REGRESSION NETWORKS: PRECISION AND UNCERTAINTY ESTIMATION

Stefano Forte
Università di Milano & INFN

with tutorials by

Tommaso Giani
NIKHEF

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

Istituto Nazionale di Fisica Nucleare

Machine Learning in Particle Theory

MITP, July 6, 2023

IV: VALIDATION AND TESTING

- GAUSSIANITY
  - THE "HESSIAN" PROJECTION
  - ASSESSING GAUSSIANITY
- CLOSURE TESTING
  - THE CLOSURE TEST AND ITS METRICS
  - THE NATURE OF UNCERTAINTIES
- FUTURE TESTING
  - THE IDEA
  - THE RESULTS
- CORRELATIONS
  - THE NATURE OF PDF CORRELATIONS
  - DATA-INDUCED VS. METHDOLOGY-INDUCED CORRELATIONS

GAUSSIANITY

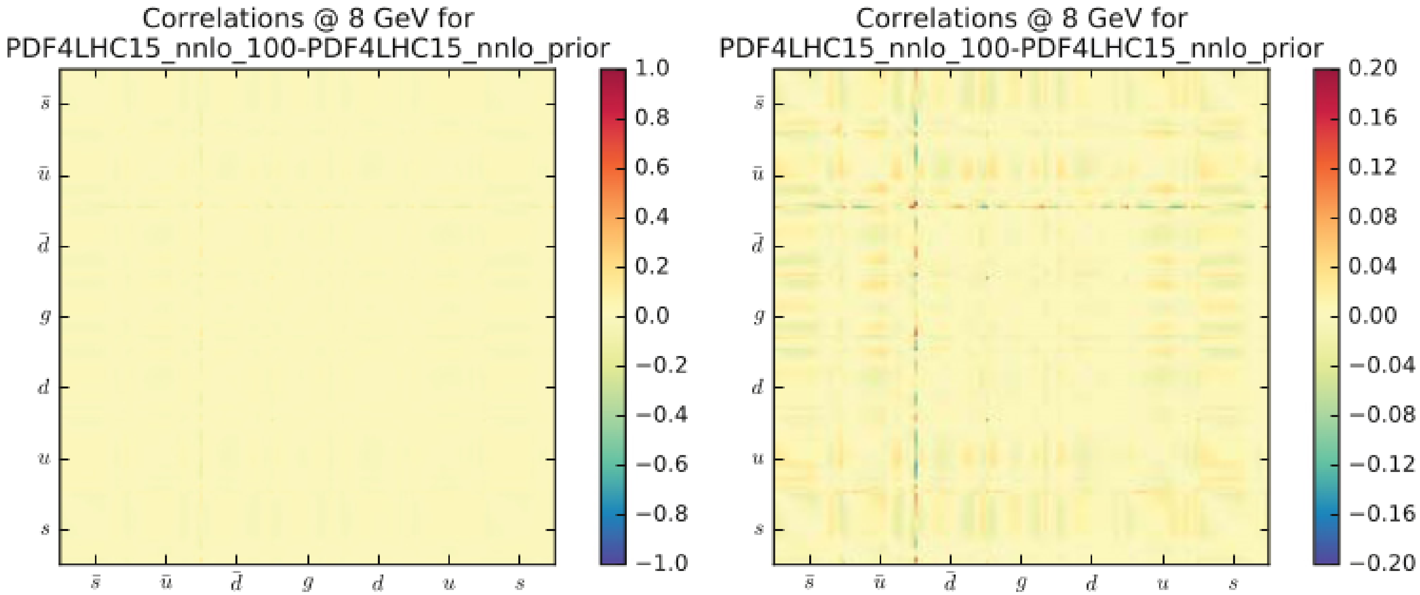
MULTIGAUSSIAN REPRESENTATION

- PARAMETRIC REGRESSION $\Rightarrow$ MAP MULTIGAUSSIAN IN PARAMETER SPACE
  "HESSIAN" $\Rightarrow C^{-1}_{ij} = \partial_i \partial_j \chi^2$

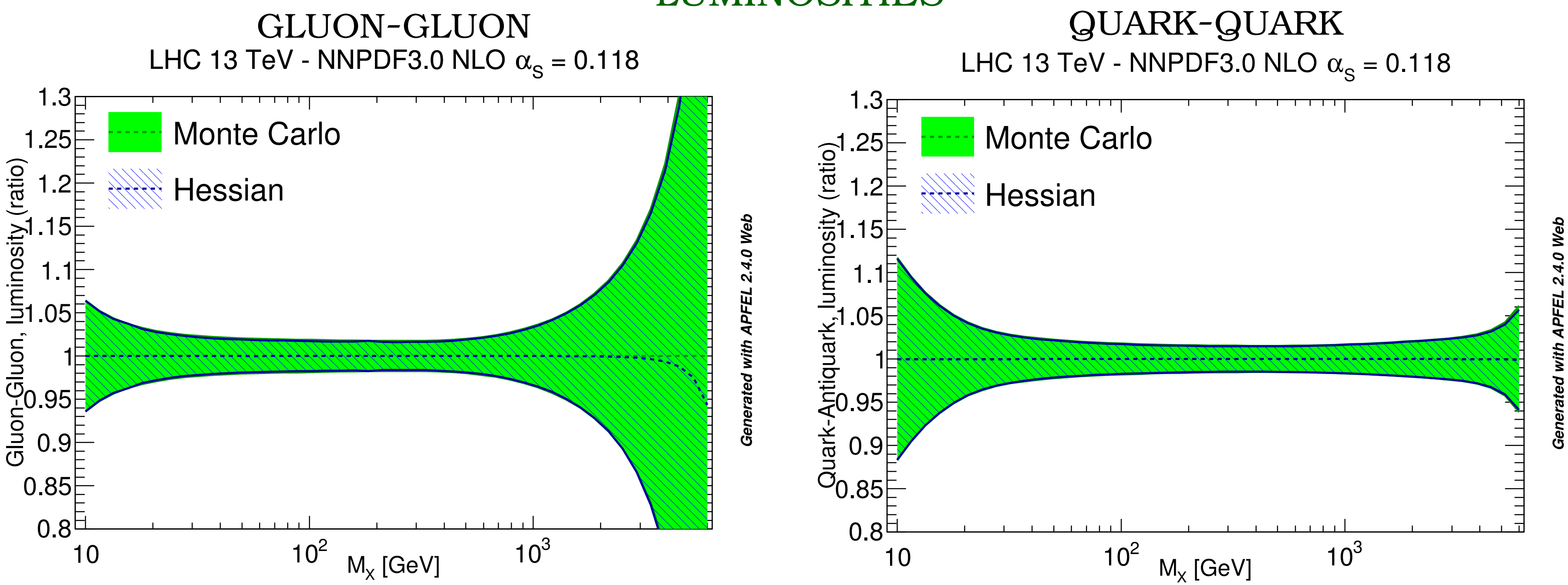
- HESSIAN REPRESENTATION OF MC POSTERIOR:
  - SAMPLE k-TH PDF REPLICA OVER SET OF N_p POINTS $f_i^{(k)}(x_j)$
    i runs over PDF flavors; $\{ij\} = \{p\}$, $p = 1, \ldots, N_p \times N_f$
    $X_{pk} = f_i^{(k)}(x_j) - f_i^{(0)}(x_j)$; $f_i^{(0)}(x_j) \equiv \langle f_i^{(k)}(x_j) \rangle$ REPLICA AVERAGE
  - $C_{pp'} = \frac{1}{N_{\rm rep}} X X^t$ (Cholesky)
  - $X = USV^t$; $U \Rightarrow$ EIGENVECTORS OF C $N_p \times N_{rep}$; $S \Rightarrow$ DIAGONAL NONZERO EIGENVALUE SQRT MATRIX; $V \Rightarrow$ ORTHOGONAL $N_{\rm rep} \times N_{\rm rep}$ (SVD)
  - $C = \frac{1}{N_{\rm rep}} X X^t = \frac{1}{N_{\rm rep}} (US)(US)^t \Rightarrow$ KEEP LARGEST EIGENVALUES

MULTIGAUSSIAN REPRESENTATION vs. MONTECARLO

PDF CORRELATIONS

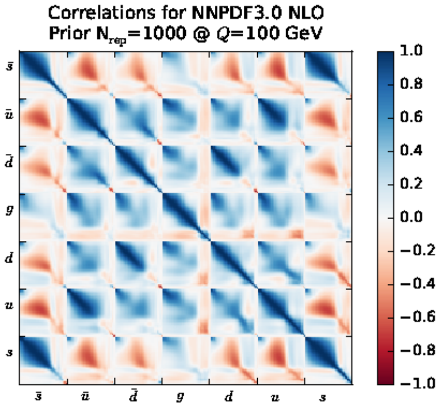


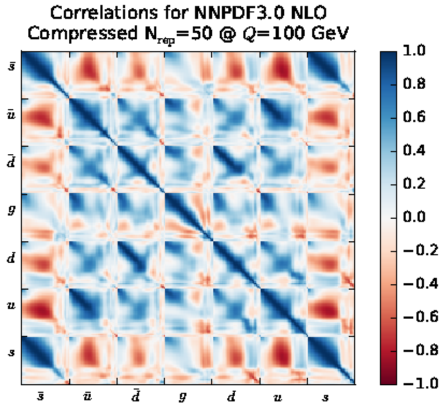
LUMINOSITIES



COMPRESSION

MONTECARLO

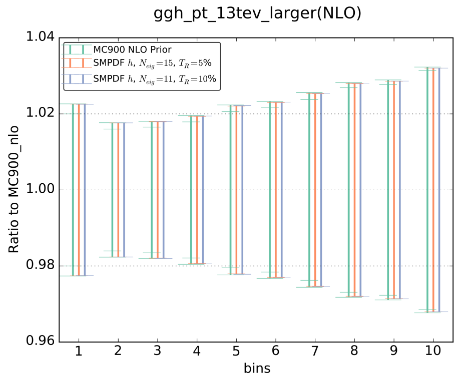


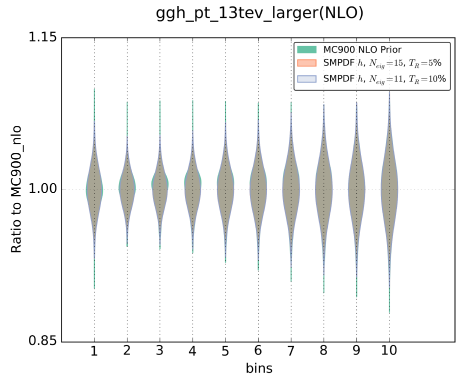


- CONSTRUCT A VERY LARGE REPLICA SAMPLE
- SELECT BY GENETIC ALGORITHM A SUBSET OF REPLICAS WHOSE STATISTICAL FEATURES ARE AS CLOSE AS POSSIBLE TO THOSE OF THE PRIOR
- $\Rightarrow$ FOR ALL PDFS ON A GRID OF POINTS// MINIMIZE DIFFERENCE OF: FIRST FOUR MOMENTS, CORRELATIONS; OUTPUT OF KOLMOGOROV-SMIRNOV TEST (NUMBER OF REPLICAS BETWEEN MEAN AND σ, 2σ, INFINITY)
- 50 COMPRESSED REPLICA REPRODUCE 1000 REPLICA SET TO PRECENT ACCURACY

MULTIGAUSSIAN

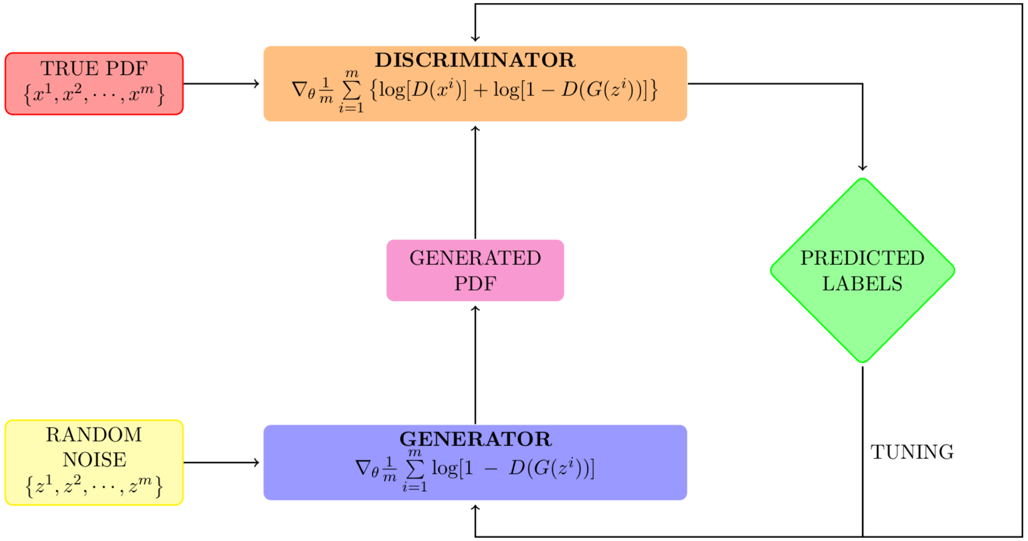
- SELECT SUBSET OF THE COVARIANCE MATRIX CORRELATED TO A GIVEN SET OF PROCESSES
- PERFORM SVD ON THE REDUCED COVARIANCE MATRIX, SELECT DOMINANT EIGENVECTOR, PROJECT OUT ORTHOGONAL SUBSPACE
- ITERATE UNTIL DESIRED ACCURACY REACHED
- 15 EIGENVECTORS DESCRIBE ALL HIGGS MODES + JETS + W, Z PRODUCTION





GAN ENHANCEMENT

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

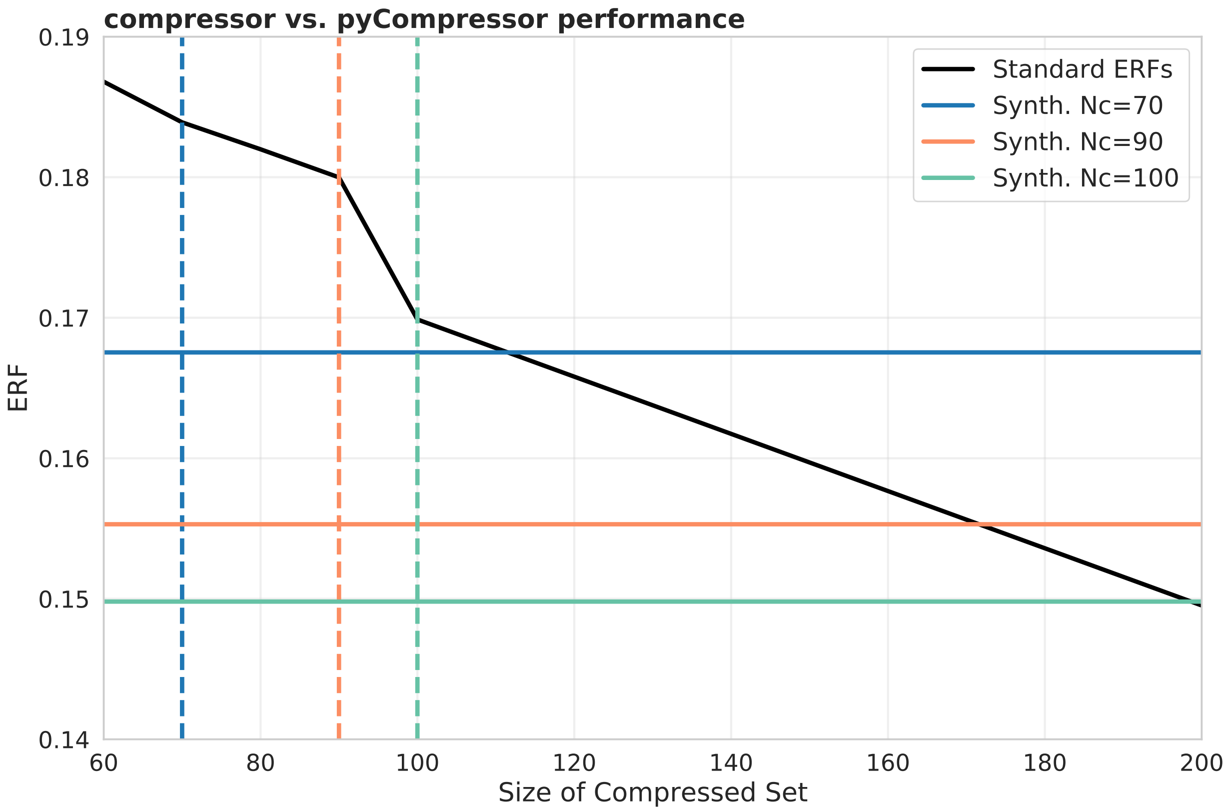


- TRAIN A NETWORK TO SIMULATE THE TRUE DISTRIBUTION (GENERATOR)
- TRAIN A NETWORK TO DISCRIMINATE TRUTH FROM SIMULATION (DISCRIMINATOR)
- TRAIN THE GENERATOR TO TRICK THE DISCRIMINATOR

GAN ENHANCEMENT

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

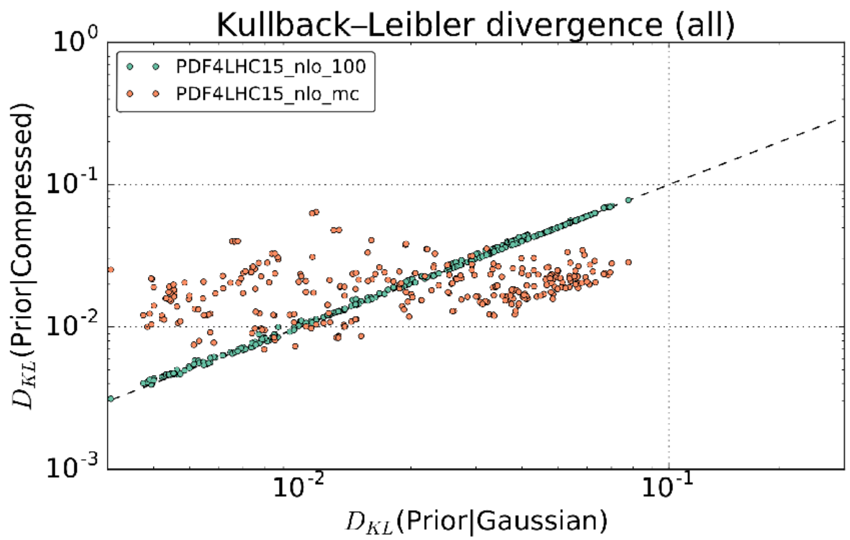
PERFORMANCE



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

ARE UNCERTAINTIES GAUSSIAN?

- REPLICA HISTOGRAM i-TH DATAPOINT z_i FROM MC $\Rightarrow$ CONTINUOUS DISTRIBUTION WITH KDE
  - POINT $\Rightarrow$ KERNEL: $P(z) = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} K\left(z - z_i\right)$;
  - Gaussian kernel $K(z - z_i) \equiv \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{(z-z_i)^2}{h}\right)$
  - Silverman bandwidth $h = \sigma_i \left(\frac{4}{3N_{\rm rep}}\right)^{\frac{1}{5}}$ $\Rightarrow$ MINIMIZES DIFFERENCE TO GAUSSIAN
- DEFINE KULLBACK-LEIBLER DIVERGENCE
  $D_{\rm KL} = \int_{-\infty}^{\infty} P(x) \ln \frac{P(x)}{Q(x)}\, dx$
  BETWEEN A PRIOR P AND ITS REPRESENTATION Q
- COMPUTE $D_{\rm KL}$ MC PRIOR VS REPRESENTATION & MC PRIOR VS GAUSSIAN
- REPRESENTATIONS: MULTIGAUSS OR MC COMPRESSION



- $D_{\rm KL}$ TO GAUSSIAN GENERALLY SMALL
- ONLY FOR FEW POINTS COMPRESSION MORE EFFICIENT THAN MULTIGAUSS CONVERSION

PDF UNCERTAINTIES: DATA

THE CLOSURE TEST:
THE BASIC IDEA

- POSTERIOR REPLICA DISTRIBUTION $\Rightarrow$ APPROXIMATELY GAUSSIAN
- CAN DETERMINE CONFIDENCE LEVEL OF TRUTH ABOUT PREDICTION
  - DATA SPACE
    - IN SAMPLE (USED FOR TRAINING)
    - OUT OF SAMPLE (PREDICTIONS)
  - PDF SPACE
- MEASURABLE BASED ON ASSUMED UNDERLYING TRUTH = RUNS OF THE UNIVERSE
- $n\sigma$ GAUSSIAN CONFIDENCE INTERVAL: $\xi_{n\sigma} = \mathrm{erf}\left(\frac{n\sigma}{\sqrt{2}\sigma^0}\right)$

  $n\sigma \to$ CONFIDENCE INTERVAL; $\sigma^0 \Rightarrow$ WIDTH OF GAUSSIAN

STATISTICAL INDICATORS

- BIAS $b = \frac{1}{N_{\text{points}}} \sum_{i=1}^{N_{\text{points}}} (\mathcal{G}_i(f) - z_i)^2$

  $(\mathcal{G}_i(f) = \langle \mathcal{G}_i(f) \rangle \equiv \frac{1}{N_{\text{replicas}}} \sum_{j=1}^{N_{\text{replicas}}} \mathcal{G}_i(f_j)$ prediction, z_i true);

  NORMALIZED: $\frac{1}{N_{\text{points}}} |\mathcal{G}(f) - z|_C^2,$

  C COVARIANCE MATRIX:
  - DATA $\Rightarrow$ FROM EXPERIMENT
  - PDF $\Rightarrow$ FROM REPLICAS
- VARIANCE $v = \frac{1}{N_{\text{points}}} \sum_{i=1}^{N_{\text{points}}} \sigma_i^2$; $\sigma_i = \langle (\mathcal{G}_i(f) - \langle \mathcal{G}_i(f) \rangle)^2 \rangle$; NORMALIZED

  $v = |\mathcal{G}(f) - \langle \mathcal{G}(f) \rangle|_C^2$
- BIAS-VARIANCE RATIO $R_{bv} = \sqrt{\frac{b}{v}}$: AVERAGED OVER RUNS OF THE UNVERSE (RUs)
- EMPIRICAL CONFIDENCE LVL $\xi_{n\sigma} = \frac{1}{N_{\text{points}}} \sum_{i=1}^{N_{\text{points}}} I_{[-n\sigma, n\sigma]} (\langle \mathcal{G}_i(f) \rangle - z_i)$ OVER RUs

CLOSURE TEST IMPLEMENTATION

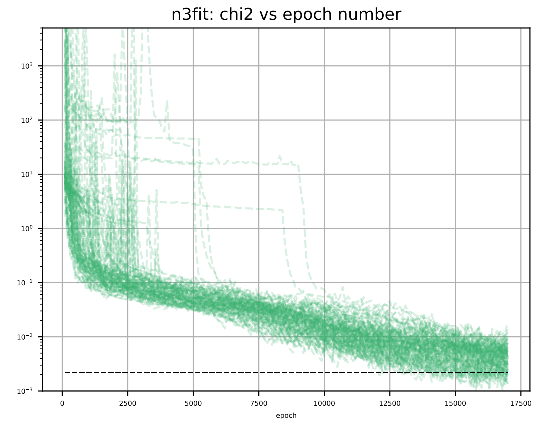
- ASSUME UNDERLYING "TRUTH" PDF (SAY A RANDOM PDF REPLICA)
- GENERATE DATA ACCORDING TO STATISTICAL AND CORRELATED SYSTEMATICS (SAY FOR NNPDF4.0 DATASET)
- DETERMINE PDFS & COMPARED TO "TRUTH" BASED ON INDICATORS

THE NATURE OF UNCERTAINTIES

- LEVEL 0:
  - EACH DATAPOINT EQUAL TO THE "TRUTH VALUE"; ZERO UNCERTAINTY
  - FIT $\rightarrow$ MUST FIND $\chi^2 = 0$ (GET BACK "TRUTH")
  - $\chi^2 \approx 0$ BOTH REPLICA TO REPLICA AND AVERAGE TO TRUTH
  - INTERPOLATION/EXTRAPOLATION UNCERTAINTY
- LEVEL 1:
  - EACH PSEUDO- DATAPOINT IS OBTAINED AS A RANDOM FLUCTUATION WITH GIVEN COVARIANCE MATRIX ABOUT "TRUTH"
    $\Rightarrow$ "RUN OF THE UNIVERSE"
  - FIT DATA OVER AND OVER AGAIN
  - $\chi^2 \approx 1$ BOTH REPLICA TO REPLICA AND AVERAGE TO TRUTH
  - FUNCTIONAL UNCERTAINTY
- LEVEL 2:
  - DATA AS IN LEVEL 1
  - GENERATE DATA REPLICAS OF THESE "DATA"
  - FIT PDF REPLICAS TO DATA REPLICAS
  - $\chi^2 \approx 2$ REPLICA TO REPLICA; $\chi^2 \approx 1$ AVERAGE TO TRUTH
  - DATA UNCERTAINTY

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

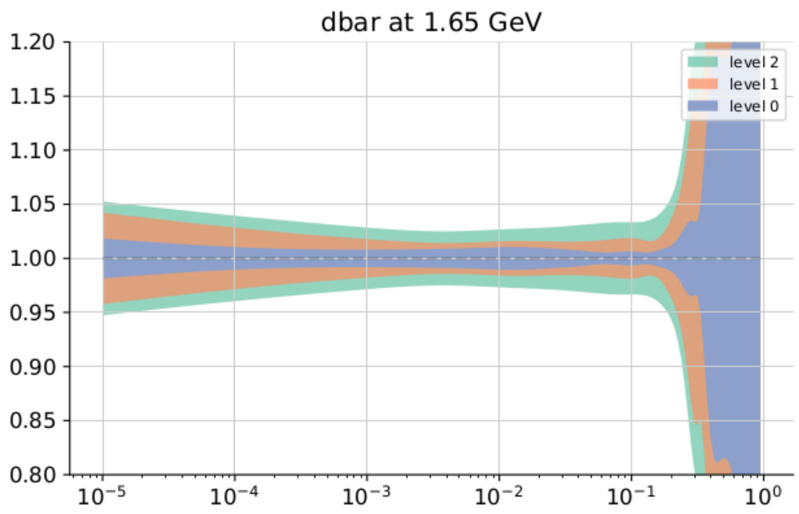
LEVEL 0 χ^2 VS TRAINING



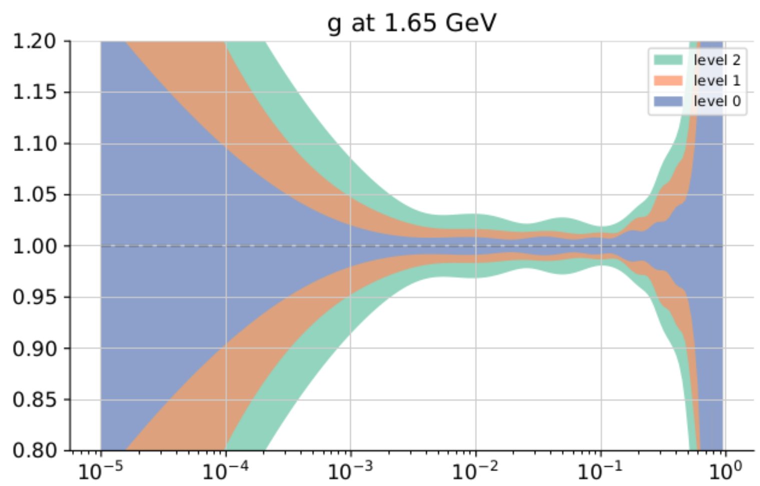
- LEVEL 0 (TRUTH DATA) $\Rightarrow$ $\chi^2 \approx 0$, YET UNCERTAINTY NONZERO
  $\Rightarrow$ NEURAL NETS $\Leftrightarrow$ MANY FUNCTIONAL FORMS
- LEVEL 1 (RUNS OF UNIVERSE) $\Rightarrow$ REPLICAS ALL FITTED TO SAME DATA,
  YET UNCERTAINTY NONZERO
  $\Rightarrow$ DITTO
- LEVEL 0, 1 AND 2 UNCERTAINTIES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

LEVEL 0/1/2 UNCERTAINTIES

ANTIDOWN



GLUON



TESTING: THE INDICATORS

BIAS/VARIANCE RATIO AND ONE-σ QUANTILE

DATA-SPACE, DATA COVARIANCE MATRIX, OUT-OF-SAMPLE

Dataset	$\sqrt{\mathrm{b/v}}$	$\xi^{(\mathrm{data})}_{1\sigma}$	$\mathrm{erf}(R_{bv}/\sqrt{2})$
DY	0.99 ± 0.08	0.69 ± 0.02	0.69 ± 0.04
Top-pair	0.75 ± 0.06	0.75 ± 0.03	0.82 ± 0.03
Jets	1.14 ± 0.05	0.63 ± 0.03	0.62 ± 0.02
Dijets	0.99 ± 0.07	0.70 ± 0.03	0.69 ± 0.04
Direct photon	0.71 ± 0.06	0.81 ± 0.03	0.84 ± 0.03
Single top	0.87 ± 0.07	0.69 ± 0.04	0.75 ± 0.04
Total	1.03 ± 0.05	0.68 ± 0.02	0.67 ± 0.03

PDF-SPACE & COV MATRIX

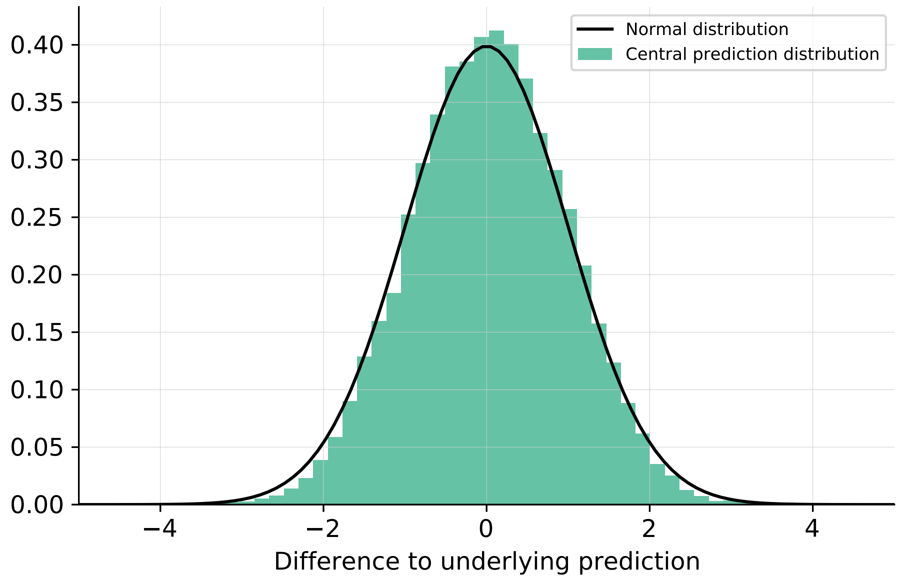
flavour	$\xi^{(\mathrm{pdf})}_{1\sigma}$
Σ	0.82 ± 0.04
g	0.70 ± 0.05
V	0.65 ± 0.05
V_3	0.63 ± 0.05
V_8	0.72 ± 0.04
T_3	0.71 ± 0.05
T_8	0.71 ± 0.05
Total	0.71 ± 0.02

- 25 “UNIVERSE RUNS”, 45 REPLICAS EACH
- IN-SAMPLE DATA: PRE 2015
- OUT OF SAMPLE DATA: 2015-2020, MOSTLY LHC
- PDFS HIGHLY CORRELATED $\Rightarrow$ SAMPLED AT 4 POINTS EACH

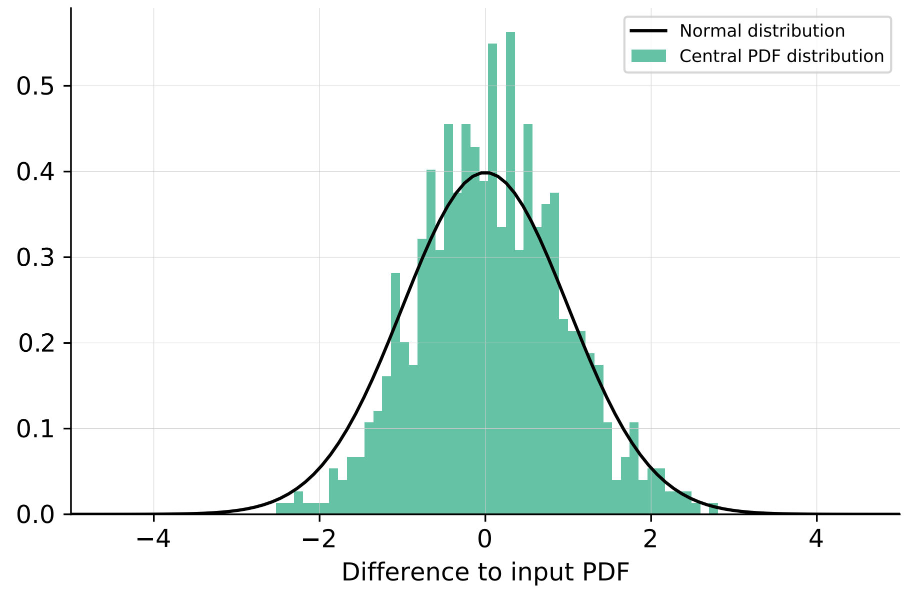
TESTING: THE INDICATORS

DISTRIBUTION OF DEVIATIONS FROM TRUTH

DATA SPACE (OUT OF SAMPLE)



PDF SPACE



- PDF-SPACE MORE NOISY THAN DATA SPACE

ASIDE: ERRORS IN MC ESTIMATES

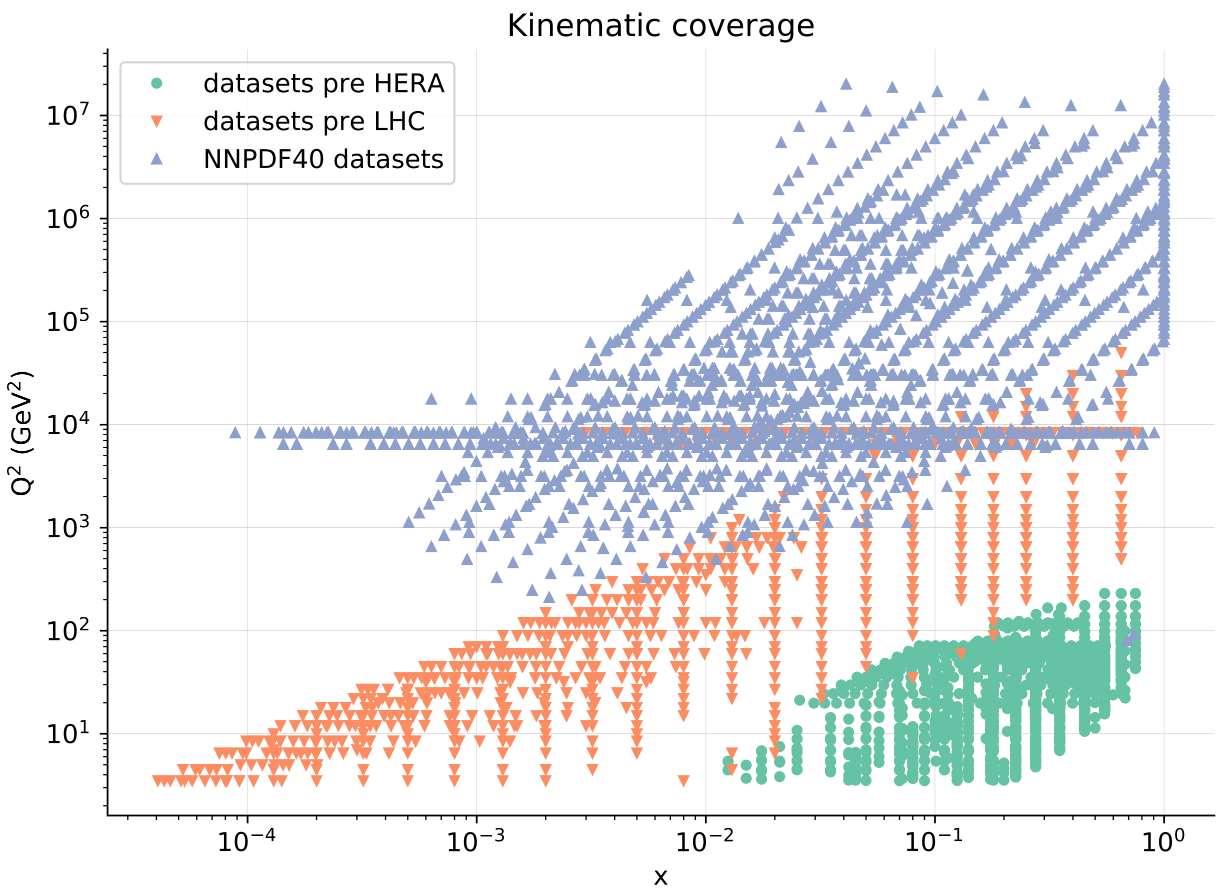
THE JACKNIFE/BOOSTRAP METHOD

- GIVEN $N_{\rm est}$ ESTIMATES x^i OF x, COMBINED ESTIMATE $x = \langle x \rangle \pm \sigma$, $\langle x \rangle = \frac{1}{N_{\rm est}} \sum_{i=1}^{N_{\rm est}} x^i$; $\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle$.
- GIVEN A MC SAMPLE OF ESTIMATES, EXTRACT RANDOMLY $n < N_{\rm est}$ FROM IT WITH REPLACEMENT $\Rightarrow$ EXTRACTIONS INDEPENDENT, REPETITIONS ALLOWED
  JACKNIFE: $n = N_{\rm est} - 1$
- REPEAT EXTRACTION N_B TIMES $\Rightarrow$ N_b SAMPLES OF n REPLICAS
- COMPUTE AVERAGE FOR EACH EXTRACTED n-REPLICA SAMPLE
- AVERAGE OF THESE EQUAL TO THE STARTING SAMPLE AVERAGE
- COMPUTE $\langle x \rangle$ FROM FULL SAMPLE, ESTIMATE UNCERTAINTY ON IT FROM VARIANCE OF BOOTSTRAP EXTRACTIONS

PDF UNCERTAINTIES:
EXTRAPOLATION

FUTURE TESTS

IDEA: USE (REAL) HIERARCHICAL DATASETS



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

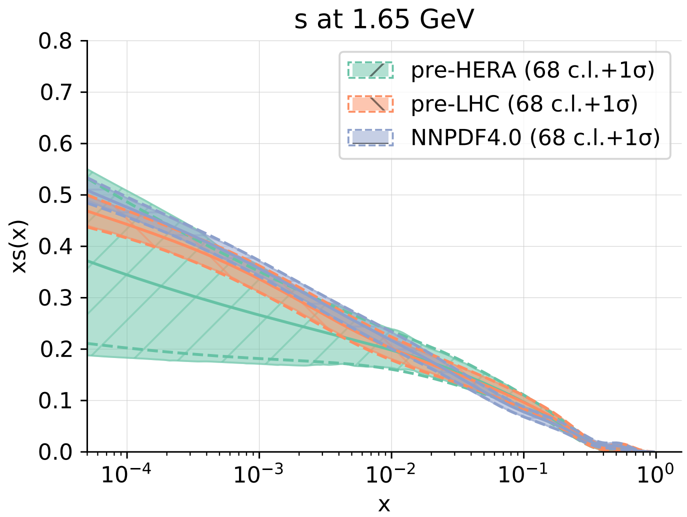
ASSESSING EXTRAPOLATION UNCERTAINTIES

FUTURE TEST RESULTS (NNPDF4.0)

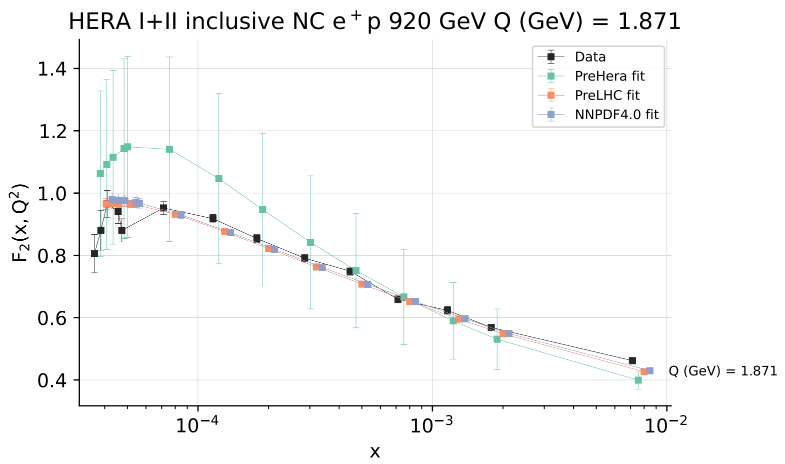
χ^2: FITTED VS EXTRAPOLATED: WITHOUT/WITH PDF UNC.

Process	pre-HERA	pre-LHC	NNPDF4.0
FT DIS (NC)	1.05	1.18	1.23
FT DIS (CC)	0.80	0.85	0.87
FT DY	0.92	1.27	1.59
HERA	27.20/1.23	1.22	1.20
Coll. DY (Tev.)	5.52/1.02	0.99	1.11
Coll. DY (LHC)	18.91/1.31	2.63/1.58	1.53
Top quark	20.01/1.06	1.30/0.87	1.01
Jets	2.69/0.98	2.12/1.10	1.26
Total out of sample	19.48/1.16	2.10/1.15	–

strange PDF



HERA F_2^p



PDFs ARE FUTURE-COMPATIBLE!

PDF CORRELATIONS

CORRELATION BETWEEN MODEL FEATURES

example: up vs down PDFs

COVARIANCE: $\mathrm{Cov}[u,d](x,x') = \langle u(x,Q_0^2) d(x',Q_0^2)\rangle - \langle u(x,Q_0^2)\rangle\langle d(x',Q_0^2)\rangle;$

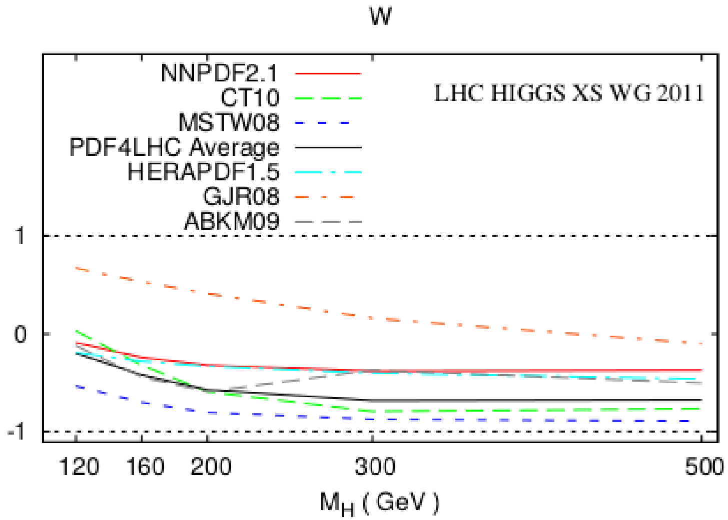
CORRELATION: $\rho[u,d](x,x') = \frac{\mathrm{Cov}[u,d](x,x')}{\sqrt{\mathrm{Var}[u](x)\mathrm{Var}[d](x')}}$

COMPUTATION IN MC APPROACH: $\langle u(x,Q_0^2) d(x',Q_0^2)\rangle = \frac{1}{N}\sum_{r=1}^{N} u^{(r)}(x,Q_0^2) d^{(r)}(x',Q_0^2);$
$u^{(r)}(x,Q_0^2)$ REPLICAS

- CORRELATION INDUCED BY DATA, THEORY (E.G. SUM RULES), METHODOLOGY (E.G. ASSUMPTIONS ON EXTRAPOLATION)

- USED E.G. TO ASSESS CORRELATION BETWEEN SIGNAL AND BACKGROUND PROCESSES

PDF-INDUCED CORRELATIONS
BETWEEN HIGGS SIGNAL & BACKGROUND PROCESSES (HXSWG, YR2, 2011)
Higgs in gluon fusion vs. W production



CORRELATIONS BETWEEN MODELS

CORRELATE PDFs IN DIFFERENT SETS

example: up NN model vs down parametric model

$$\mathrm{Cov}[u^N, d^P](x, x') = \langle u^{\mathrm{N}}(x, Q_0^2) d^{\mathrm{P}}(x', Q_0^2)\rangle - \langle u^{\mathrm{N}}(x, Q_0^2)\rangle\langle d^{\mathrm{P}}(x', Q_0^2)\rangle$$

S-CORRELATION VS F-CORRELATION

$\rho[u^N, u^P]$ DIFFERENT SETS, SAME PDF VS. $\rho[u^N, d^N]$ SAME SET, DIFFERENT PDFs

- SAME REPLICA MUST BE USED FOR NONZERO CORRELATION:

  IF REPLICAS UNCORRELATED $\langle u(x, Q_0^2) d(x, Q_0^2)\rangle \overset{?}{=} \frac{1}{N}\sum_{r=1}^N u^{(r)}(x, Q_0^2) d^{(r')}(x, Q_0^2) = \langle u\rangle\langle d\rangle$

  THEN CORRELATION VANISHES

REPLICA CORRELATION

- FIT PDF REPLICAS $f_i^{(r,\,\mathrm{N})}(x, Q_0^2)$ & $f_i^{(r,\,\mathrm{P})}(x, Q_0^2)$ for all x, i TO SAME DATA REPLICA

- COMPUTE COVARIANCE & CORRELATION USING

$$\langle u(x, Q_0^2) d(x, Q_0^2)\rangle = \frac{1}{N}\sum_{r=1}^N u^{(r,\,\mathrm{N})}(x, Q_0^2) d^{(r,\,\mathrm{P})}(x, Q_0^2)$$

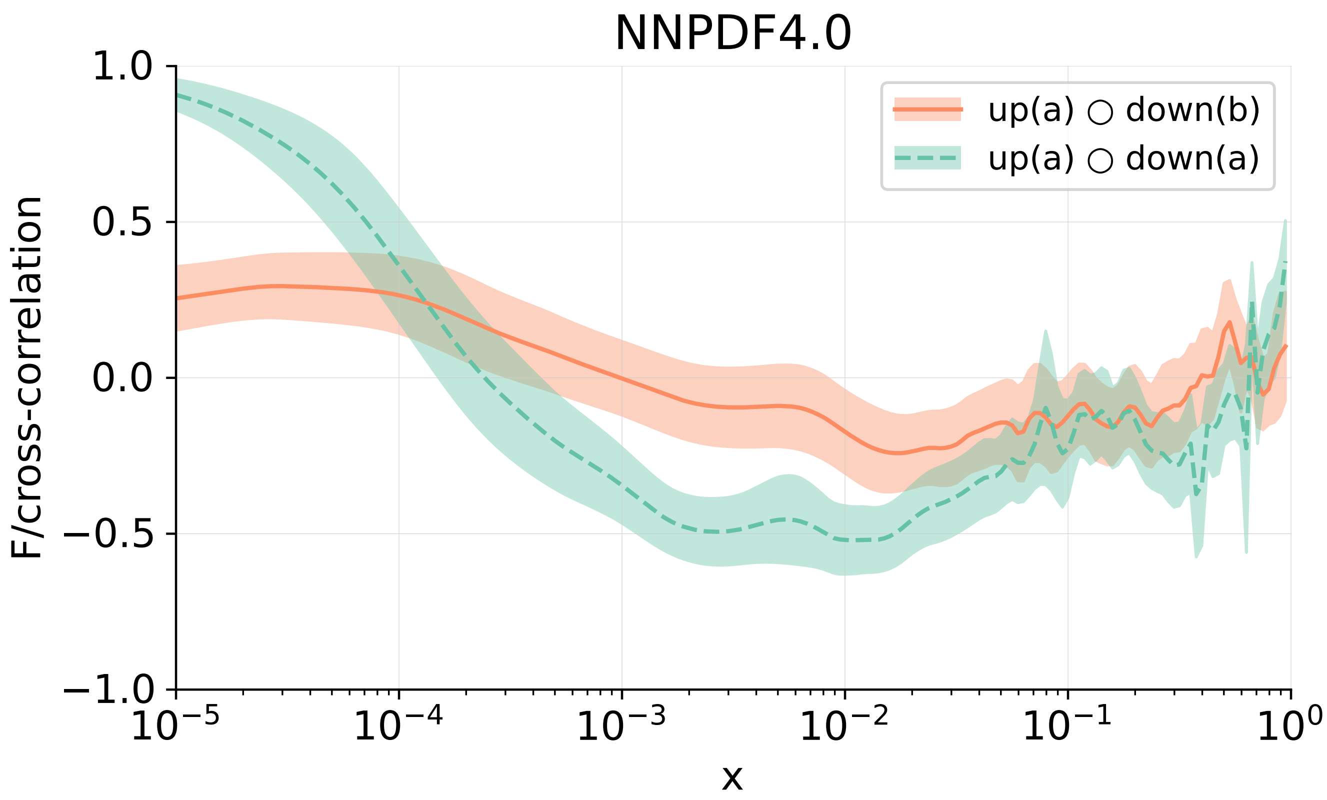
DATA VS METHOOLOGY CORRELATION

- NONZERO LEVEL-1 UNCERTAINTY $\Rightarrow$ DATA REPLICA DOES NOT DETERMINE UNIQUELY THE PDF REPLICA
- IN PRINCIPLE FULL CORRELATION: $r \Leftrightarrow$ DATA REPLICA AND $r' \Leftrightarrow$ LEVEL-1 (METHDOLOLOGY) REPLICAS
  REPLICAS (UP QUARK) $u^{(r,r')}(x, Q_0^2)$;

$$\left| \frac{1}{N}\sum_{r=1}^{N} u^{(r,r')}(x,Q_0^2) d^{(r,r'')}(x,Q_0^2) - \langle u\rangle\langle d\rangle \right| \leq \left| \frac{1}{NM}\sum_{r=1}^{N}\sum_{r'=1}^{M} u^{(r,r')}(x,Q_0^2) d^{(r,r')}(x,Q_0^2) - \langle u\rangle\langle d\rangle \right|$$

- IN PRACTICE METHODOLOGY CORRELATION NOT INCLUDED $\Rightarrow$ CORRELATION LOSS

FULL VS DATA-INDUCED



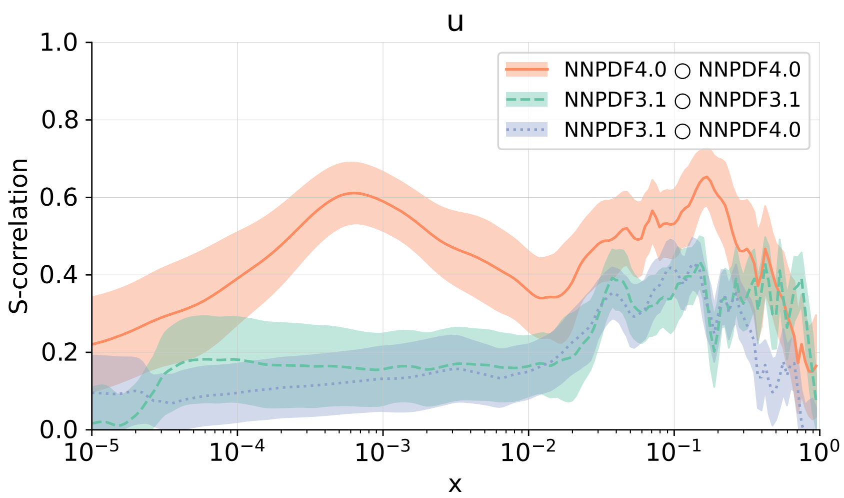
MEASURING METHODOLOGY DECORRELATION

- SELF-CORRELATION: S-CORRELATION OF A PDF SET TO ITSELF
  = F-CORRELATION OF A PDF TO ITSELF
- USE TWO DIFFERENT SETS OF PDF REPLICAS FITTED TO THE SAME DATA REPLICAS

$$\langle u(x,Q_0^2)u(x,Q_0^2)\rangle = \frac{1}{N}\sum_{r=1}^{N} u^{(r,r')}(x,Q_0^2)u^{(r,r'')}(x,Q_0^2)$$

- DEVIATION OF CORRELATION FROM 100% MEASURES THE CORRELATION LOSS $\Rightarrow$ UNCORRELATED FUNCTIONAL UNCERTAINTY
- HIGHER CORRELATION $\Rightarrow$ MORE EFFICIENT METHODOLOGY

up quark



gluon