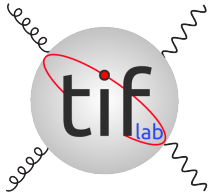
MACHINE LEARNING

THE UNKNOWN

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA



CRACOW SCHOOL OF TH. PHYSICS

NOVEMBER 20, 2020

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

PRECISION PHYSICS AND UNCERTAINTIES

AN EXAMPLE: HIGGS IN GLUON FUSION

~~2015~~ VS. NOW

~~(APPR)~~ N^{3}LO+N^{3}LL QCD (EFT); NLO ~~PURE~~ EW; NLO EXACT HQ; NNLO APPROX TOP; NNLO PDFS

$$\sigma(\text{LHC13},\ m_H = 125\ \text{GeV}) = 48.58\ \text{pb} \pm \sout{2.2^{\text{TH}}\ (4.5\%) \pm 1.6^{\text{PDF}+\alpha_s}\ (3.2\%)}$$
$$\sigma(\text{LHC13},\ m_H = 125\ \text{GeV}) = \quad \text{pb} \pm 1.6^{\text{TH}}\ (3.3\%) \pm 1.4^{\text{PDF}+\alpha_s}\ (2.8\%)$$

PDF+α_s UNCERTAINTY

PDF: ~~± 0.9 pb (1.9%)~~ ± 0.5 pb (1%)

α_s: ± 1.3 pb (2.6%)

- UNCERTAINTY RAPIDLY DECREASING
- TOWARD 1% UNCERTAINTIES!

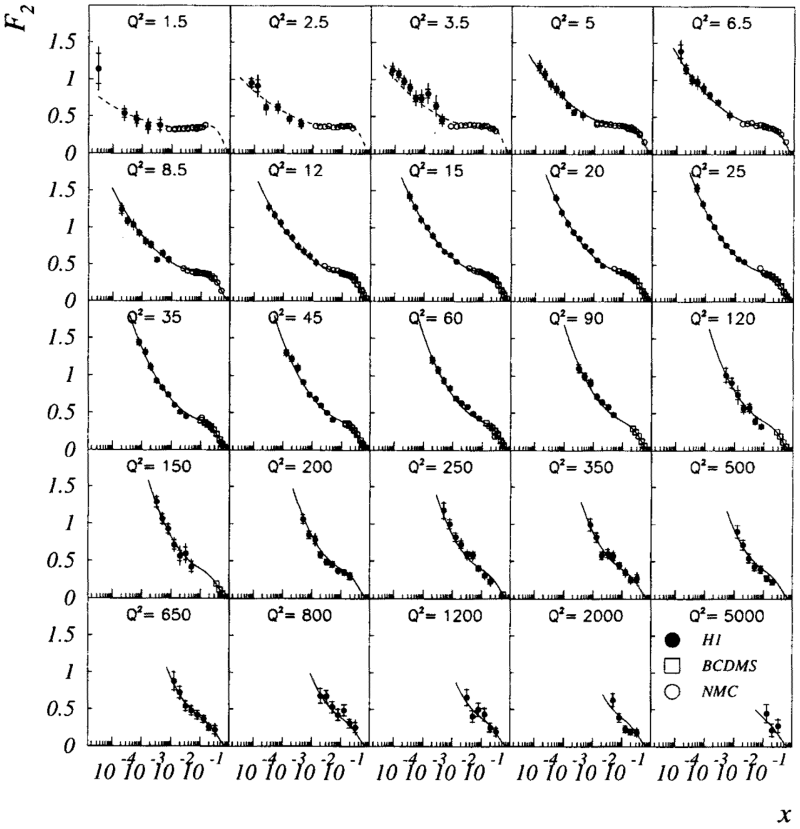
UNCERTAINTY ESTIMATION:

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

DATA OUTSIDE THE DATA REGION

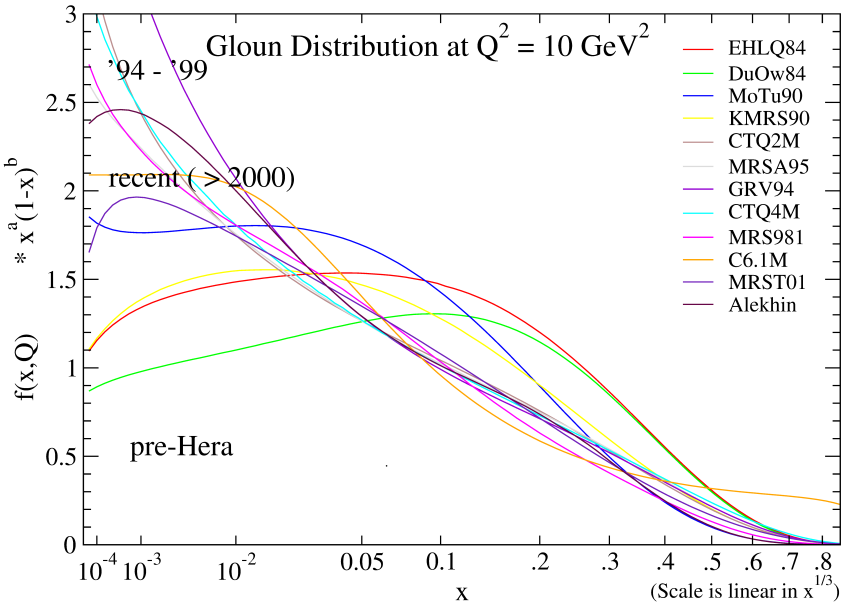
1995: THE RISE OF STRUCTURE FUNCTIONS AT HERA

FIRST HERA DATA VS OLDER DATA



A. de Roeck, Cracow epiphany conf. 1996

HISTORICAL COMPILATION OF GLUON PDFS

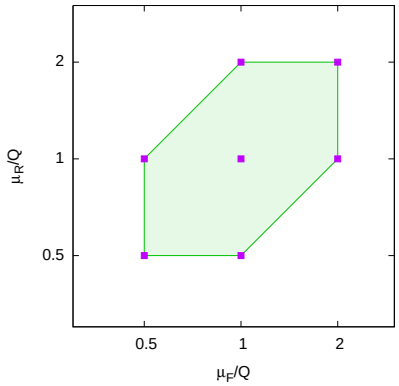


W.K.Tung, DIS 2004

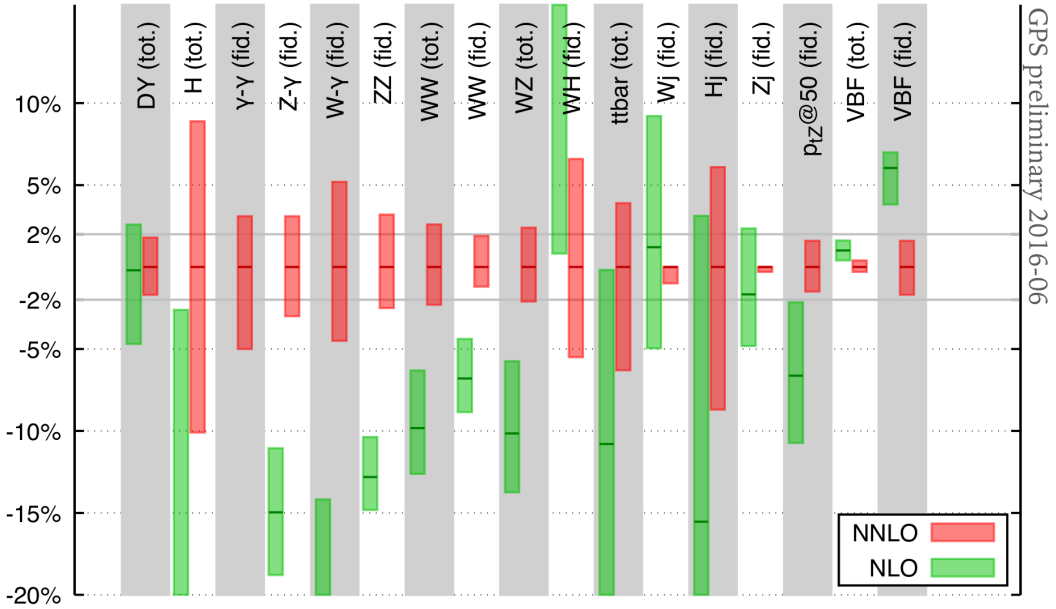
- RISE OF F_2 AT HERA CAME $\Rightarrow$ SURPRIZE
- UNCERTAINTY $\Leftrightarrow$ BIAS

"THEORY UNCERTAINTIES"
MISSING HIGHER ORDER CORRECTIONS

- TRADITIONALLY ESTIMATED BY "SEVEN POINT" SCALE VARIATION
- TRADITIONALLY, VARIATION BY FACTOR 2



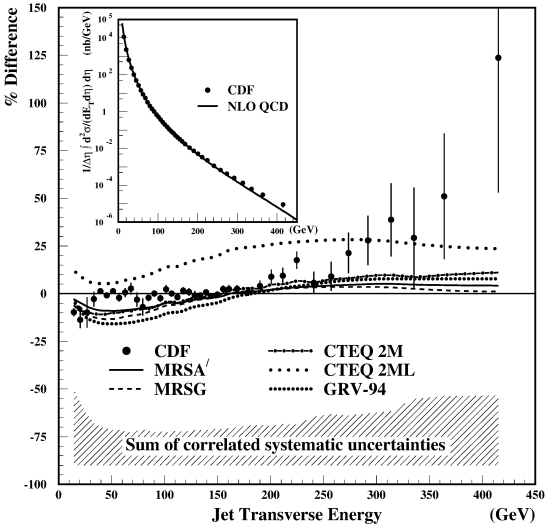
HOW WELL DOES IT WORK?



- NNLO WITHIN 7-POINT NLO BAND IN 3/17 CASES
- KNOWN ISSUES: SCALE VARIATION DOES NOT ACCOUNT FOR NEW CHANNELS, STRUCTURES

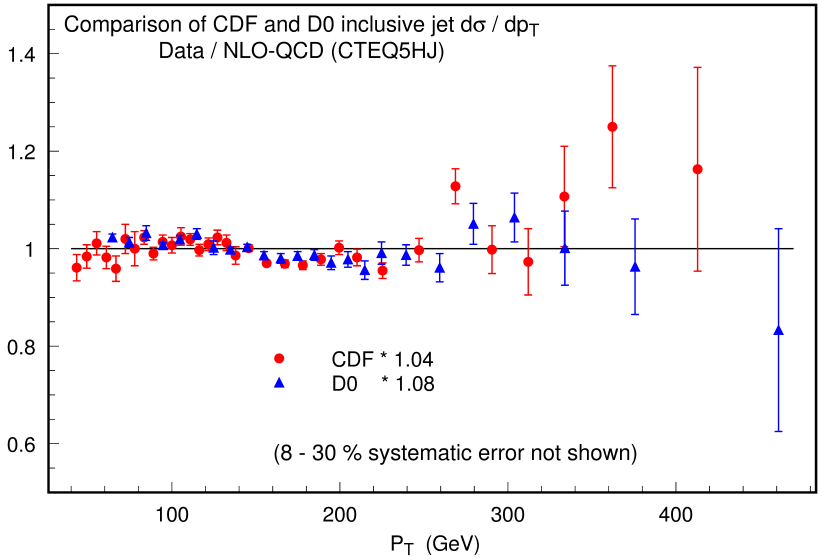
DOES IT MATTER?
THE "DISCOVERY" OF QUARK COMPOSITENESS

- 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 $x \gtrsim 0.05$!

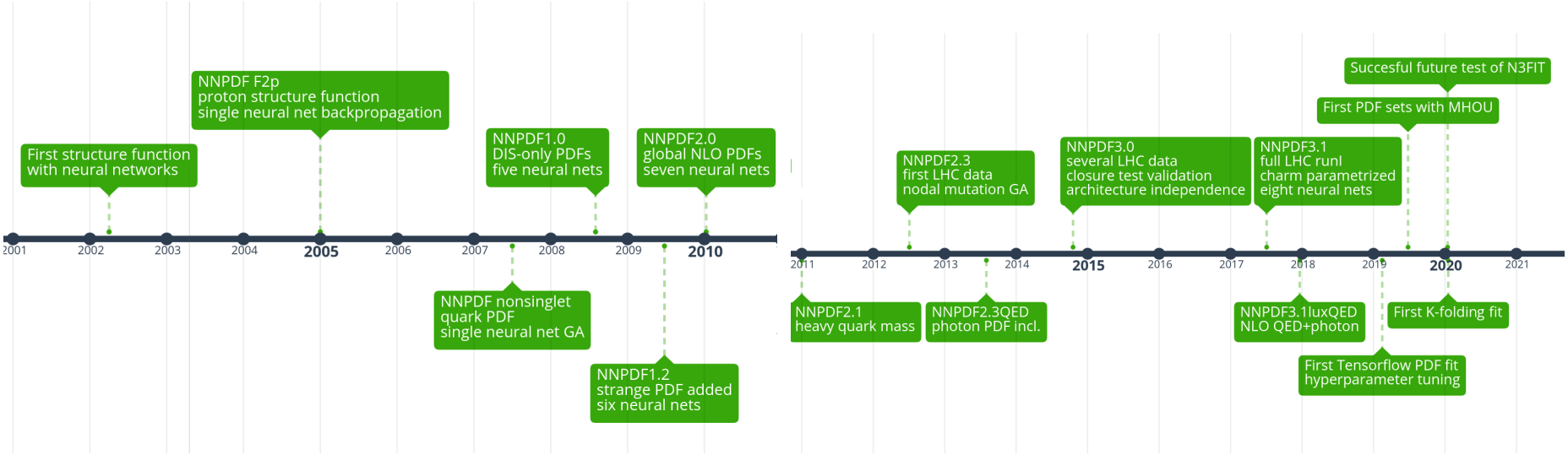


DISCREPANCY REMOVED IF JET DATA USED FOR GLUON DETERMINATION

NEW CTEQ GLUON (1998)



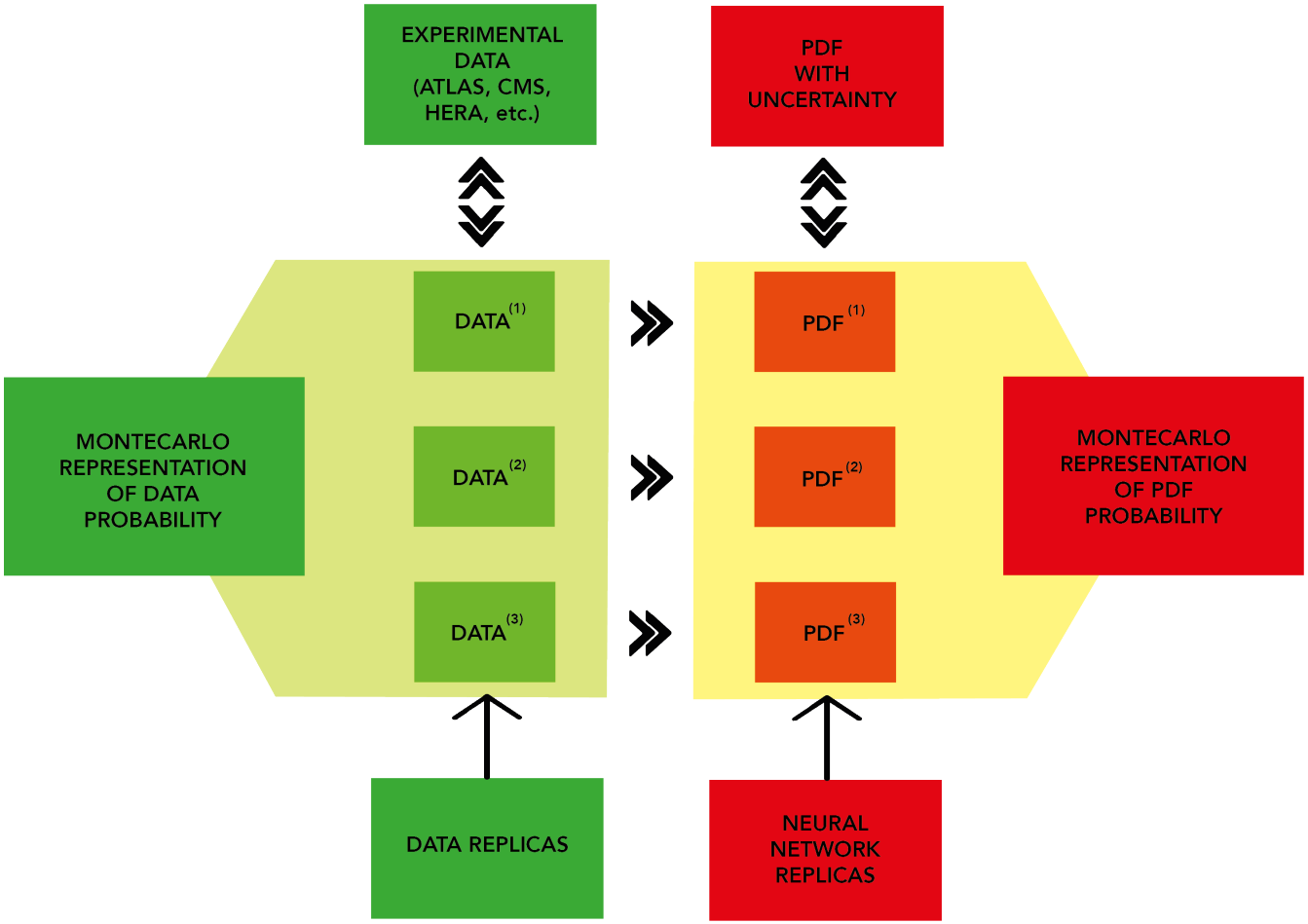
UNCERTAINTIES AS AN AI PROBLEM:
NNPDF



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



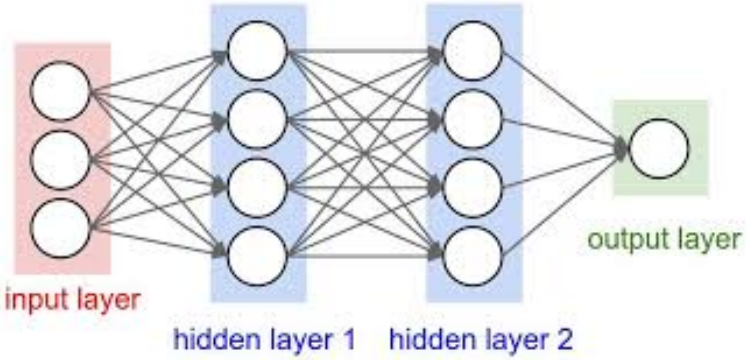
FINAL PDF SET: $f_i^{(a)}(x,\mu)$;

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

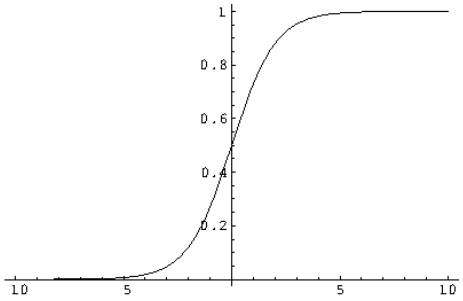
ARTIFICIAL INTELLIGENCE

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

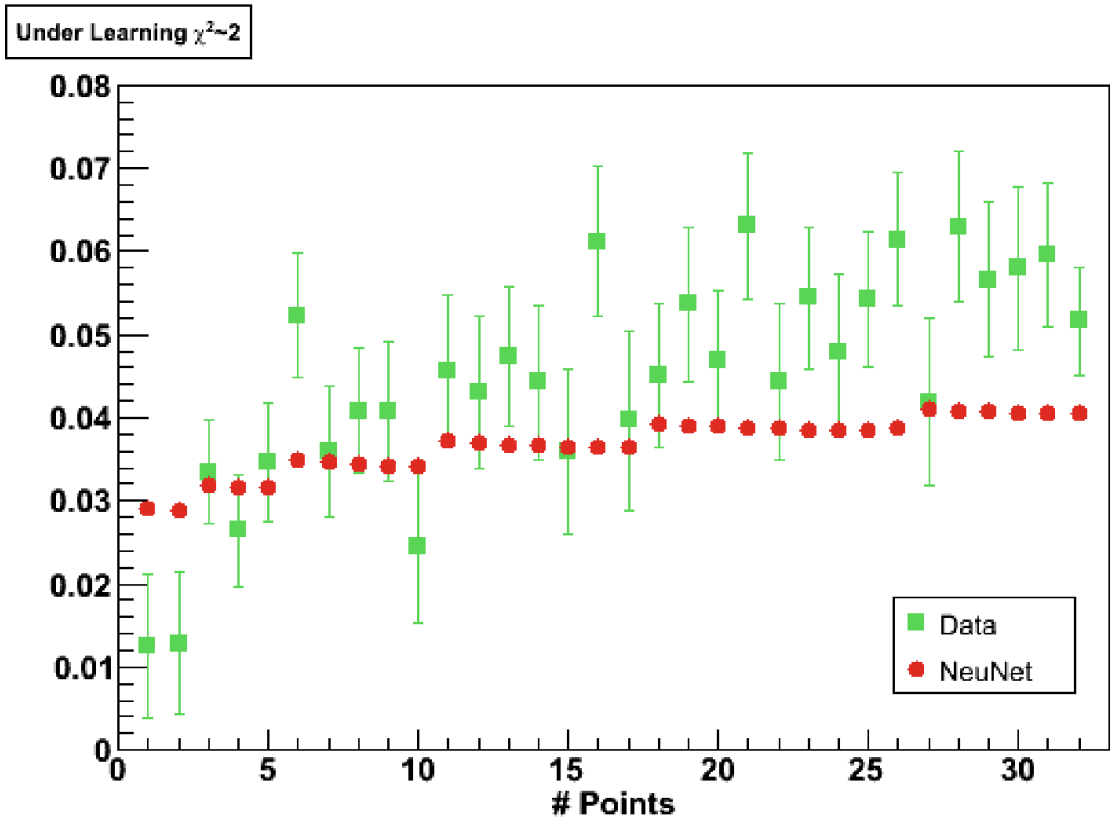
$$f(x) = \frac{1}{1+e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - x\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - x\omega_{21}^{(1)}}}}}$$

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

NEURAL LEARNING

- ✔ FEATURES LEARNT GRADUALLY
- ✗ UNTIL LEARNING NOISE

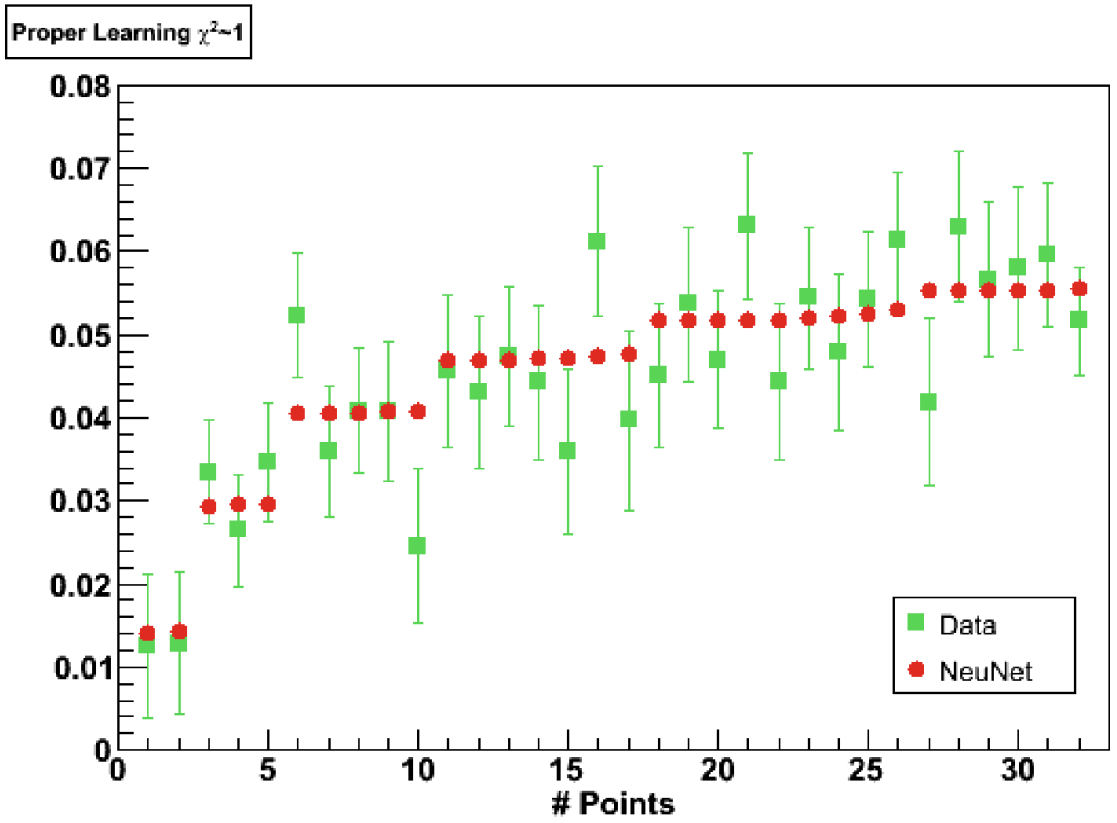
UNDERLEARNING



NEURAL LEARNING

- ✔ FEATURES LEARNT GRADUALLY
- ✗ UNTIL LEARNING NOISE

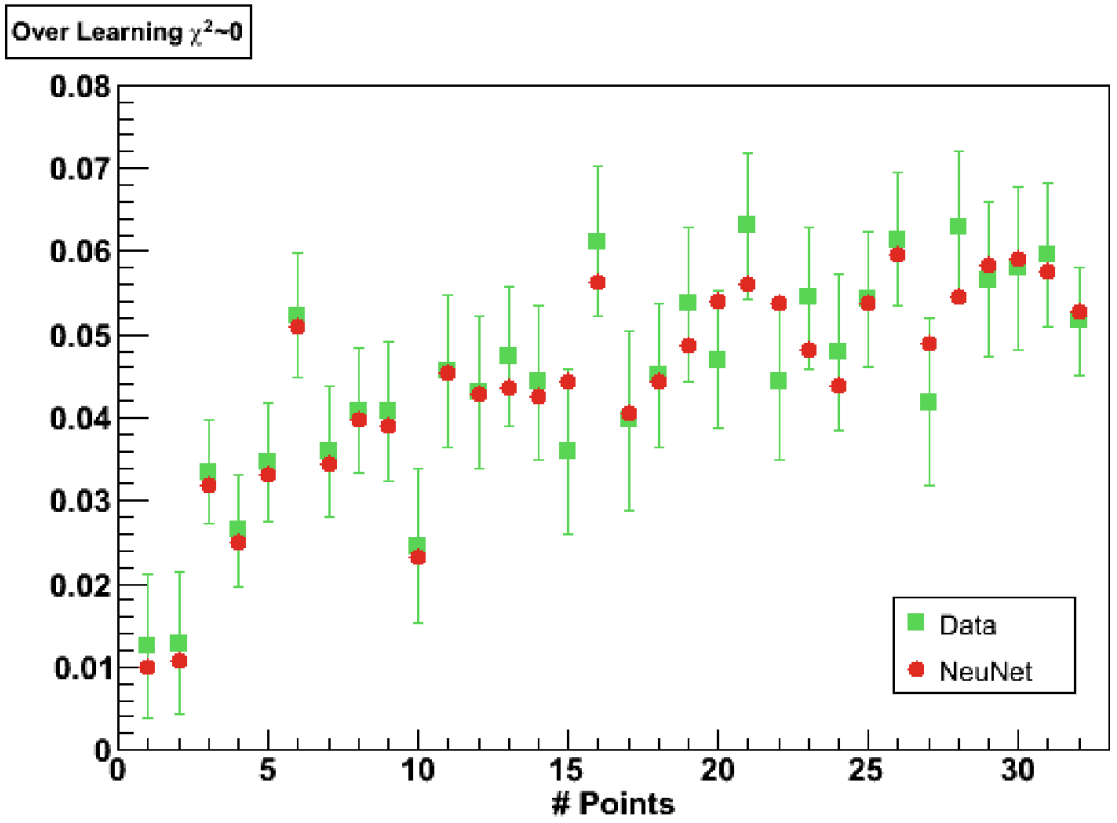
PROPER LEARNING



NEURAL LEARNING

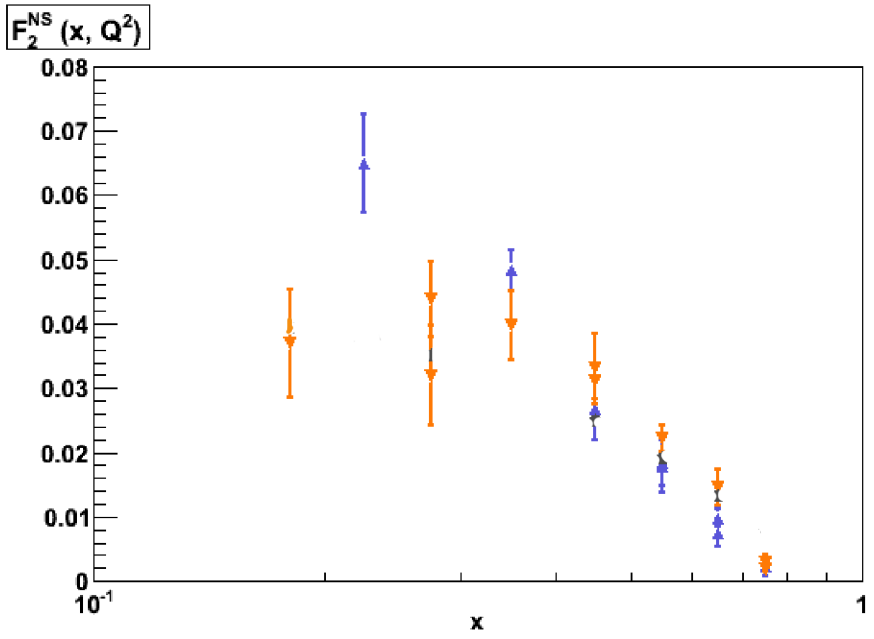
- ✔ FEATURES LEARNT GRADUALLY
- ✗ UNTIL LEARNING NOISE

OVERLEARNING



QUALITY CONTROL: CROSS-VALIDATION

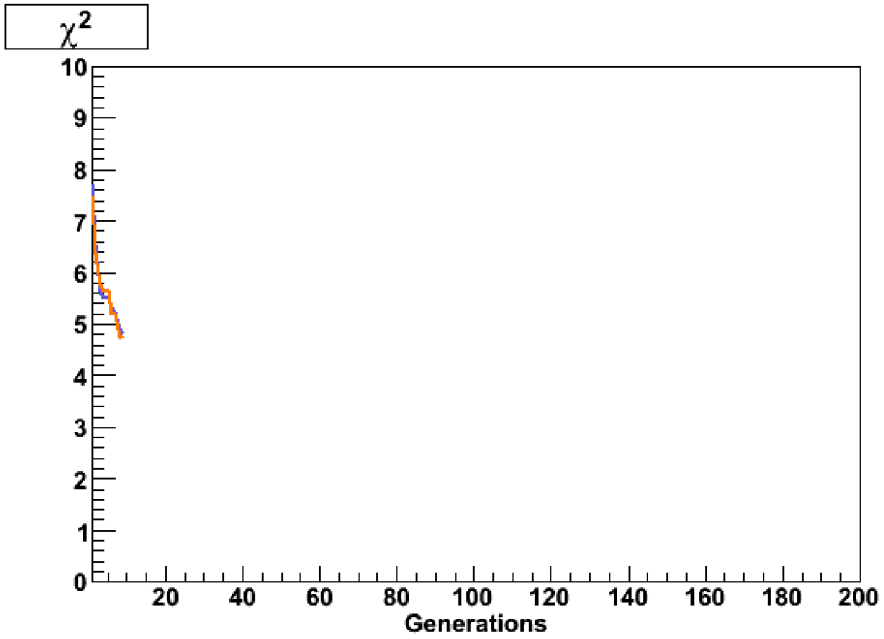
- DIVIDE THE DATA IN TWO SETS: TRAINING AND VALIDATION
- MINIMIZE THE χ^2 OF THE DATA IN THE TRAINING SET
- AT EACH ITERATION, COMPUTE THE χ^2 FOR THE DATA IN THE VALIDATION SET (NOT USED FOR FITTING)
- WHEN THE VALIDATION χ^2 STOPS DECREASING, STOP THE FIT

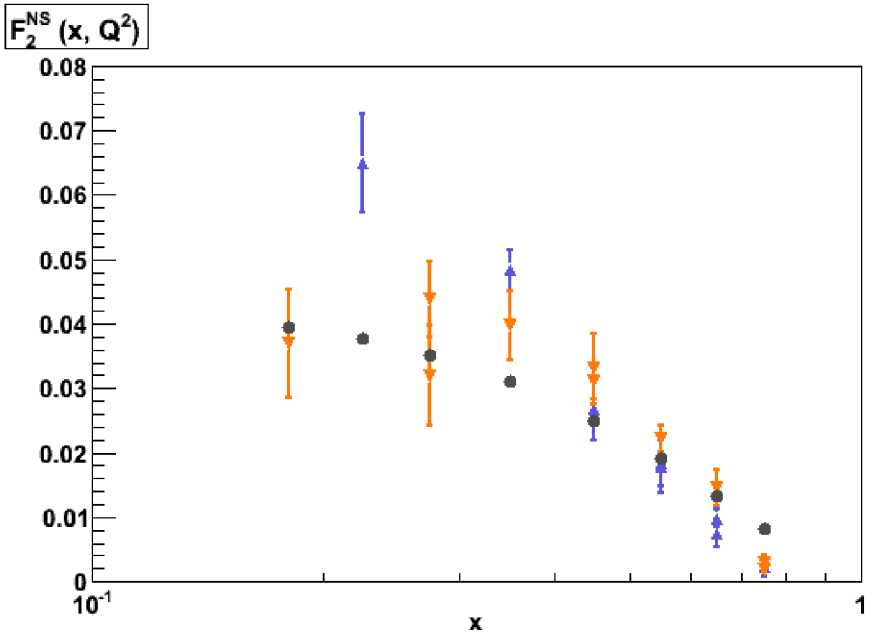


QUALITY CONTROL: CROSS-VALIDATION

- 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!

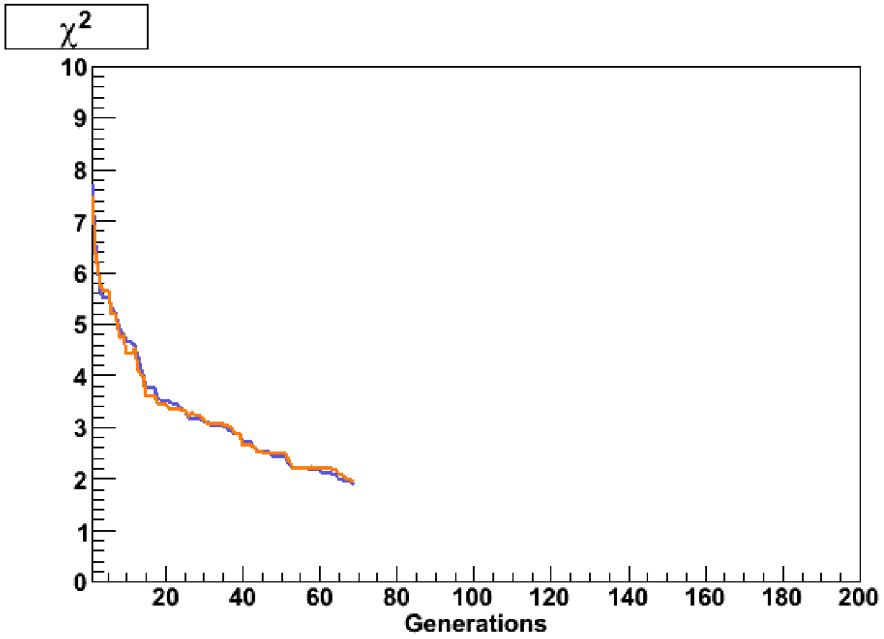


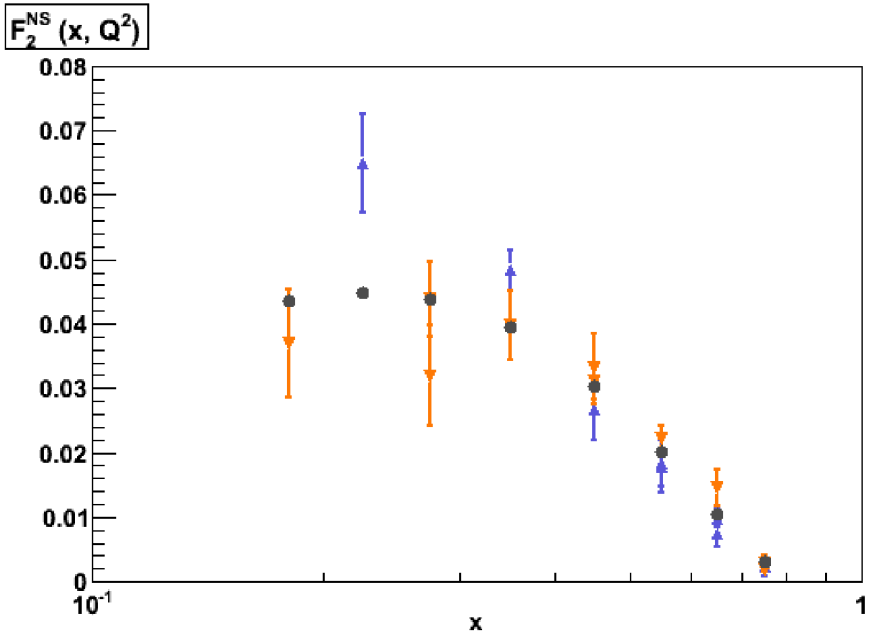


QUALITY CONTROL: CROSS-VALIDATION

- 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!

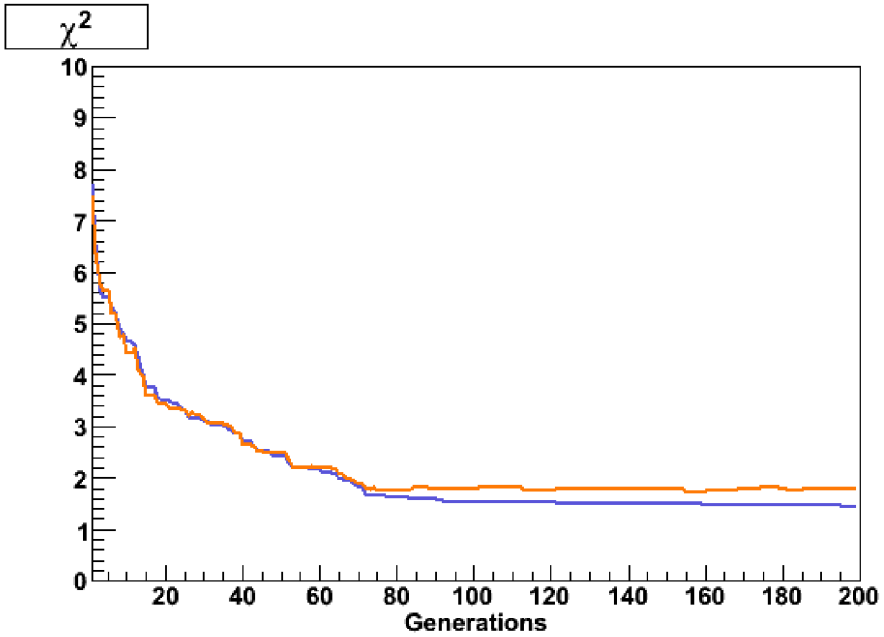


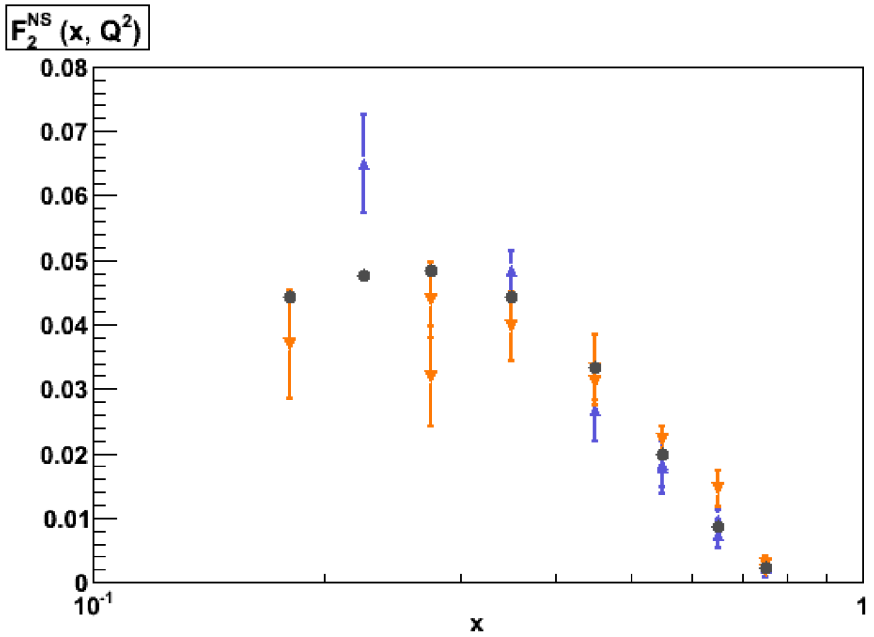


QUALITY CONTROL: CROSS-VALIDATION

- 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!

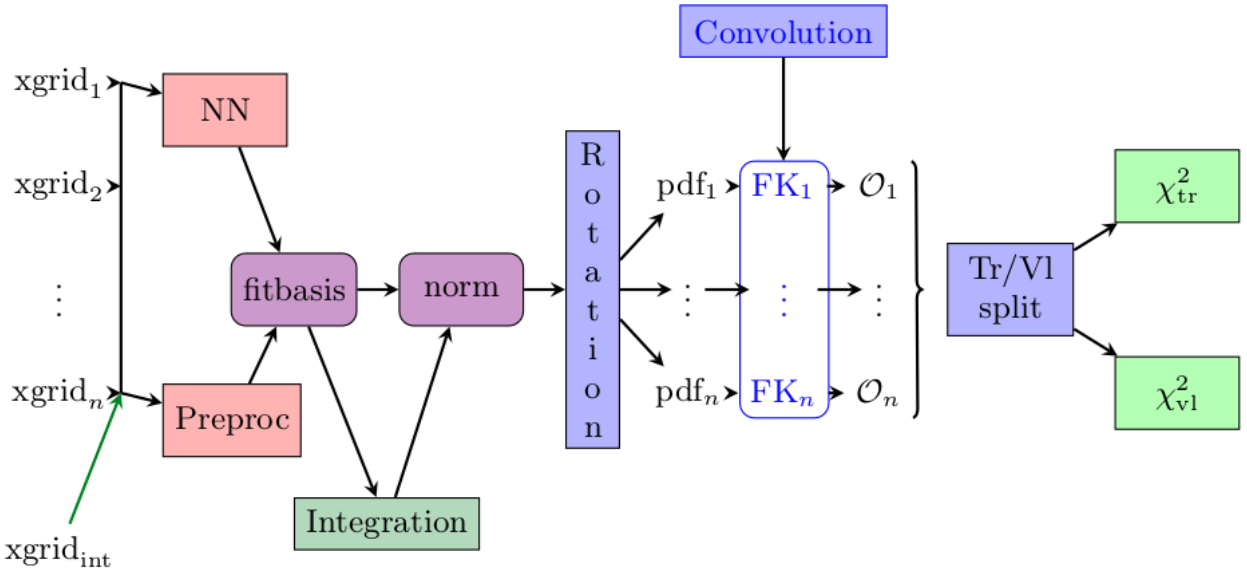




LEARNING THE METHODOLOGY

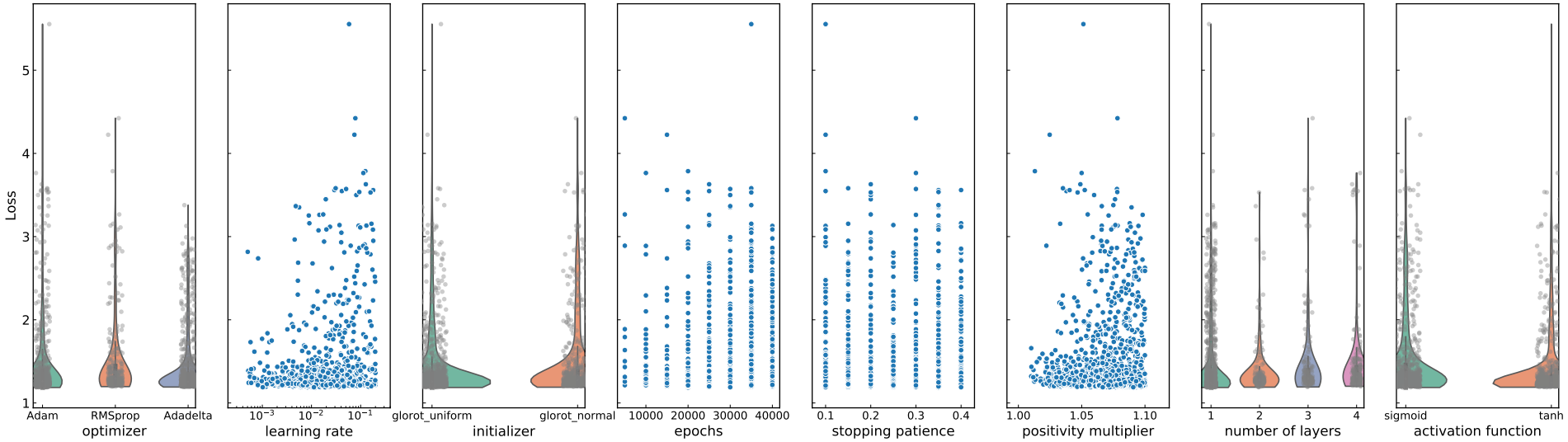
HOW DO WE KNOW THAT THE METHODOLOGY IS THE BEST?
"ACCUMULATED WISDOM" INEFFICIENT AND SLOW

HYPEROPTIMIZATION
N3FIT



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

FITTING THE METHODOLOGY
HYPEROPTIMIZATION SCANS



HYPEROPT PARAMETERS

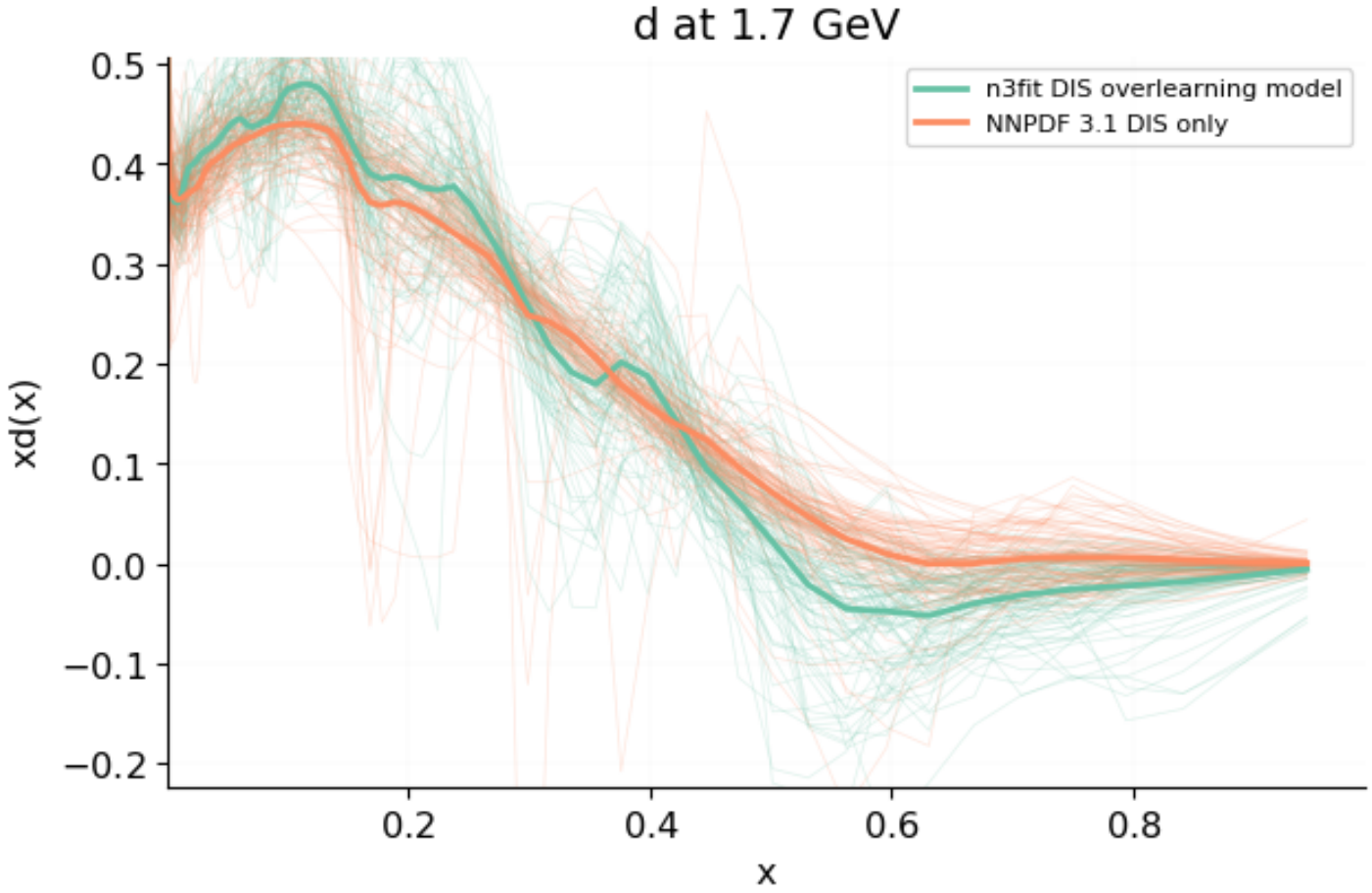
Neural Network	Fit options
Number of layers (*)	Optimizer (*)
Size of each layer	Initial learning rate (*)
Dropout	Maximum number of epochs (*)
Activation functions (*)	Stopping Patience (*)
Initialization functions (*)	Positivity multiplier (*)

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

FITTING THE METHODOLOGY

THE OVERFITTING PROBLEM

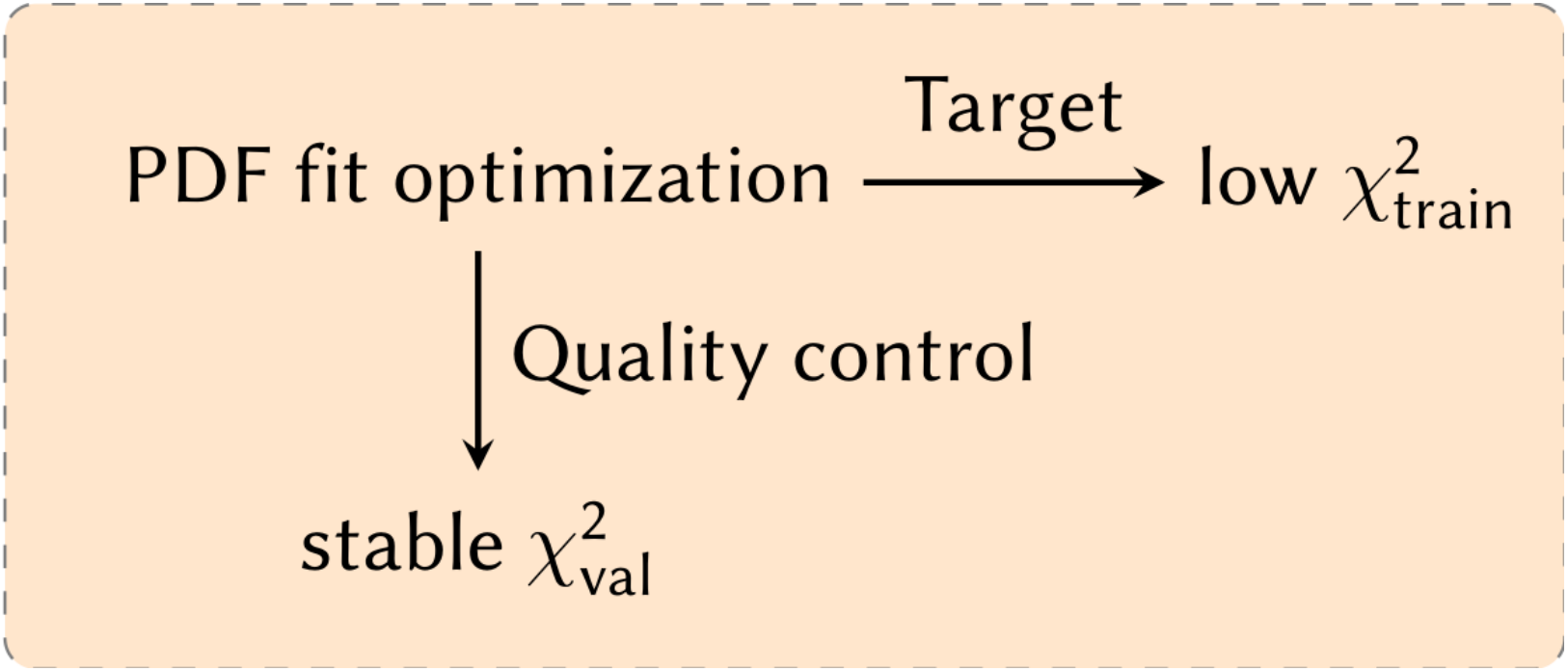
DOWN QUARK: HYPEROPTIMIZED VS. STANDARD



- 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?

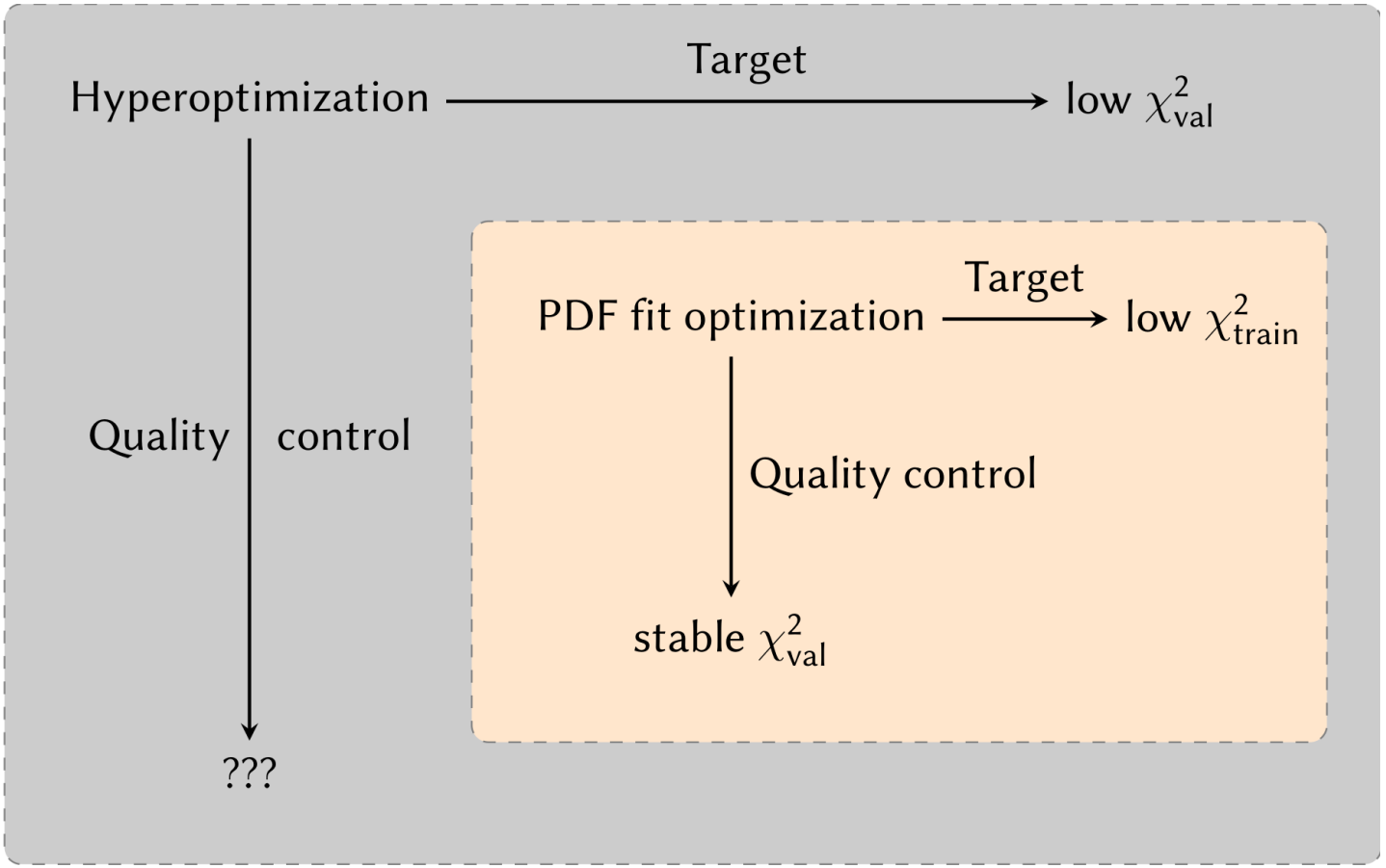
OPTIMIZATION



CROSS-VALIDATION SELECTS THE OPTIMAL MINIMUM

WHAT HAPPENED?

HYPEROPTIMIZATION

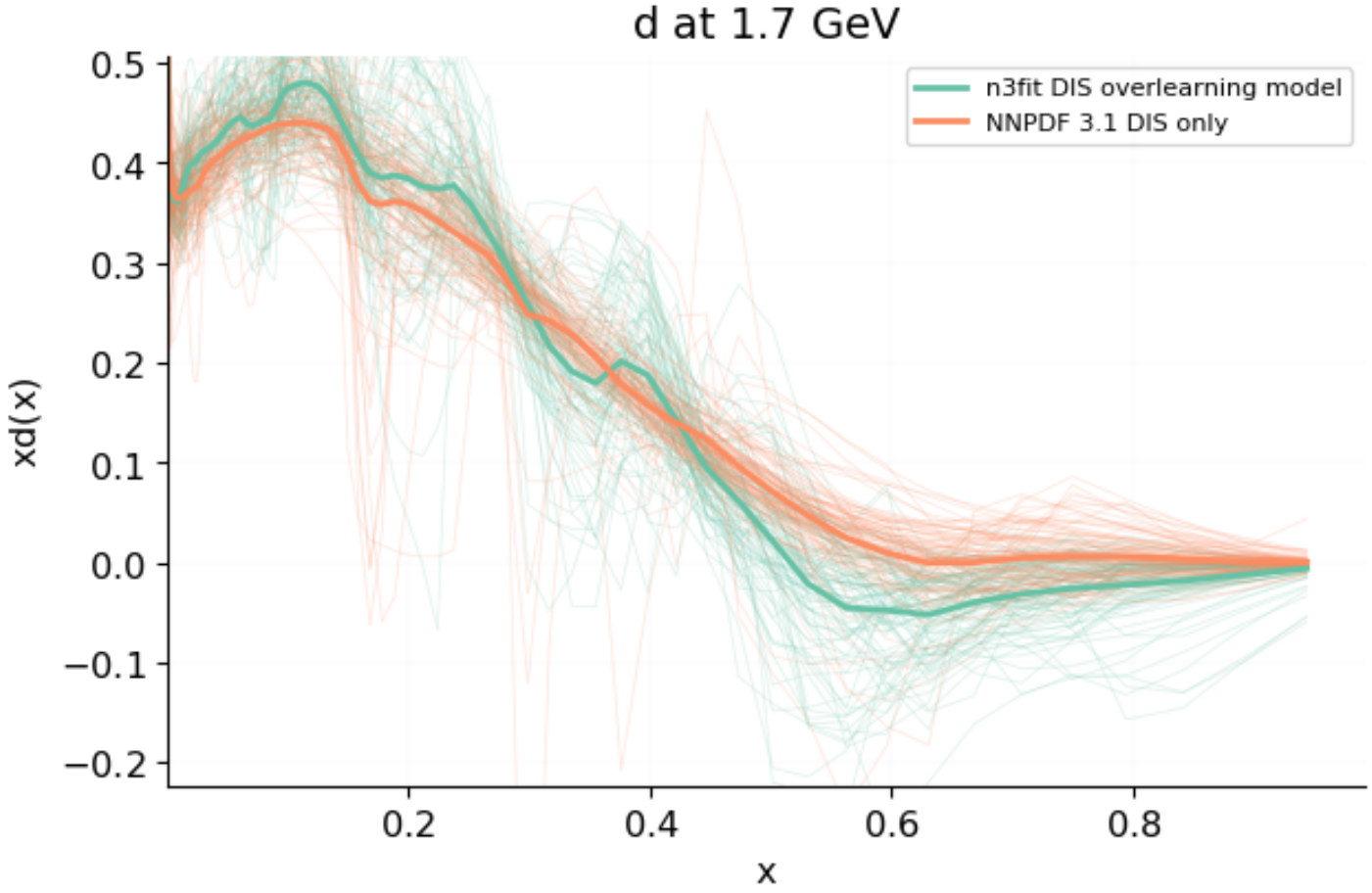


WE ARE MISSING A SELECTION CRITERION

FITTING THE METHODOLOGY

THE OVERFITTING PROBLEM

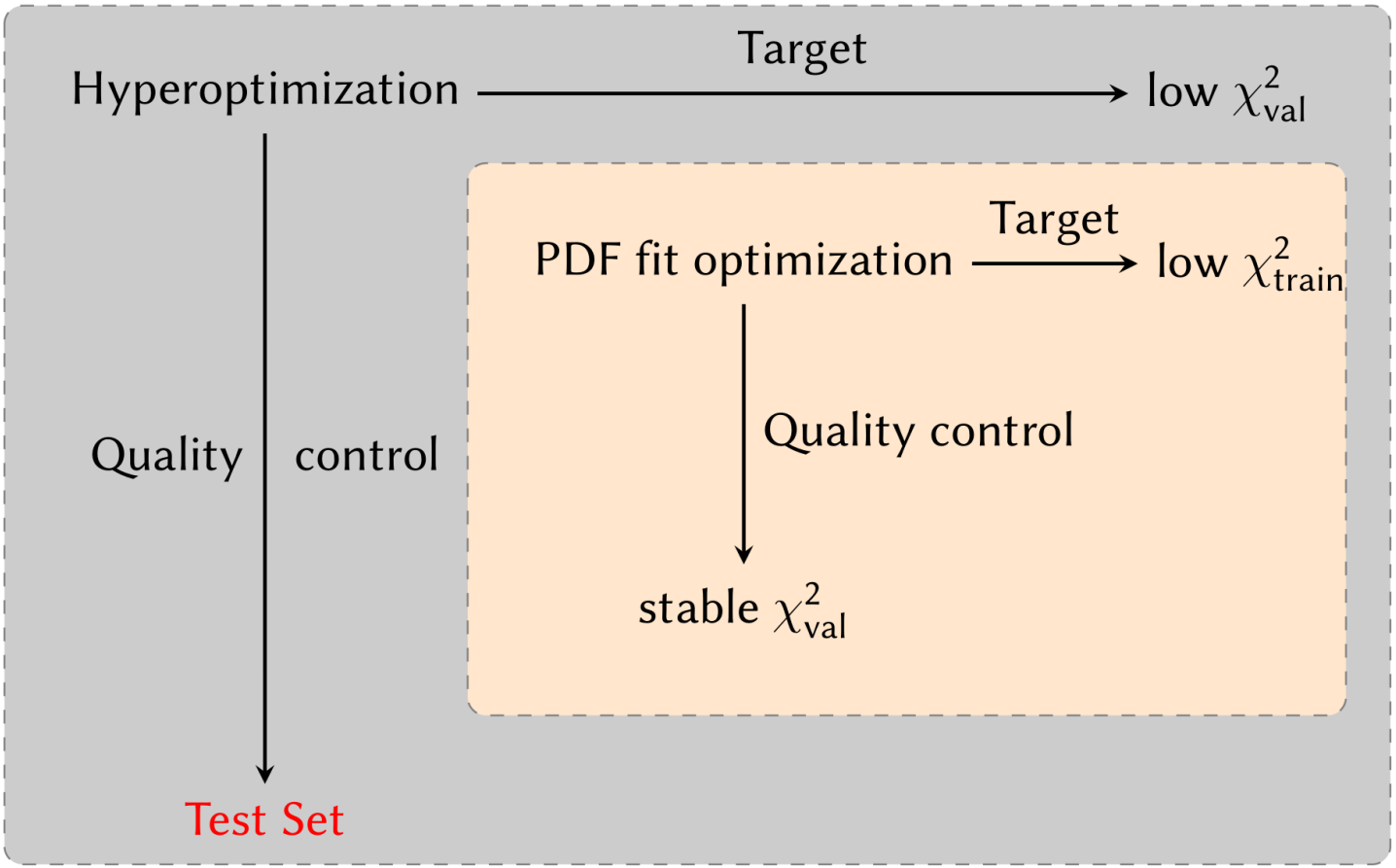
DOWN QUARK: HYPEROPTIMIZED VS. STANDARD



- 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

THE SOLUTION

TUNED HYPEROPTIMIZATION



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

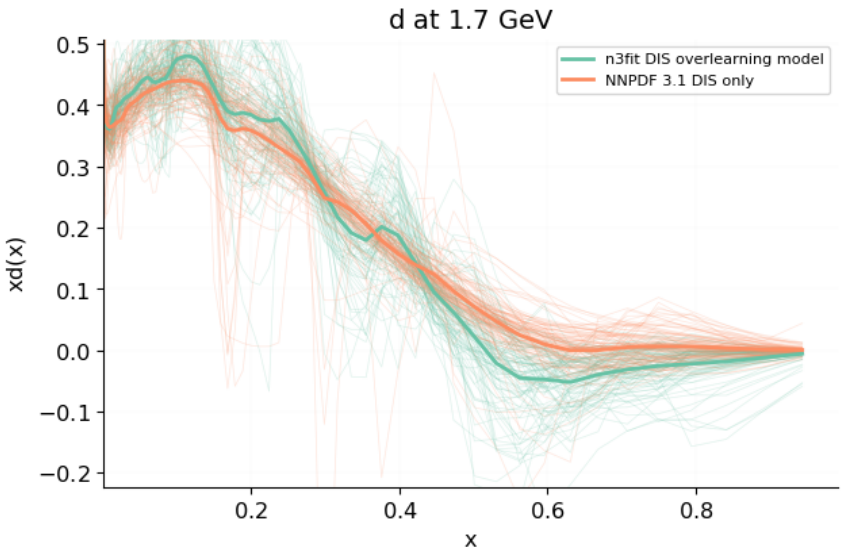
THE TEST SET METHOD

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

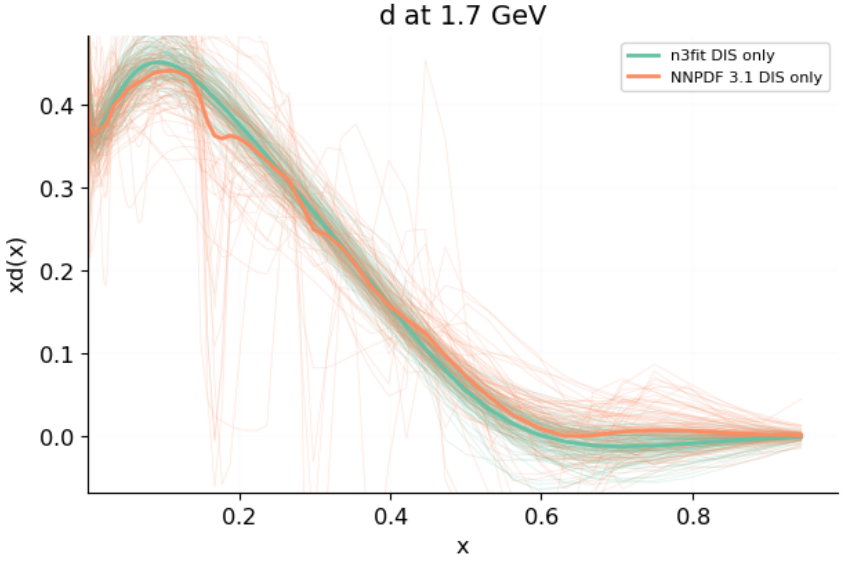
HYPEROPTIMIZED PDFS

DOWN QUARK

N3 OVERFIT VS NNPDF3.1



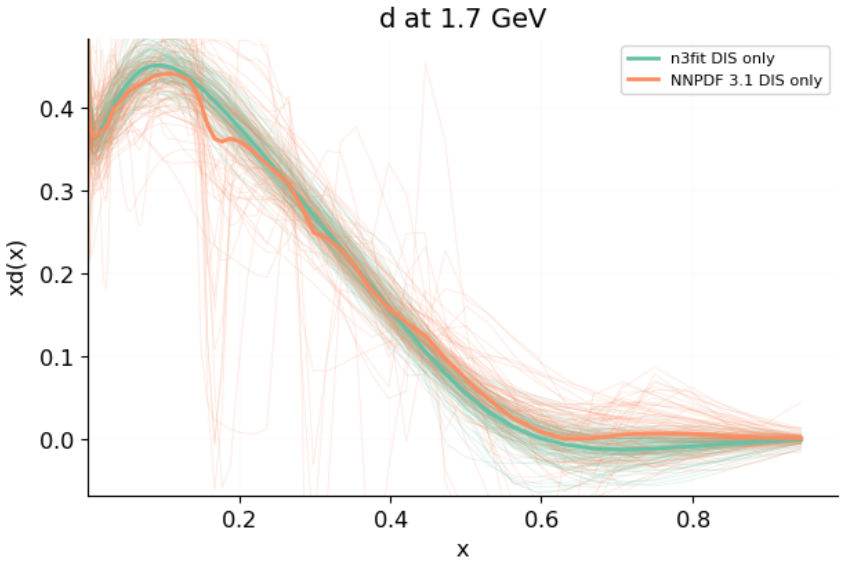
N3FIT VS NNPDF3.1



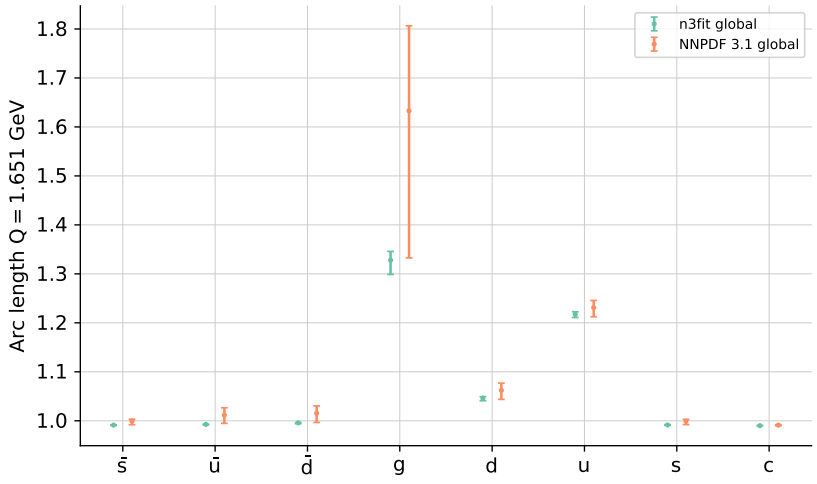
THE TEST SET METHOD

N3FIT VS NNPDF3.1

DOWN PDF



ARCLENGTHS

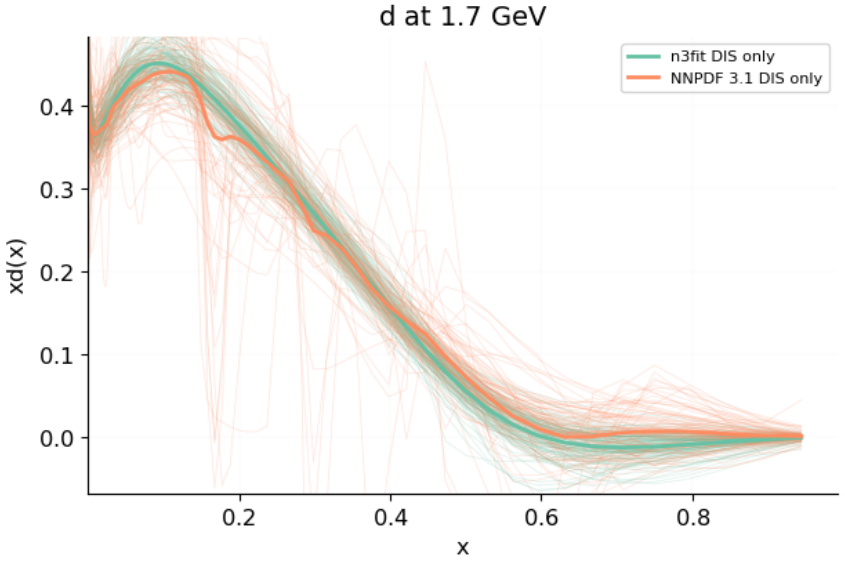


- NO OVERFITTING
- COMPARED TO NNPDF3.1
  - MUCH GREATER STABILITY ⇒ FEWER REPLICAS FOR EQUAL ACCURACY
  - UNCERTAINTIES SOMEWHAT REDUCED

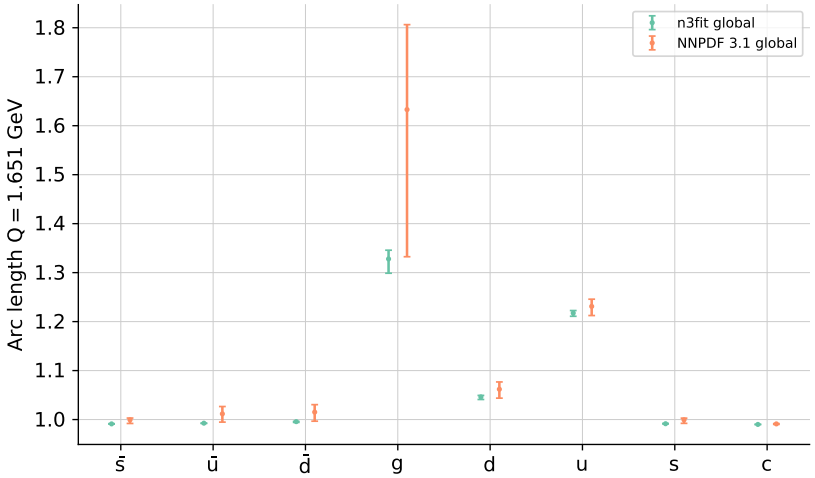
THE TEST SET METHOD

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

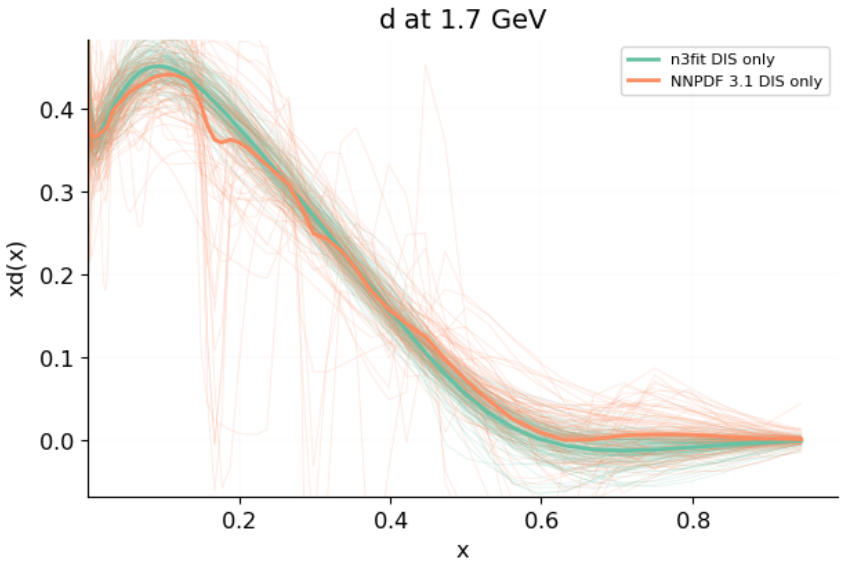
WHO PICKS THE TEST SET?

AUTOMATIC GENERALIZATION
K-FOLDINGS
THE BASIC IDEA:

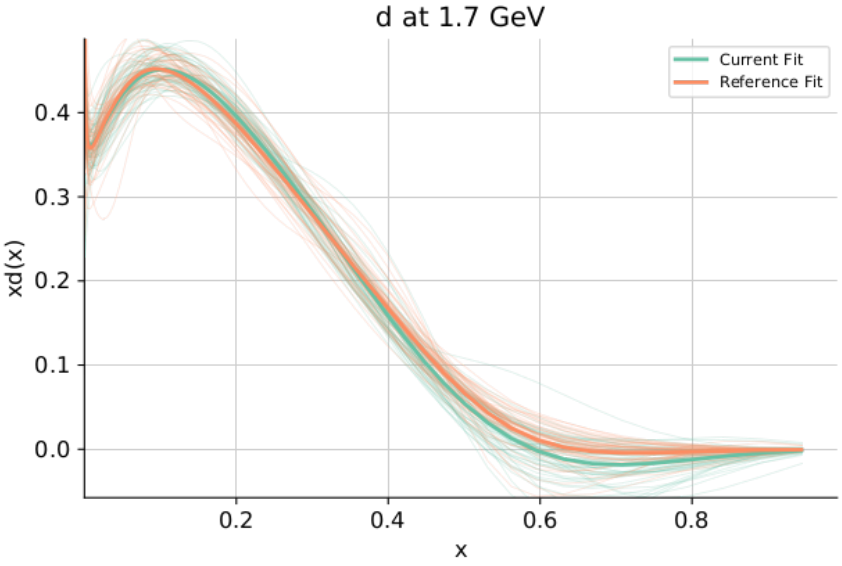
- DIVIDE THE DATA INTO n REPRESENTATIVE SUBSETS
  EACH CONTAINING PROCESS TYPES, KINEMATIC RANGE OF FULL SET
- FIT $n-1$ SETS AND USE n-TH SET AS TEST
  $\Rightarrow$ n VALUES OF $\chi^2_{\text{test, i}}$
- HYPEROPTIMIZE ON MEAN AND STANDARD DEVIATION OF $\chi^2_{\text{test, i}}$
  $\rightarrow$ GOOD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

FOLDED PDFS
DOWN QUARK

N3FIT VS NNPDF3.1



N3FIT-K VS. N3FIT

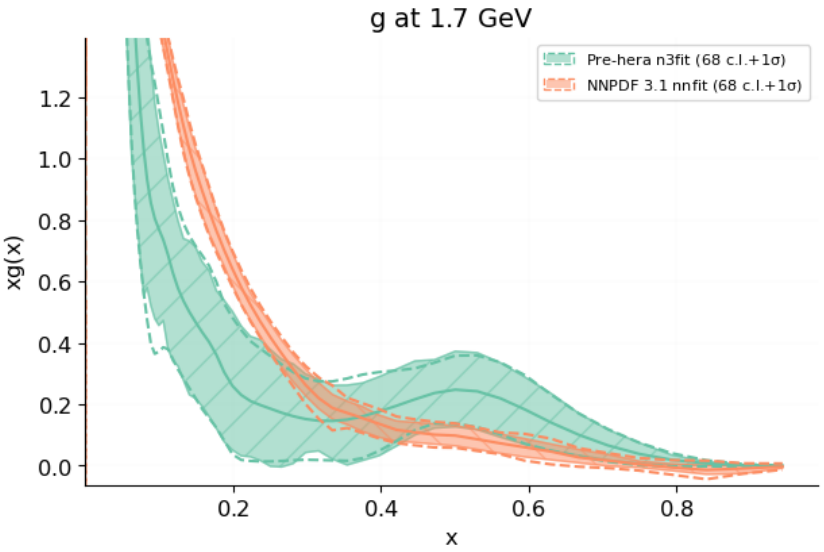


DOES IT WORK?:
THE "FUTURE TEST"

COULD WE "PREDICT" THE RISE OF F_2 AT HERA?

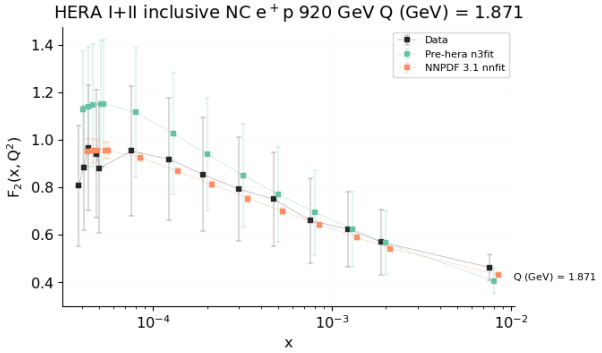
FIT PDFS TO PRE-HERA DATA ONLY

PREDICTED VS TRUE GLUON

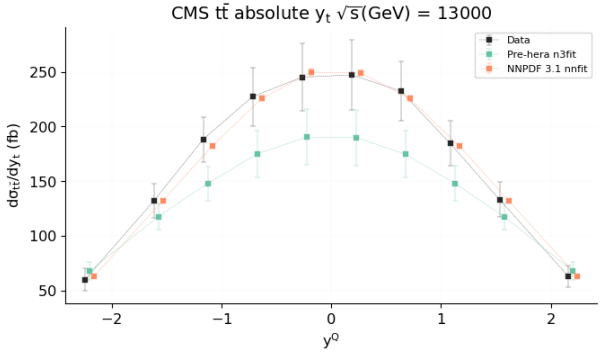


PREDICTION COMPARED TO DATA

HERA F_2



CMS TOP



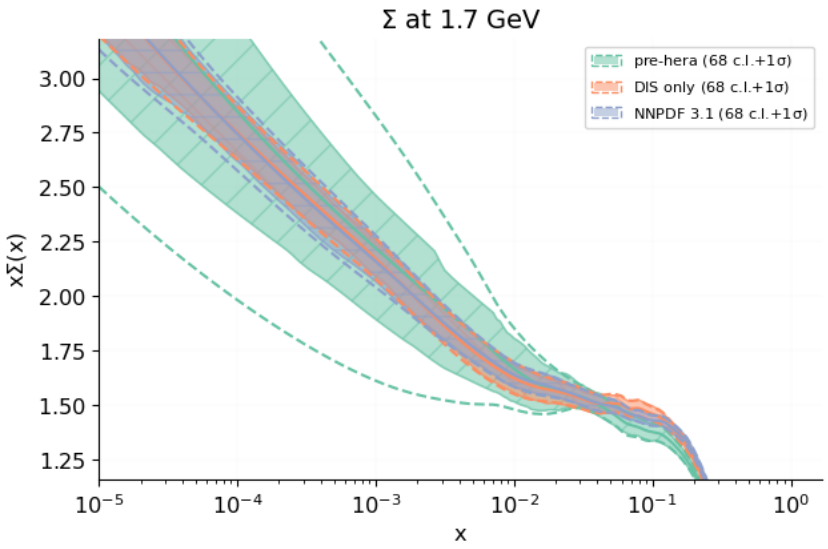
- PDFS ARE FUTURE-COMPATIBLE
- THE DATA ARE WITHIN SHRINKING UNCERTAINTIES
- PREDICTED χ^2/dat=1.20 (WITH PDF UNCERTAINTIES),
  COMPARE TO FITTED χ^2/dat=1.16 (WITHOUT UNCERTAINTIES)

DOES IT WORK?: THE "FUTURE TEST"

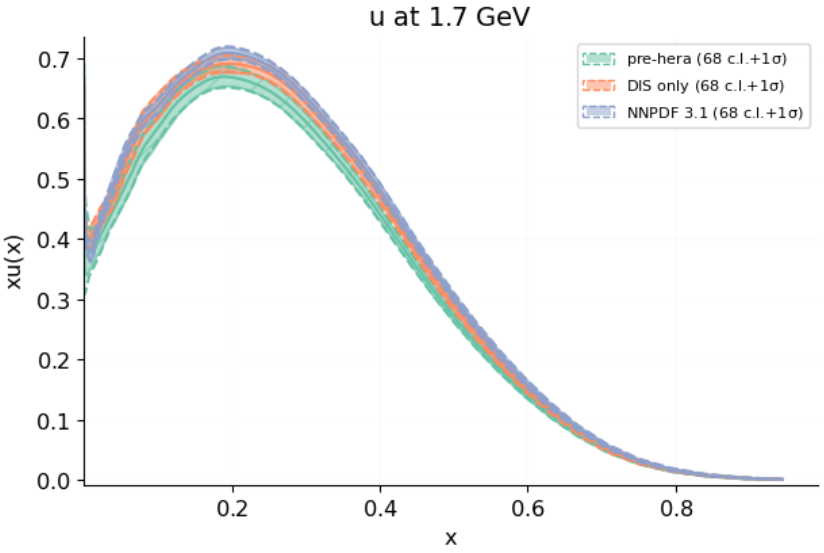
SEQUENTIAL FUTURE TEST DATASETS:

- PRE-HERA
- POST-HERA, PRE-LHC
- LHC RUN I (NNPDF3.1)

QUARK SINGLET



UP QUARK



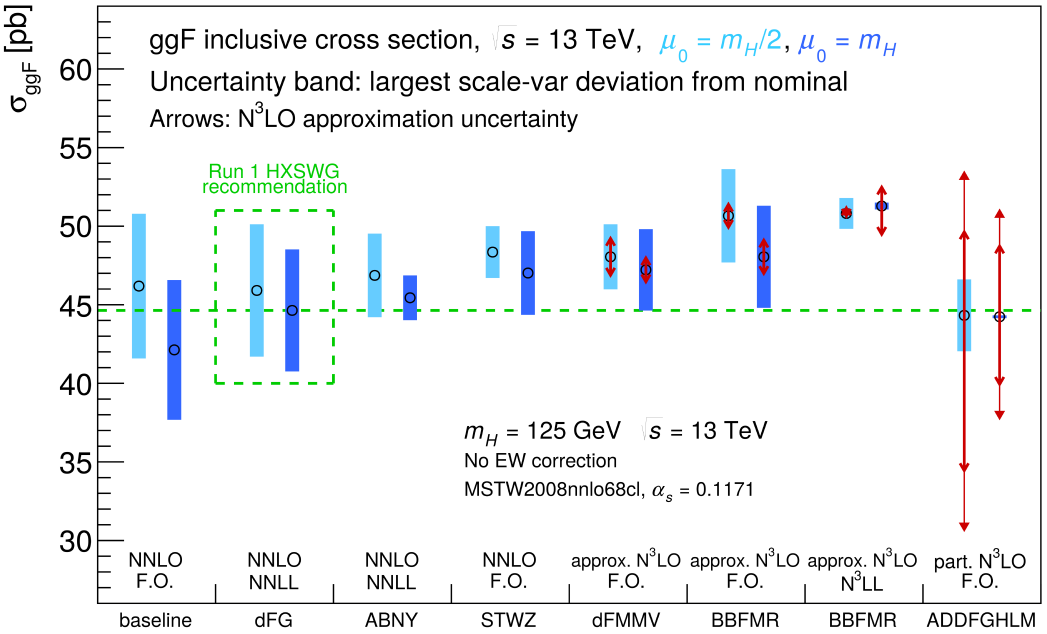
- PDFS ARE FUTURE-COMPATIBLE
- GENERALIZATION FAITHFUL

WHAT ABOUT MISSING HIGHER ORDERS?

MISSING HIGHER ORDERS FROM ASYMPTOTICS

THE GLUON FUSION HIGGS CROSS SECTION: APPROXIMATE N^{3}LO (LHC 13)

HXSWG 2015



APPROXIMATE N^{3}LO+N^{3}LL (Bonvini, Marzani, Muselli, Rottoli, 2016): $48.5^{+1.5}_{-1.9}$PB

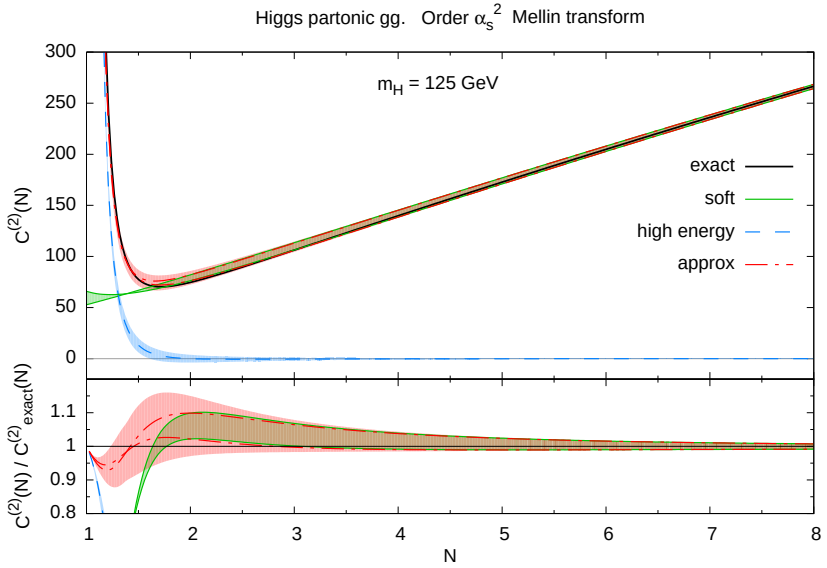
EXACT N^{3}LO+N^{3}LL+LLx: 48.9 ± 1.9PB (HL-LC AND HL-LHC YR, 2019)

MISSING HIGHER ORDERS FROM ASYMPTOTICS

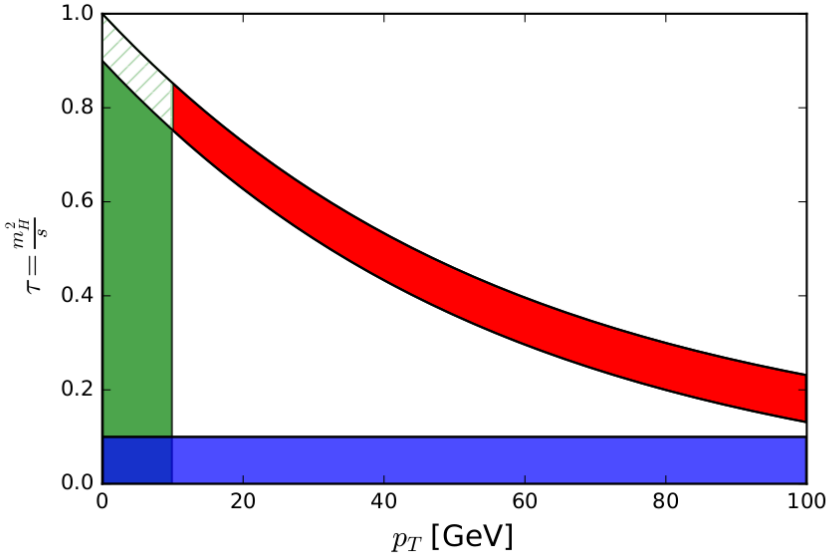
HOW DOES IT WORK?

- TOTAL XSECT: HIGHER ORDERS KNOWN IN VARIOUS KINEMATIC LIMITS FROM RESUMMATION
- CAN IT BE EXTENDED TO DIFFERENTIAL OBSERVABLES?
- CAN WEMACHINE LEARN MHO?

NNLO N-SPACE GGHIGGS
ANALYTIC APPROX VS. EXACT



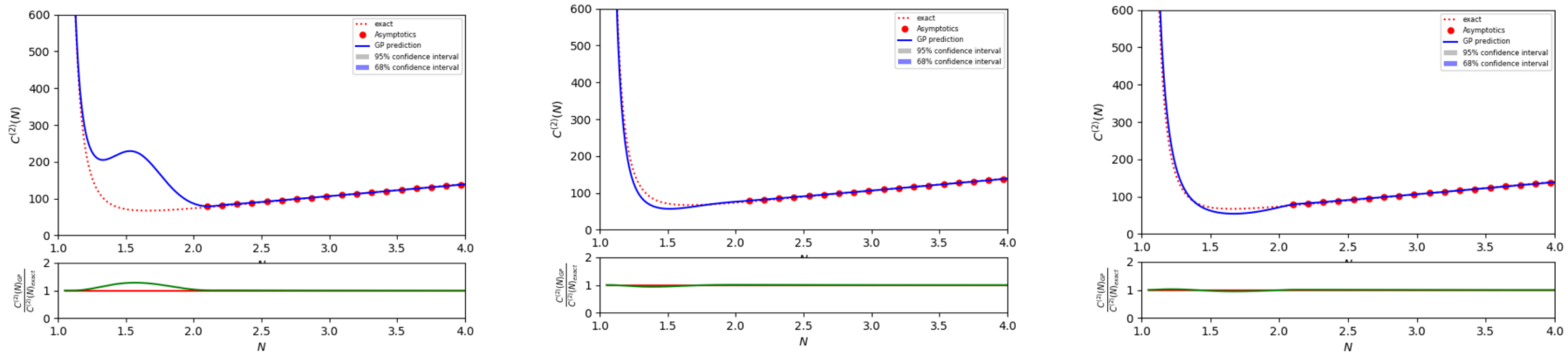
(τ, p_T) RESUMMATION REGIONS



ML EXTRAPOLATION
THE GAUSSIAN PROCESS

- ASSUME $\sigma(x)$ MULTIGAUSSIAN IN FUNCTION SPACE
- DETERMINE THE CORRELATION IN KNOWN REGION ASSUMING KERNEL
- DETERMINE CONDITIONAL DISTRIBUTION IN EXTRAPOLATION
- HYPEROPTIMIZE KERNEL CHOICE AND PARAMETERS BASED ON KNOWN CASES

NNLO N-SPACE GGHIGGS: GAUSSIAN KERNEL INTERPOLATIONS



- TOO FEW DATA $\Rightarrow$ RESULTS UNSTABLE, DEPEND ON CHOICE OF KERNEL

TRANSFER LEARNING?

THE BASIC IDEA:

- PERTURBATIVE DEPENDENCE KNOWN UP TO NNLO FOR MANY PROCESSES
- LEARN PERTURBATIVE DEPENDENCE FROM KNOWN CASES
- ADD FINAL LAYER WHICH EXTRAPOLATES FROM ASYMPTOTICS

....STAY TUNED!

“Estrema temerità mi è parsa sempre quella di coloro che voglion far la capacità umana misura di quanto possa e sappia operar la natura”

“I always found reckless the attitude of those who think that the human capabilities are a measure of what could and might be accomplished by Nature”

Galileo Galilei, “Dialogo sopra i due massimi sistemi del mondo”