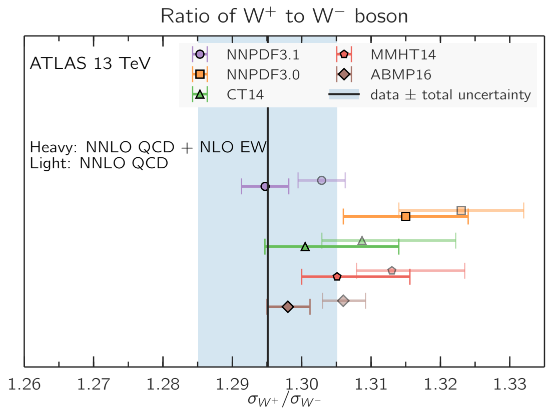
CHARM FROM DATA

IMPACT ON PHENOMENOLOGY

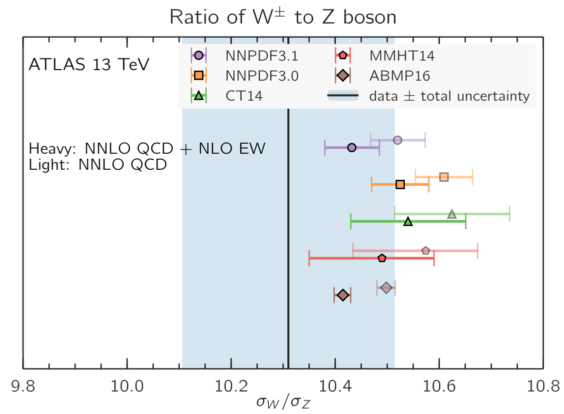
DRELL-YAN XSECTS



W^+/W^- XSECT RATIO



W/Z XSECT RATIO



- W, Z CROSS-SECTIONS AT 13 TEV IN PERFECT AGREEMENT WITH DATA THANKS TO FITTED CHARM!

LESSONS:

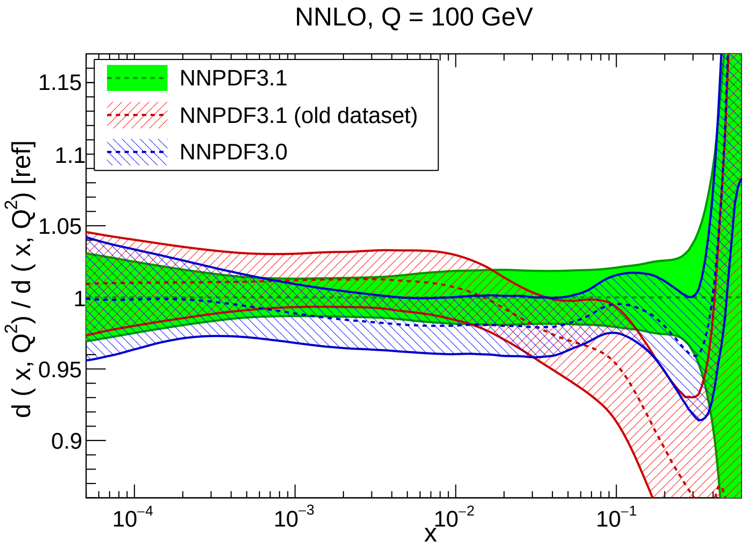
- TENSIONS CAN REVEAL METHODOLOGICAL ISSUES
- MORE LIKELY AS DATASET INCREASES, EXPERIMENTAL UNCERTAINTIES DECREASE
- RESOLVED BY MORE COMPLEX METHODOLOGY

DATA VS. METHODOLOGY

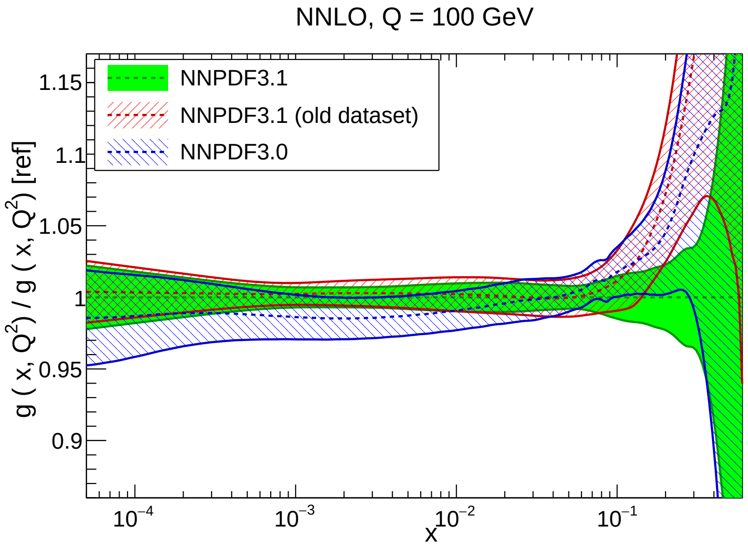
- NEW DATA ⇒ MAJOR METHODOLOGICAL CHOICES ⇒ SIGNIFICANT IMPACT
- NNPDF3.1 VS NNPDF3.0: DATA AND METHODOLOGY HAVE SIMILAR IMPACT

NNPDF3.0 VS. NNPDF3.1 VS. NNPDF3.1 W/ NNPDF3.0 DATASET

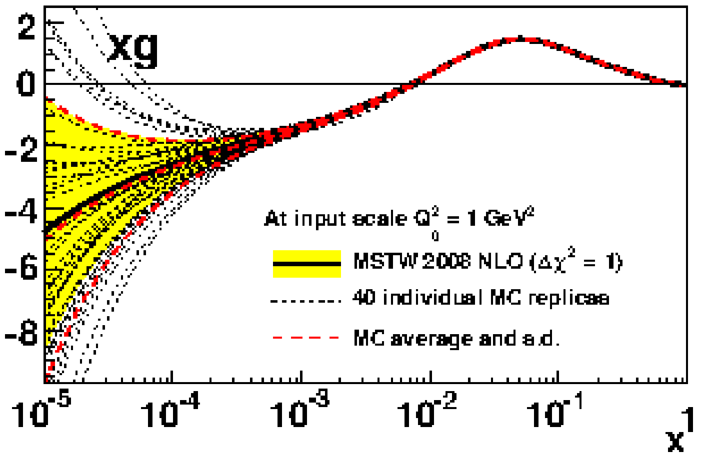
DOWN



GLUON

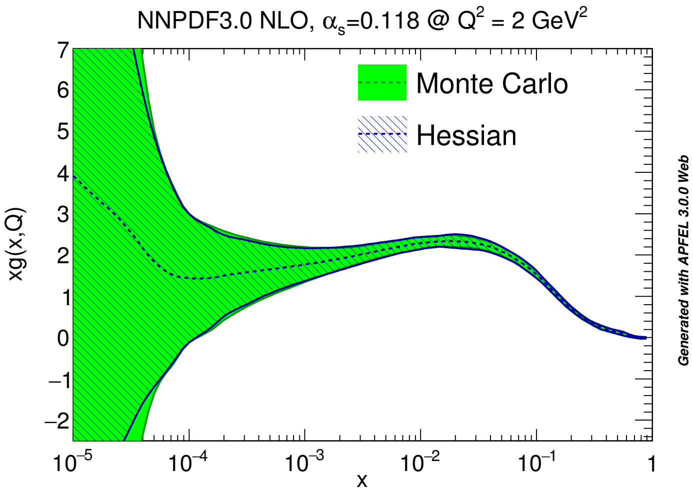


TOOLS I
MC $\Leftrightarrow$ HESSIAN



(Thorne, Watt, 2012)

- TO CONVERT HESSIAN INTO MONTECARLO GENERATE MULTIGAUSSIAN REPLICAS IN PARAMETER SPACE
- ACCURATE WHEN NUMBER OF REPLICAS SIMILAR TO THAT WHICH REPRODUCES DATA



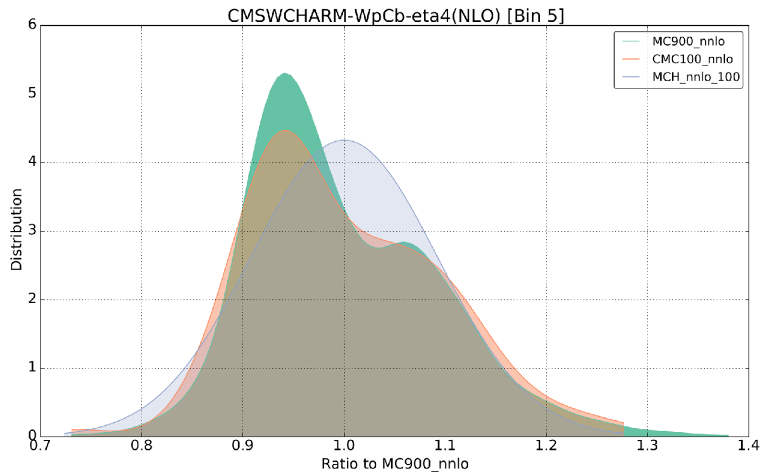
(Carrazza, SF, Kassabov, Rojo, 2015)

- TO CONVERT MONTE CARLO INTO HESSIAN, SAMPLE THE REPLICAS $f_i(x)$ AT A DISCRETE SET OF POINTS & CONSTRUCT THE ENSUING COVARIANCE MATRIX
- EIGENVECTORS OF THE COVARIANCE MATRIX AS A BASIS IN THE VECTOR SPACE SPANNED BY THE REPLICAS BY SINGULAR-VALUE DECOMPOSITION
- NUMBER OF DOMINANT EIGENVECTORS SIMILAR TO NUMBER OF REPLICAS $\Rightarrow$ ACCURATE REPRESENTATION

TOOLS II
NONGAUSSIAN BEHAVIOUR

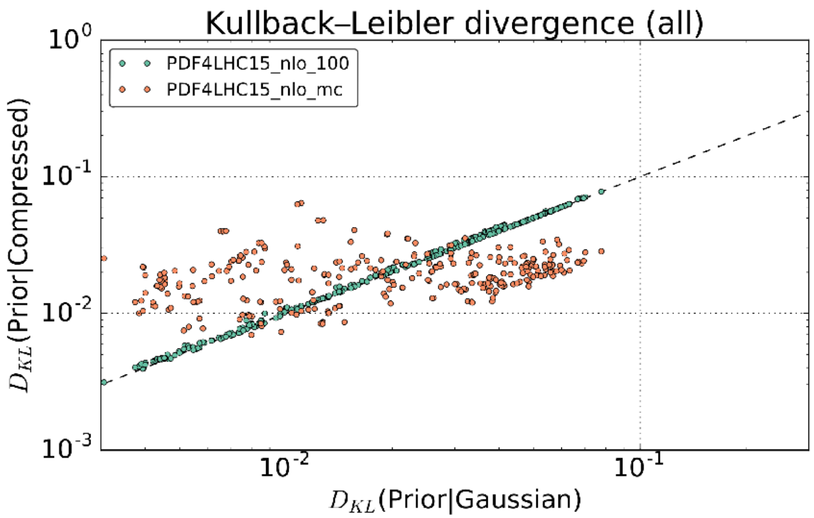
MONTE CARLO COMPARED TO HESSIAN

CMS $W + c$ production



- DEVIATION FROM GAUSSIANITY E.G. AT LARGE x DUE TO LARGE UNCERTAINTY + POSITIVITY BOUNDS
$\Rightarrow$ RELEVANT FOR SEARCHES
- CANNOT BE REPRODUCED IN HESSIAN FRAMEWORK
- WELL REPRODUCED BY COMPRESSED MC

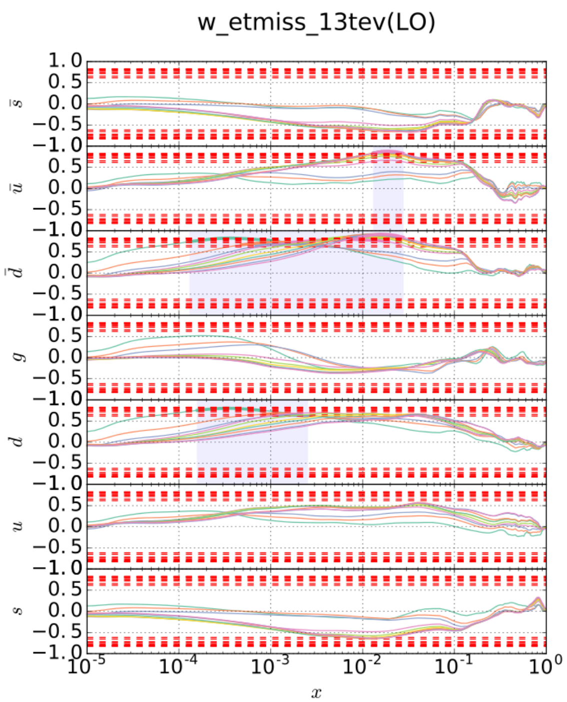
- DEFINE KULLBACK-LEIBLER DIVERGENCE $D_{\mathrm{KL}} = \int_{-\infty}^{\infty} P(x) \frac{\ln P(x)}{\ln Q(x)}\, dx$ BETWEEN A PRIOR P AND ITS REPRESENTATION Q
- D_{KL} BETWEEN PRIOR AND HESSIAN DEPENDS ON DEGREE OF GAUSSIANITY
- D_{KL} BETWEEN PRIOR AND COMPRESSED MC DOES NOT

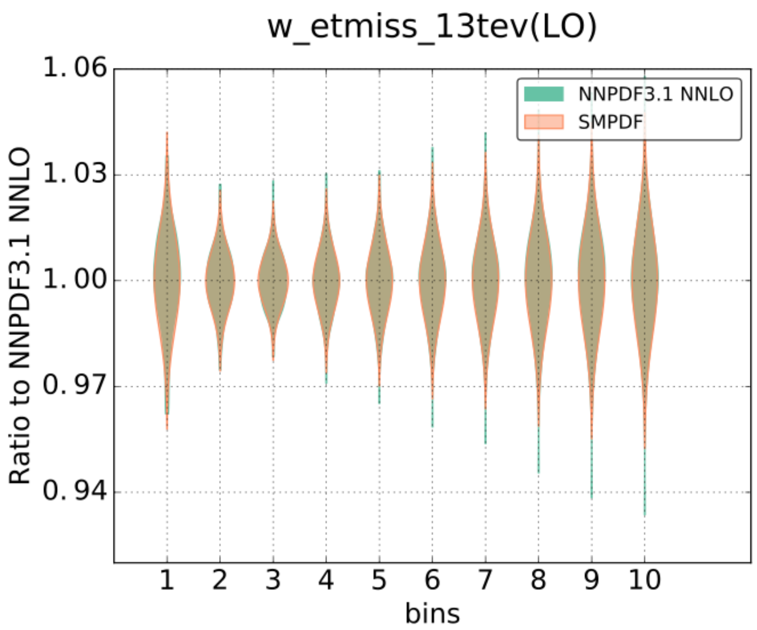


CAN (A) GAUGE WHEN MC IS MORE ADVANTAGEOUS THAN HESSIAN;
(B) ASSESS THE ACCURACY OF COMPRESSION

TOOLS III
OPTIMIZED PDFS: SMPDF

- OLD ASPIRATION: PDFS OPTIMIZED TO PROCESSES (Pumplin 2009)
- SELECT SUBSET OF THE COVARIANCE MATRIX CORRELATED TO A GIVEN SET OF PROCESSES
- PERFORM SVD ON THE REDUCED COVARIANCE MATRIX, SELECT DOMINANT EIGENVECTOR, PROJECT OUT ORTHOGONAL SUBSPACE
- ITERATE UNTIL DESIRED ACCURACY REACHED
- CAN ADD PROCESSES TO GIVEN SET; CAN COMBINE DIFFERENT OPTIMIZED SETS
- WEB INTERFACE AVAILABLE





(Carrazza, SF, Kassabov, Rojo, 2016)

- EG ggH, $Hb\bar{b}$, $W\ E_T^{\rm miss}$ $\Rightarrow$ 11 EIGENVECTORS
- STUDY CORRELATIONS OF PDFS TO DATA AND AMONG THEMSELVES!