PDFS: FROM NN TO ML

Stefano Forte
Università di Milano & INFN

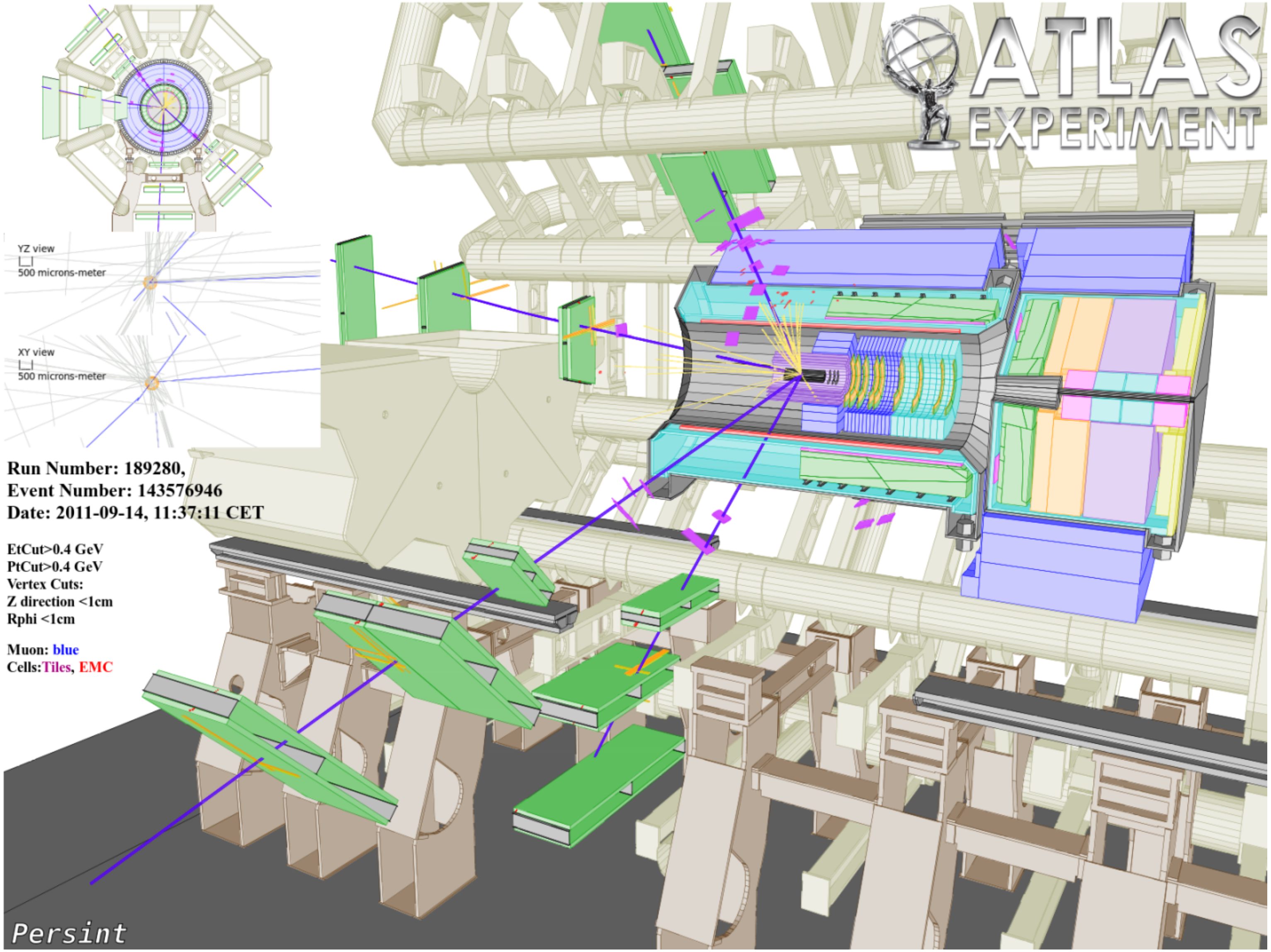
UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

INFN
Istituto Nazionale di Fisica Nucleare

Machine Learning for Pheno

IPPP, Durham, April 4,2018

PHYSICS AT THE LHC



FACTORIZATION

THE PARTON LUMINOSITY

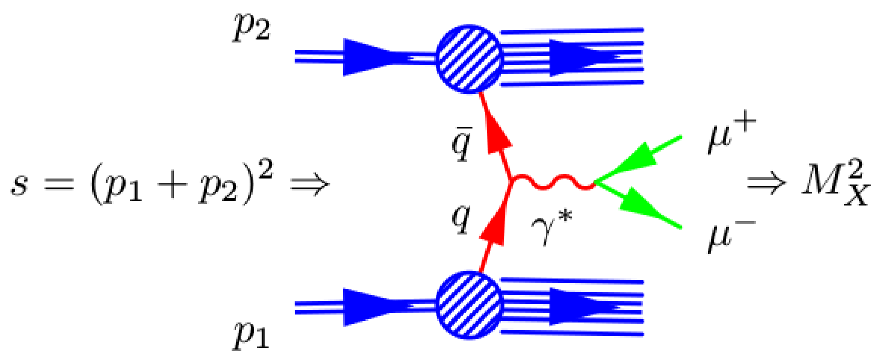
$$\sigma_X(s, M_X^2) = \sum_{a,b} \int_{x_{\min}}^1 dx_1\, dx_2\, f_{a/h_1}(x_1) f_{b/h_2}(x_2) \hat{\sigma}_{q_a q_b \to X}\left(x_1 x_2 s, M_X^2\right)$$

$$\sigma_X(s, M^2) = \sum_{a,b} \int_\tau^1 \frac{dx}{x} \mathcal{L}_{ab}\left(\frac{\tau}{x}\right) \hat{\sigma}\left(x, \alpha_s(M_H^2)\right) = \sum_{a,b} \mathcal{L}_{ab} \otimes \hat{\sigma}_{ab}$$

- PARTON LUMINOSITY $\mathcal{L}_{ab}(\tau) = \int_\tau^1 \frac{dx}{x} f_{a/h_1}(x) f_{b/h_2}(\tau/x) = f_a \otimes f_b$
- PARTONIC CROSS SECTION $\hat{\sigma}_{q_a q_b \to X}$

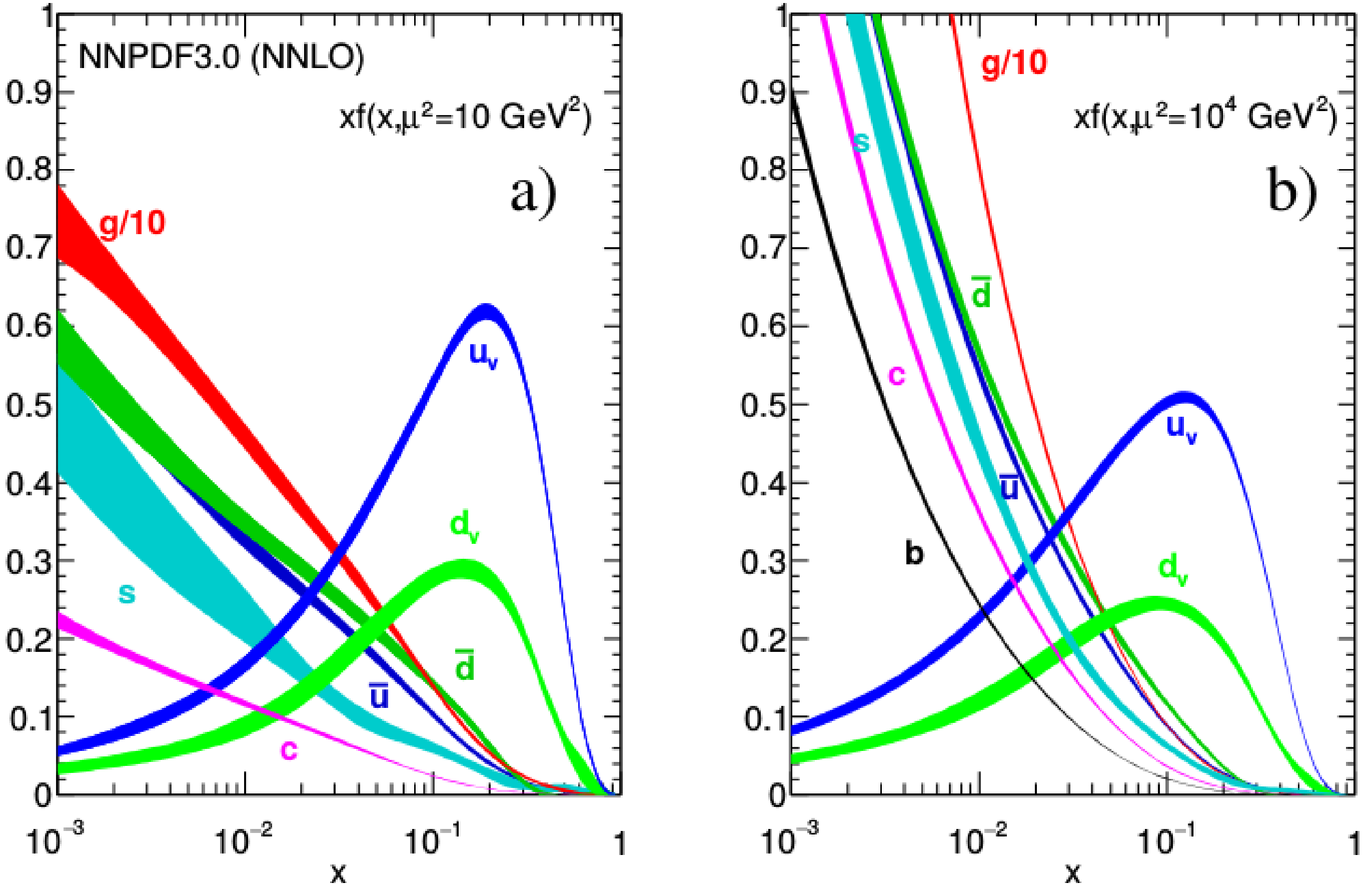
EXAMPLE: THE DRELL-YAN PROCESS

(LEADING ORDER)



- HADRONIC C.M. ENERGY: $s = (p_1 + p_2)^2$
- PARTONIC C.M. ENERGY: $\hat{s} = x_1 x_2 s$
- MOMENTUM FRACTIONS $x_{1,2} = \sqrt{\frac{\hat{s}}{s}} \exp \pm y$; AT LEADING ORDER $\hat{s} = M^2$

THE PDFs

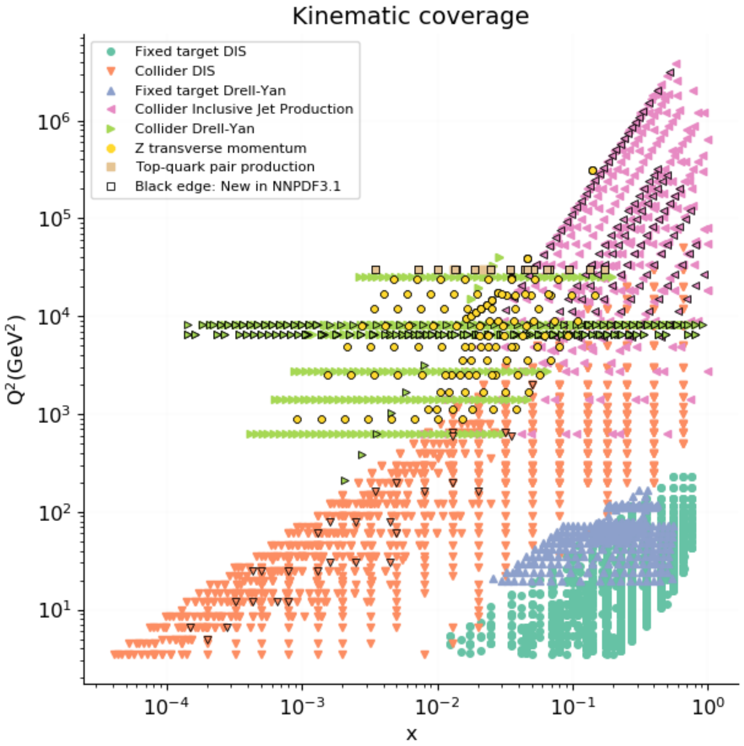


(PDG 2016)

- MOMENTUM PROBABILITY DENSITY $xf_i(x)$ AT TWO DIFFERENT SCALES (LEFT $\Rightarrow$ LOW SCALE; RIGHT $\Rightarrow$ HIGH SCALE)
- AS $x \geq 1$ KINEMATIC CONSTRAINT $f_i(x) = 0$
- VALENCE SUM RULES $\int dx(u(x) - \bar{u}(x)) = 2 \int dx(d(x) - \bar{d}(x)) = 2.$
- MOMENTUM SUM RULE $\sum \int dx x f_i(x) = 1$

PDF DETERMINATION

DATA $\rightarrow$ PARTON DISTRIBUTIONS



ISSUES:

TRIVIAL

- FROM PHYSICAL OBSERVABLES TO PDFs: SOLVE EVOLUTION EQUATIONS, CONVOLUTE WITH PARTON-LEVEL CROSS-SECTIONS
- DISENTANGLING PDFs: CHOOSE A BASIS OF PDFs ($2N_f$ QUARKS + 1 GLUON) & A SET OF SUITABLE PHYSICAL PROCESSES TO DETERMINE THEM ALL

NONTRIVIAL

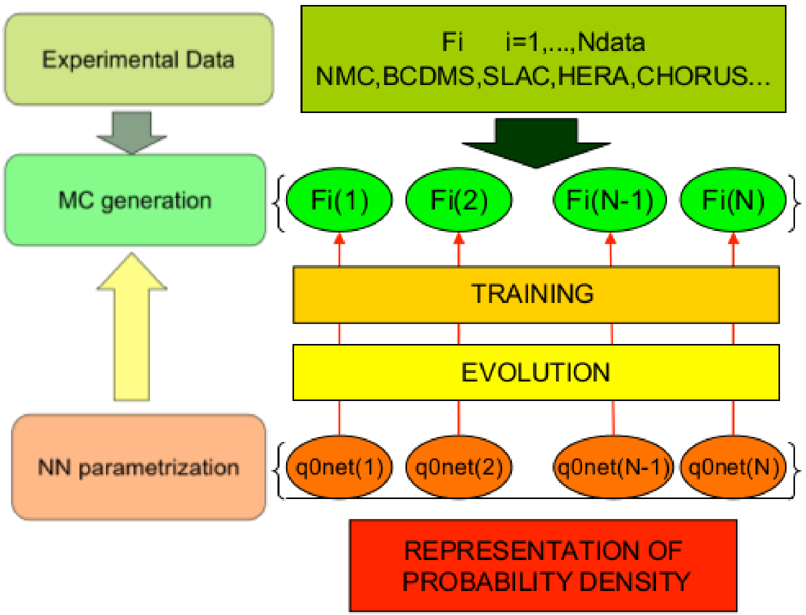
- (1) DETERMINE FUNCTIONS FROM A DISCRETE DATASET
- (2) DETERMINE A PROBABILITY FUNCTIONAL IN THE SPACE OF FUNCTIONS

THE NNPDF APPROACH

BASIC IDEA: MONTE CARLO SAMPLING OF THE PROBABILITY MEASURE IN THE (FUNCTION) SPACE OF PDFs

- GENERATE A SET OF MONTE CARLO REPLICAS $\sigma^{(k)}$ OF THE ORIGINAL DATASET $\sigma^{(\mathrm{data})}$
  $\Rightarrow$ REPRESENTATION OF $\mathcal{P}[\sigma]$ AT DISCRETE SET OF POINTS IN DATA SPACE
- FIT A PDF REPLICA TO A DATA REPLICA
  $\Rightarrow$ EACH PDF REPLICA $f_i^{(k)}$ IS A BEST-FIT PDF SET FOR GIVEN DATA REPLICA
- THE SET OF NEURAL NETS IS A REPRESENTATION OF THE PROBABILITY DENSITY:

$$\langle f_i \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} f_i^{(k)}$$



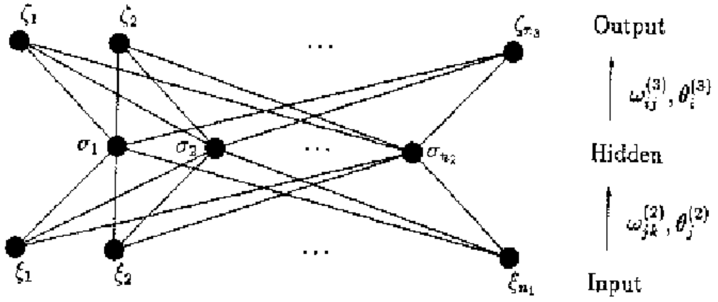
SOLUTIONS

- (1) FUNCTIONS FROM DISCRETE DATA $\Rightarrow$ NEURAL NETWORKS
- (2) PROBABILITY IN FUNCTION SPACE $\Rightarrow$ MONTE CARLO

NEURAL NETWORKS

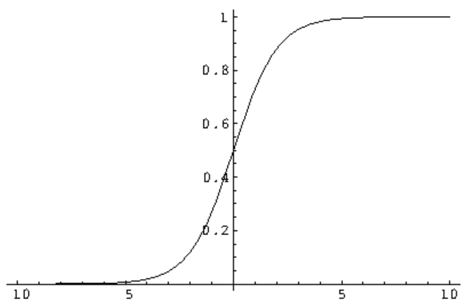
- EACH PDF REPLICA FITTED TO A DATA REPLICA
  $\Rightarrow$ NEED BEST-FIT, COVARIANCE MATRIX IN PARAMETER SPACE NOT NEEDED
- CAN USE VERY LARGE PARAMETRIZATION

NEURAL NETWORKS



MULTILAYER FEED-FORWARD NETWORKS

- Each neuron receives input from neurons in preceding layer and feeds output to neurons in subsequent layer
- Activation determined by weights and thresholds
  $$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right)$$
- Sigmoid activation function
  $g(x) = \frac{1}{1+e^{-\beta x}}$



EXAMPLE: A 1-2-1 NN

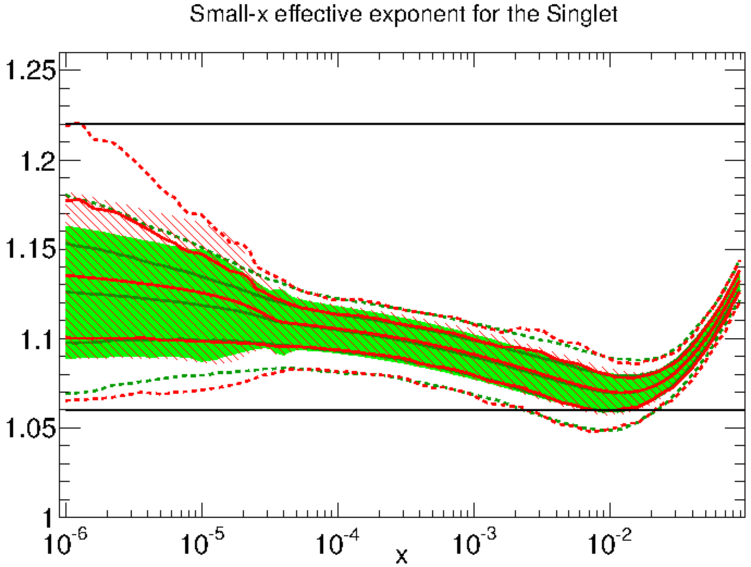
$$f(x) = \frac{1}{1+e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - x\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - x\omega_{21}^{(1)}}}}}$$

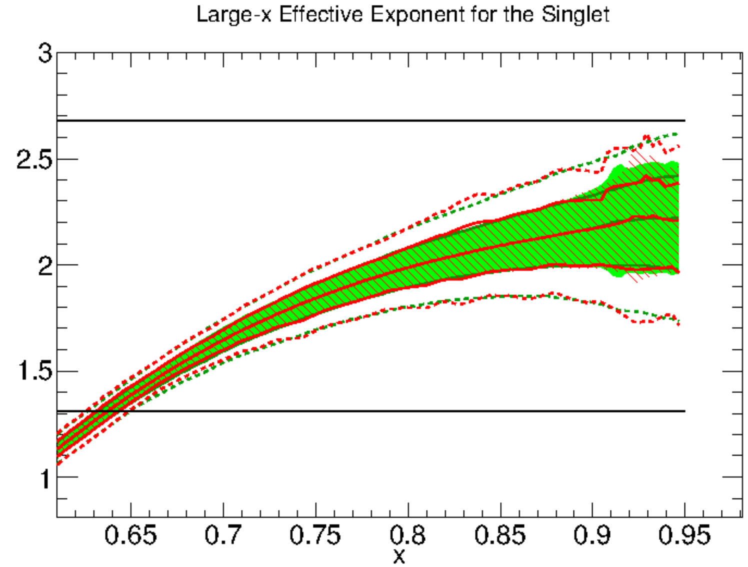
CURRENTLY: 2-5-3-1 NN FOR EACH OF 8 BASIS PDFS (37X8=296 FREE PARMS.)

PREPROCESSING

- PDFs ARE PARAMETRIZED WITH NEURAL NETWORKS TIMES PREPROCESSING FUNCTION: $f_i(x) = x^{\alpha_i}(1-x)^{\beta_i} NN(x)$
- GOAL IS TO SPEED UP TRAINING WITHOUT BIASING RESULT
- α_i, β_i RANDOM REPLICA BY REPLICA WITH UNIFORM DISTRIBUTION IN RANGE
- RANGE DETERMINED SELF-CONSISTENTLY AS TWICE THE RANGE OF EFFECTIVE EXPONENTS $\alpha_{\rm eff,i} = \frac{\ln f_i(x)}{\ln 1/x}$ $\beta_{\rm eff,i} = \frac{\ln f_i(x)}{\ln(1-x)}$
  EVALUATED AT $x = 0.95,\ 0.65$ (β); $x = 10^{-6},\ 10^{-3}$ (α)

EFFECTIVE EXPONENTS FOR QUARK SINGLET VS. PREPROCESSING RANGE
100 & 1000 REPLICAS





GENETIC MINIMIZATION

RANDOM MUTATION OF THE NN PARAMETERS STARTING FROM RANDOM VALUES

- LARGE NUMBER OF MUTANT (~ 100) PDF SETS GENERATED FROM PARENT
- FIGURE OF MERIT COMPUTED
- BEST-FIT KEPT & PASSED TO NEXT GENERATION

$$w \to w + \frac{\eta r_\delta}{N_{\rm ite}^{r_{\rm ite}}}$$

CHOICES

- MUTATION RATE η
- POINTLIKE VS. NODAL MUTATION
- NUMBER (POINTLIKE) OR PROBABILITY (NODAL) OF MUTATIONS
- TARGETED WT: WEIGTHS $p_i = E_i/E_i^{\rm targ}$
- GA EPOCHS: $N_{\rm gen}^{\rm mut}$

	$N_{\rm gen}^{\rm wt}$	$N_{\rm gen}^{\rm mut}$	$N_{\rm gen}^{\rm max}$	$E^{\rm sw}$	$N_{\rm mut}^{a}$	$N_{\rm mut}^{b}$
NNPDF 2.3	10000	2500	50000	2.3	80	30
NNPDF 3.0	–	-	30000	-	80	-

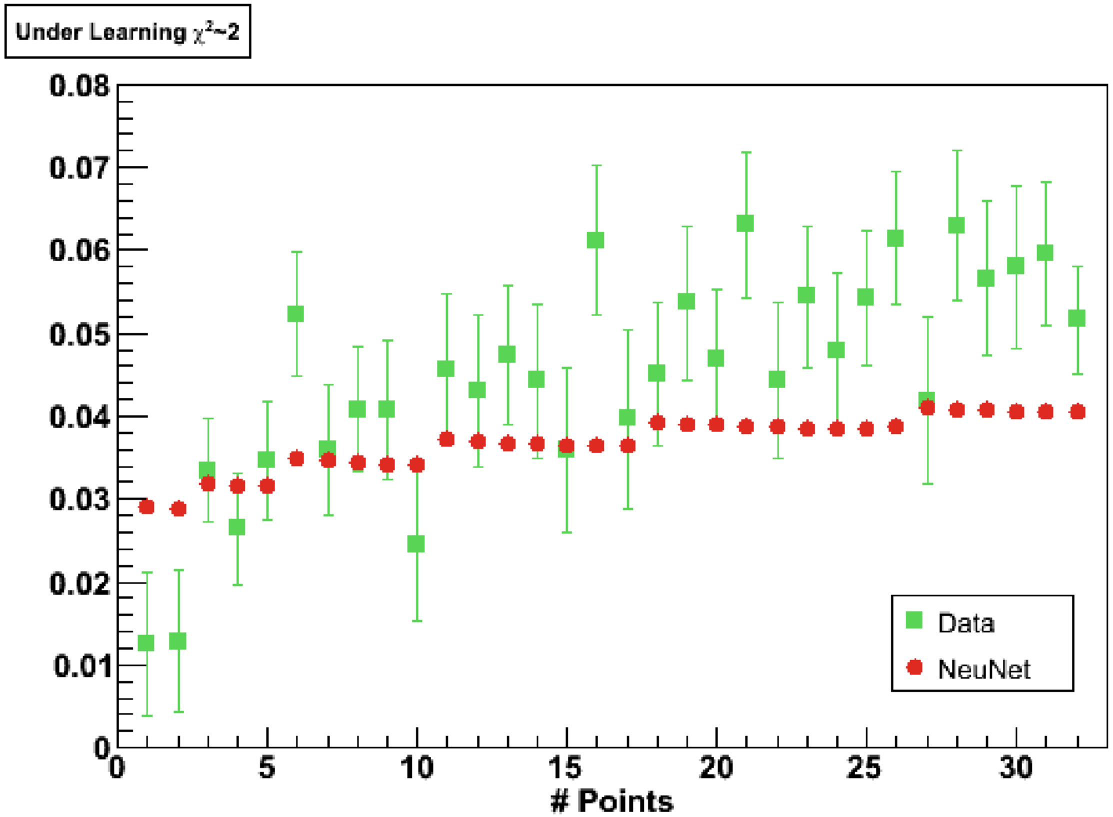
NNPDF2.3		
Single Parameter Mutation		
PDF	$N_{\rm mut}$	η
$\Sigma(x)$	2	10, 1
$g(x)$	3	10, 3, 0.4
$T_3(x)$	2	1, 0.1
$V(x)$	3	8, 1, 0.1
$\Delta_S(x)$	3	5, 1, 0.1
$s^+(x)$	2	5, 0.5
$s^-(x)$	2	1, 0.1

NNPDF3.0		
Nodal Mutation		
PDF	$P_{\rm mut}$	η
$\Sigma(x)$	5% per node	15
$g(x)$	5% per node	15
$V(x)$	5% per node	15
$V_3(x)$	5% per node	15
$V_8(x)$	5% per node	15
$T_3(x)$	5% per node	15
$T_8(x)$	5% per node	15

NN TRAINING: EXAMPLE

- HIGHLY REDUNDANT PARAMETRIZATION
- COMPLEXITY INCREASES AS THE FITTING PROCEEDS
- $\Rightarrow$ THE BEST FIT IS NOT THE ABSOLUTE MINIMUM:
  MUST LOOK FOR OPTIMAL LEARNING POINT

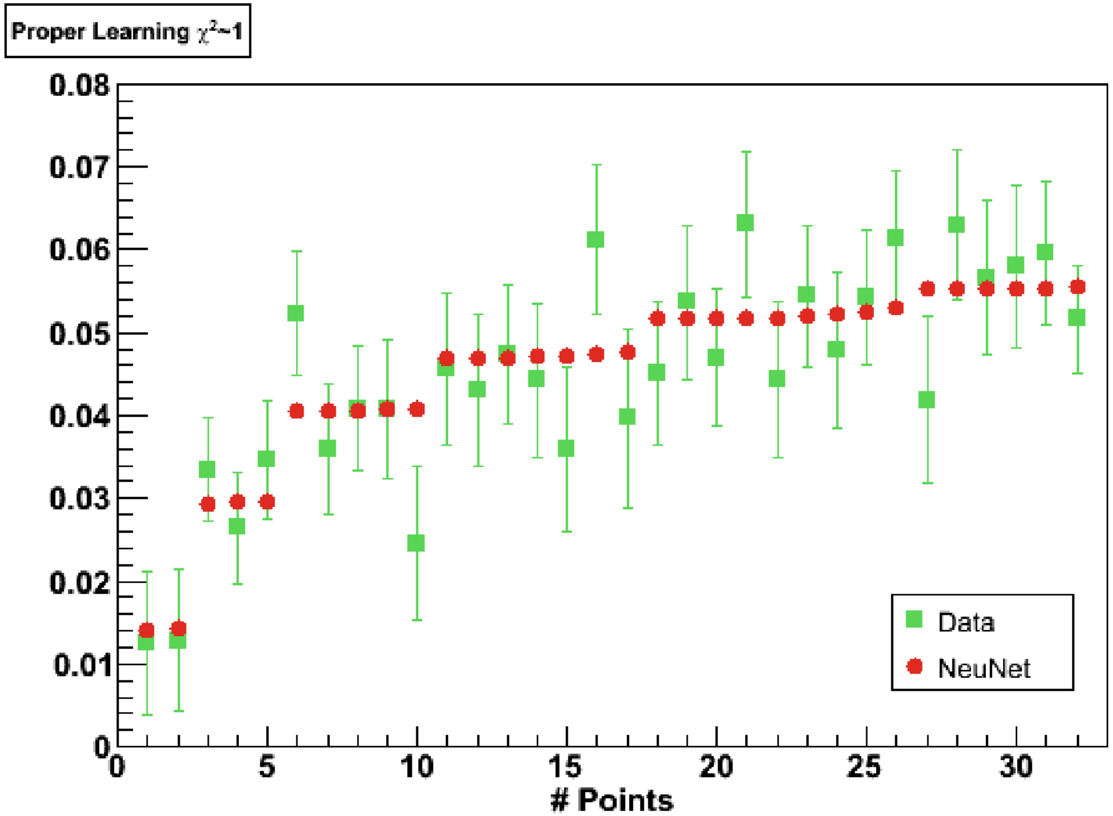
UNDERLEARNING



NN TRAINING: EXAMPLE

- HIGHLY REDUNDANT PARAMETRIZATION
- COMPLEXITY INCREASES AS THE FITTING PROCEEDS
- $\Rightarrow$ THE BEST FIT IS NOT THE ABSOLUTE MINIMUM:
  MUST LOOK FOR OPTIMAL LEARNING POINT

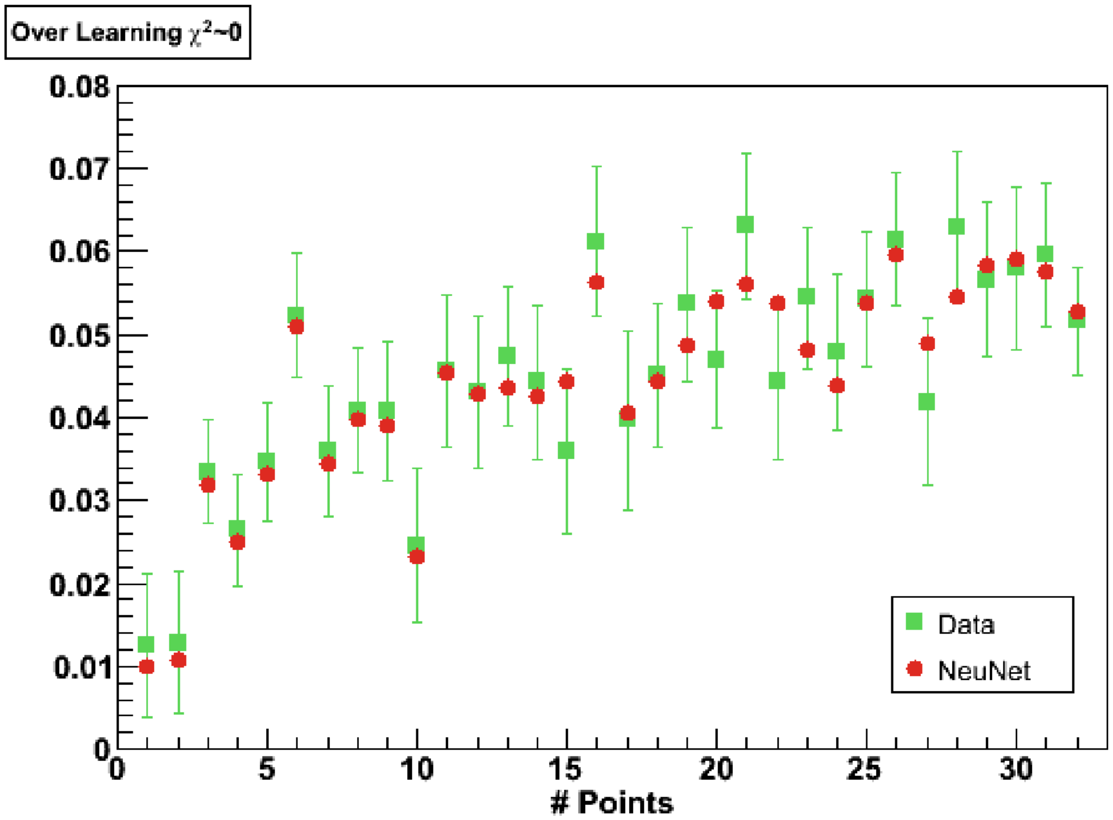
PROPER LEARNING



NN TRAINING: EXAMPLE

- HIGHLY REDUNDANT PARAMETRIZATION
- COMPLEXITY INCREASES AS THE FITTING PROCEEDS
- $\Rightarrow$ THE BEST FIT IS NOT THE ABSOLUTE MINIMUM:
  MUST LOOK FOR OPTIMAL LEARNING POINT

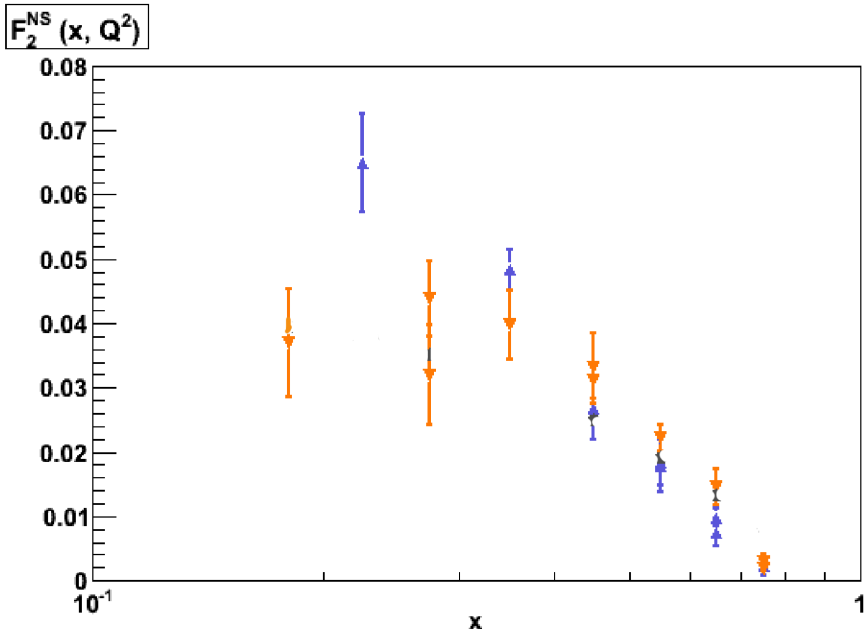
OVERLEARNING



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

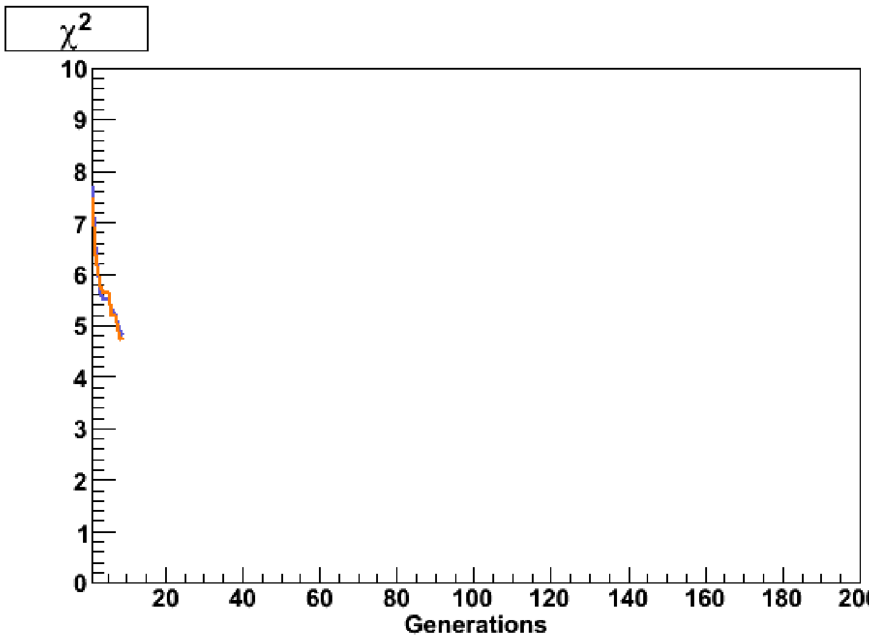


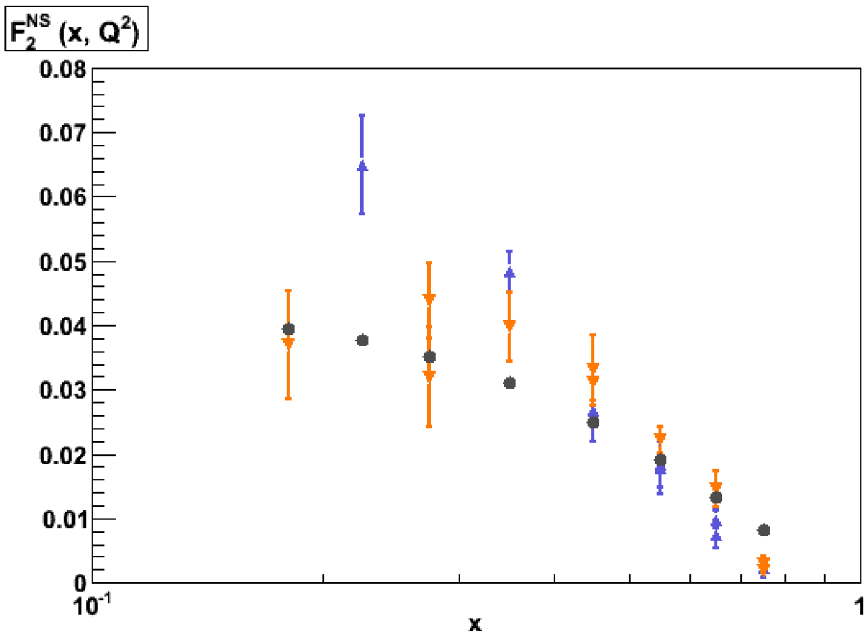
OPTIMAL FIT: CROSS-VALIDATION

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!



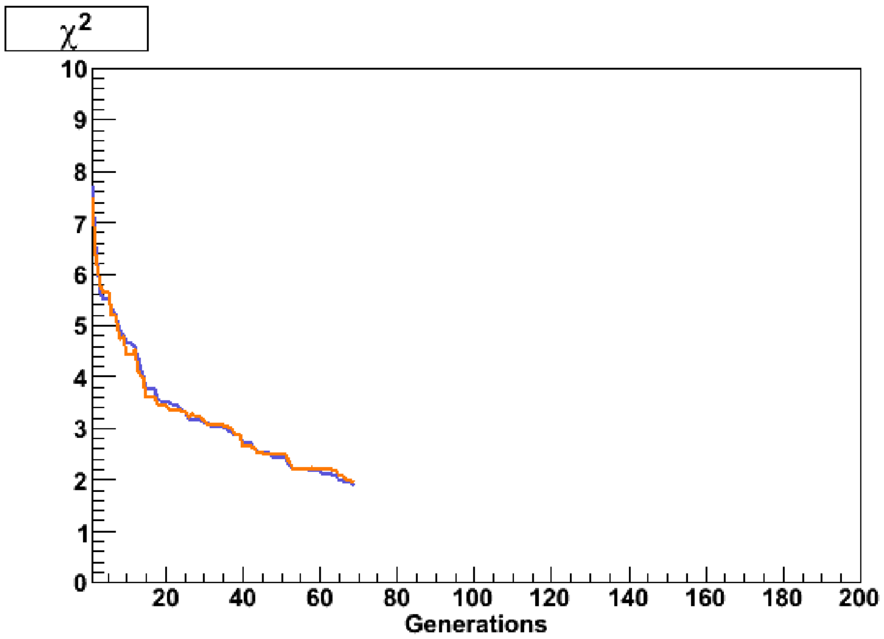


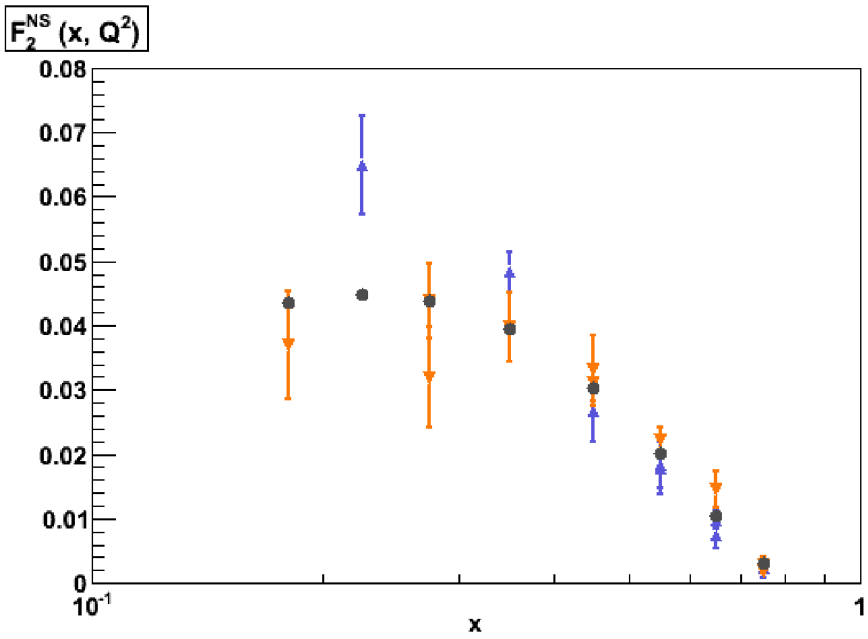
OPTIMAL FIT: CROSS-VALIDATION

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!



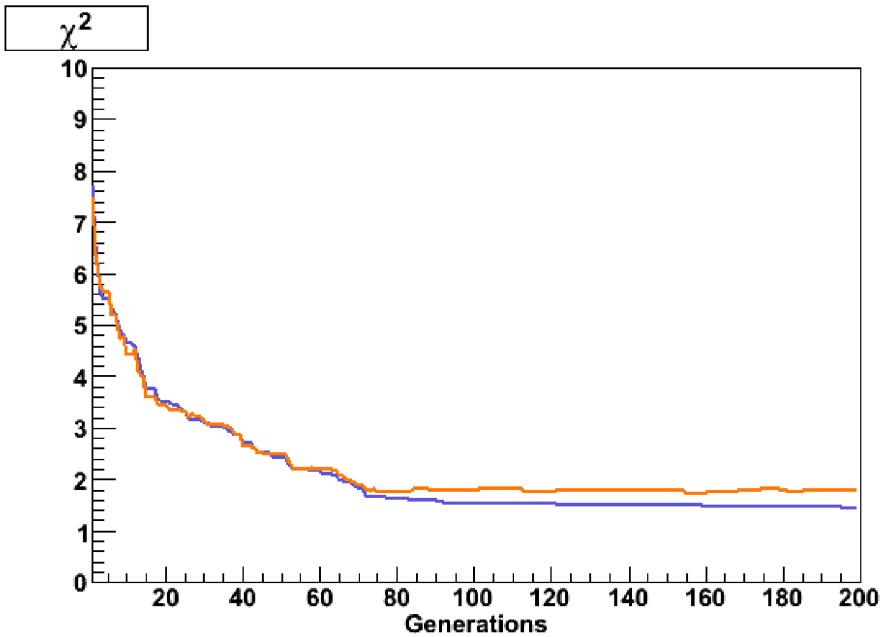


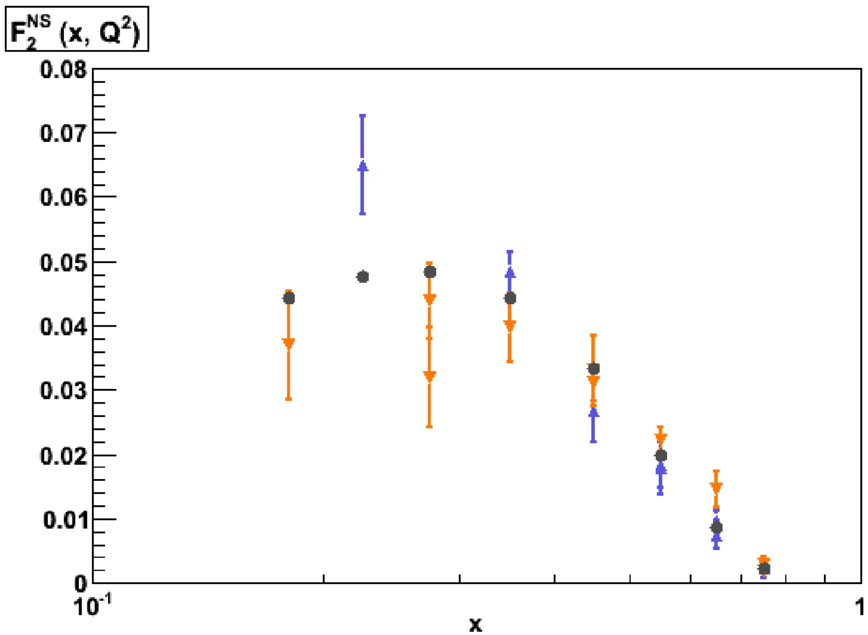
OPTIMAL FIT: CROSS-VALIDATION

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

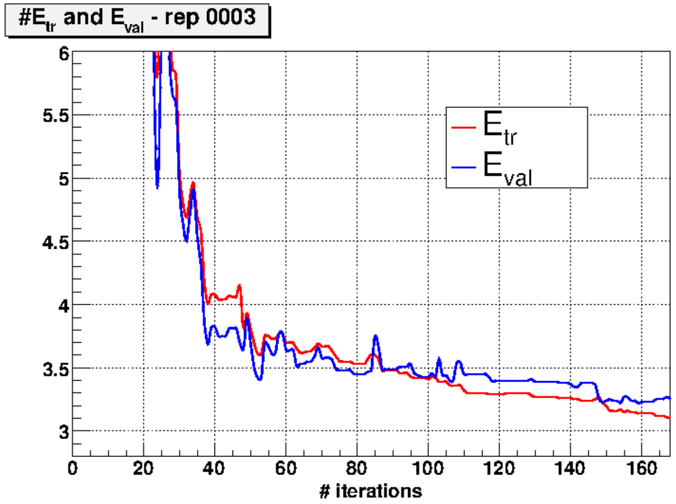
TOO LATE!

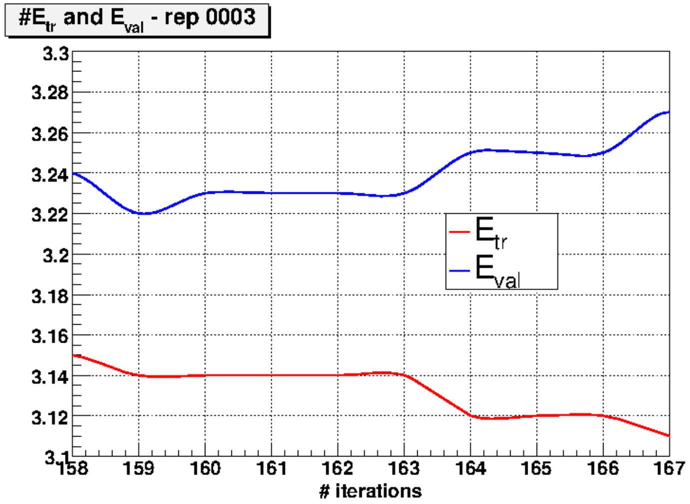




STOPPING CRITERIA

STOPPING FOR THE χ^2 OF ONE REPLICA





- UP TO NNPDF2.3 "INCREASING" AND "DECREASING" TRAINING AND VALIDATION χ^2 DEFINED IN TERMS OF THRESHOLD VALUES δ_{tr} AND δ_{val}:
  INCREASE: $\chi^2_{val}(N_{gen}+\Delta) > \chi^2(N_{gen}+\Delta) + \delta_{val}$
- FROM NNPDF3.0 USE LOOKBACK:
  FIT IS RUN UP TO SOME LARGE # OF GA GENERATIONS
  THEN ONE "LOOKS BACK" FOR ABSOLUTE MIN. OF VALIDATION χ^2
- CHECK THAT RESULTS ARE INDEPENDENT OF THE LARGE # OF GA GENS
- CHECK THAT RESULTS ARE INDEPENDENT OF FLUCTUATIONS IN VALUE OF ABSOLUTE MINIMUM
  DIFFERENT STOPPING POINTS, BUT INDISTINGUISHABLE PDFs

CLOSURE TESTS:

THE BASIC IDEA

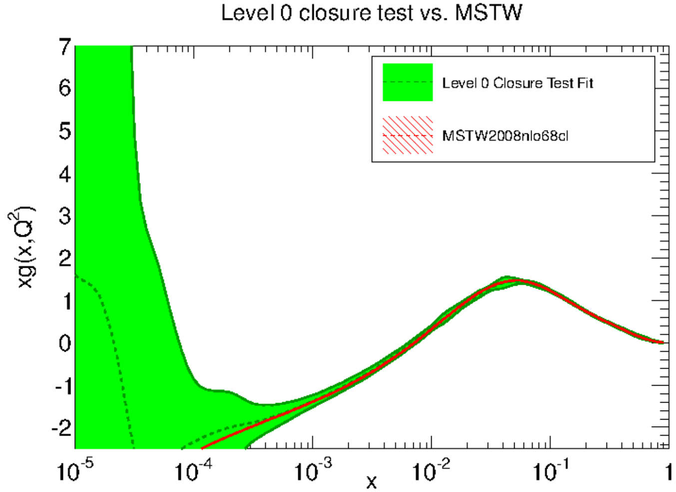
- ASSUME PDFs KNOWN: GENERATE FAKE EXPERIMENTAL DATA
- CAN DECIDE DATA UNCERTAINTY (ZERO, OR AS IN REAL DATA, OR . . .)
- FIT PDFs TO FAKE DATA
- CHECK WHETHER FIT REPRODUCES UNDERLYING "TRUTH":
  - CHECK WHETHER TRUE VALUE GAUSSIANLY DISTRIBUTED ABOUT FIT
  - CHECK WHETHER UNCERTAINTIES FAITHFUL
  - CHECK STABILITY

    (INDEP. OF METHODOLOGICAL DETAILS)

LEVEL-0 CLOSURE TESTS

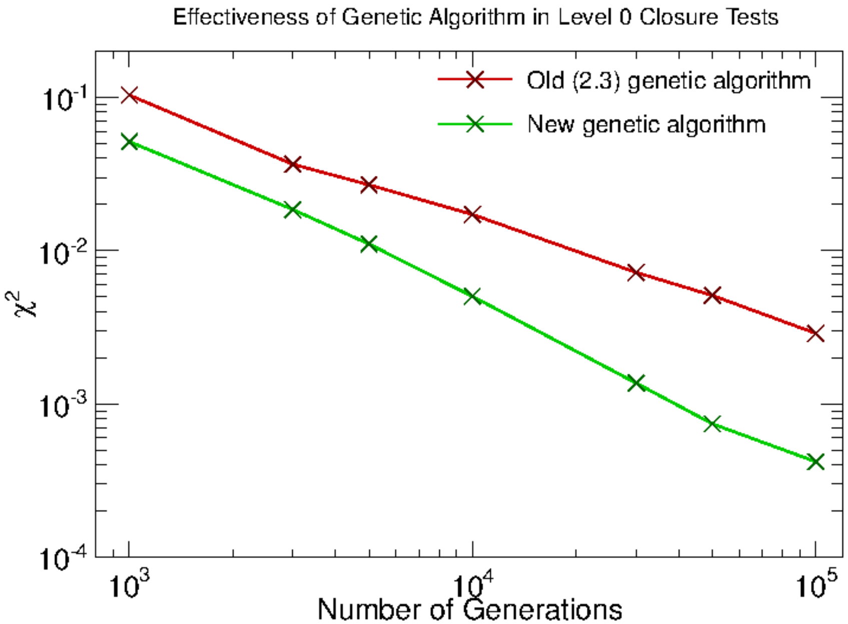
- ASSUME VANISHING EXPERIMENTAL UNCERTAINTY
- MUST BE ABLE TO GET $\chi^2 = 0$
- UNCERTAINTY AT DATA POINTS TENDS TO ZERO (NOT NECESSARILY ON PDF!)
  DEFINE $\phi \equiv \sqrt{\langle \chi^2_{rep} \rangle - \chi^2}$, EQUALS FIT UNCERTAINTY/DATA UNCERTAINTY; CHECK $\phi \to 0$
- BEST FIT ON TOP OF "TRUTH" IN DATA REGION

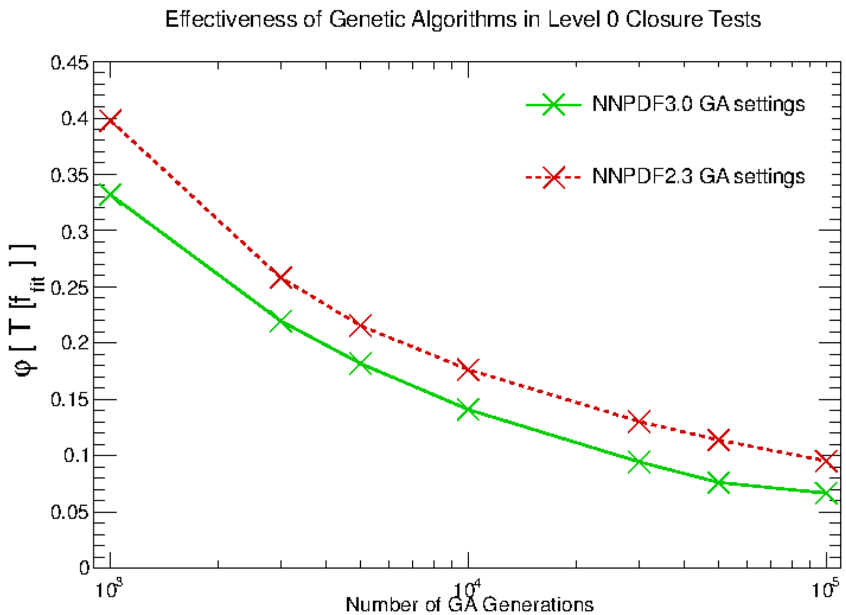
THE GLUON



FRACTIONAL UNCERTAINTY VS TRAINING LENGTH

χ^2 VS TRAINING LENGTH

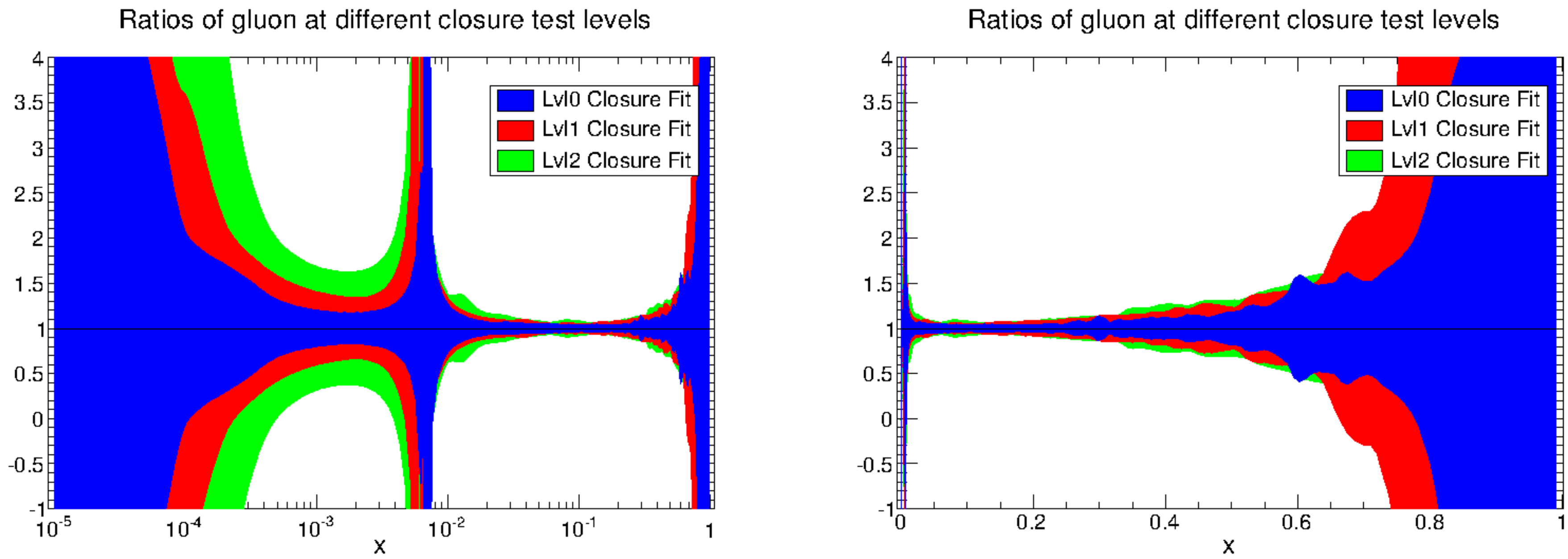




LEVEL-0, LEVEL-1 AND LEVEL-2

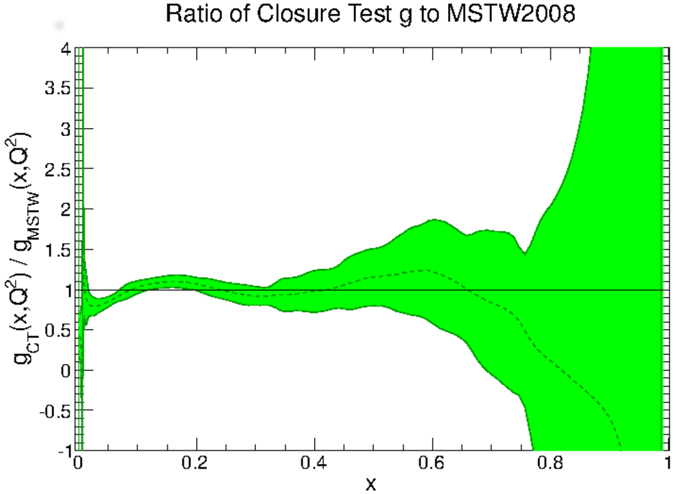
- LEVEL 0: FAKE DATA GENERATED WITH NO UNCERTAINTY
  $\rightarrow$ INTERPOLATION AND EXTRAPOLATION UNCERTAINTY
- LEVEL 1-2: FAKE DATA GENERATED WITH SAME UNCERTAINTY AS REAL DATA (INCLUDING CORRELATIONS)
- LEVEL 1: NO PSEUDODATA REPLICAS:
  $\Rightarrow$ REPLICAS FITTED TO SAME DATA OVER AND OVER AGAIN
  $\rightarrow$ FUNCTIONAL UNCERTAINTY DUE TO INFINITY OF EQUIVALENT MINIMA
- LEVEL 2: STANDARD NNPDF METHODOLOGY
  $\Rightarrow$ REPLICAS FITTED TO PSEUDODATA REPLICAS
  $\rightarrow$ DATA UNCERTAINTY
- THREE SOURCES OF UNCERTAINTY COMPARABLE IN DATA REGION

THE GLUON: LEVEL 0, LEVEL 1 AND LEVEL 2



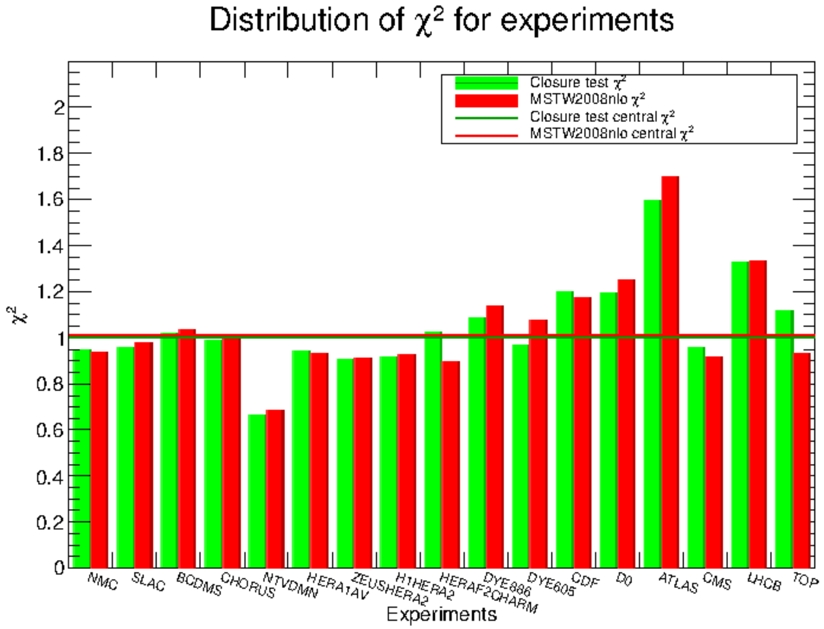
LEVEL-2: CENTRAL VALUES AND UNCERTAINTIES

THE GLUON: FITTED/"TRUE"

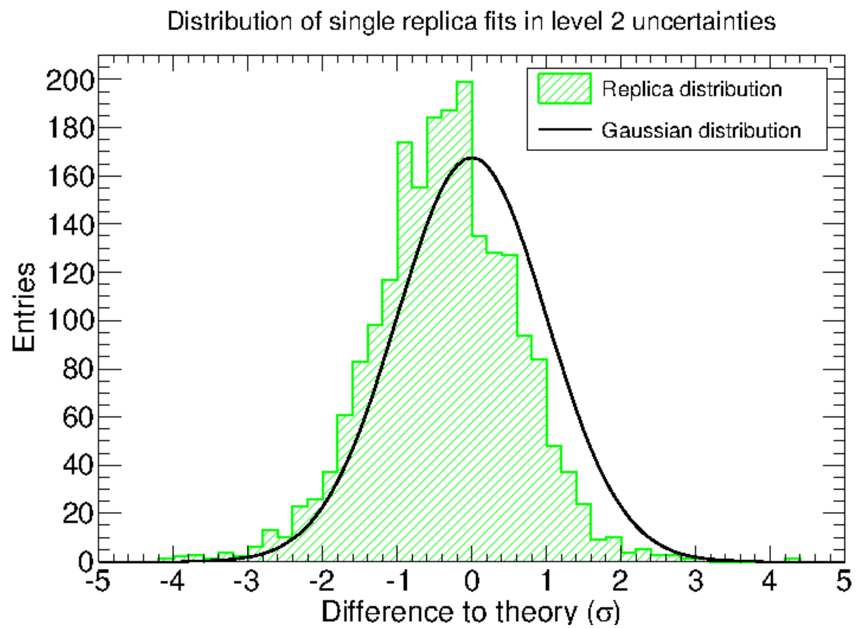


- CENTRAL VALUES:
  COMPARE FITTED VS. "TRUE" χ^2
  BOTH FOR INDIVIDUAL EXPERIMENTS
  & TOTAL DATASET
  FOR TOTAL $\Delta\chi^2 = 0.001 \pm 0.003$
- UNCERTAINTIES: DISTRIBUTION OF DEVIATIONS BETWEEN FITTED AND "TRUE" PDFs
  SAMPLED AT 20 POINTS BETWEEN 10^{-5} AND 1
  FIND 0.699% FOR ONE-SIGMA,
  0.948% FOR TWO-SIGMA C.L.

LEVEL-2 FITTED χ^2 VS "TRUE"



NORM. DISTRIBUTION OF DEVIATIONS



LEVEL-2 STABILITY TESTS

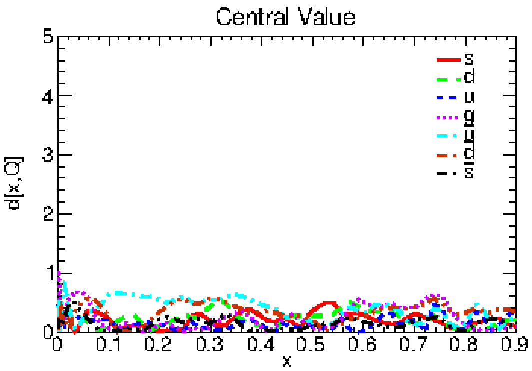
- INCREASE MAXIMUM GA TRAINING LENGTH TO 80K
  TESTS EFFICIENCY OF CROSS-VALIDATION
- INCREASE NN ARCHITECTURE TO 2-20-15-1
  NUMBER OF FREE PARAMETRES INCREASE BY MORE THAT $10\times$
- CHANGE PDF PARAMETRIZATION BASIS
  OLD: ISOTRIPLET, $\bar{u} - \bar{d}$, $s + \bar{s}$, $s - \bar{s}$;
  NEW: ISOTRIPLET, SU(3)-OCTET, BOTH TOTAL $(q + \bar{q})$ & VALENCE $(q - \bar{q})$

STATISTICAL EQUIVALENCE!

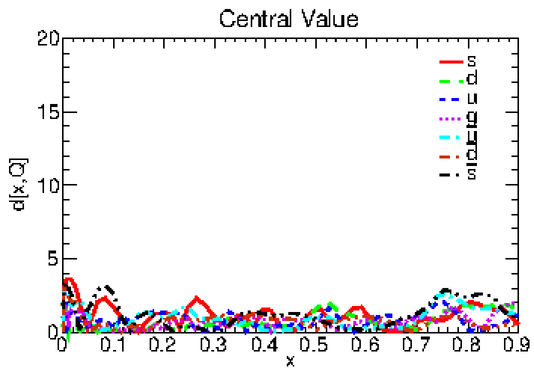
DISTANCES BETWEEN REF. AND NEW FIT:

difference in units of standard deviation of the mean

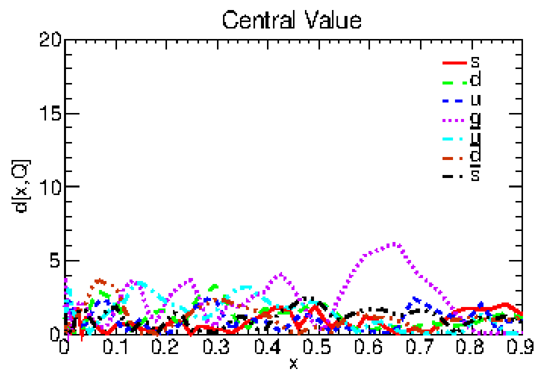
30K GENS VS 80K GENS



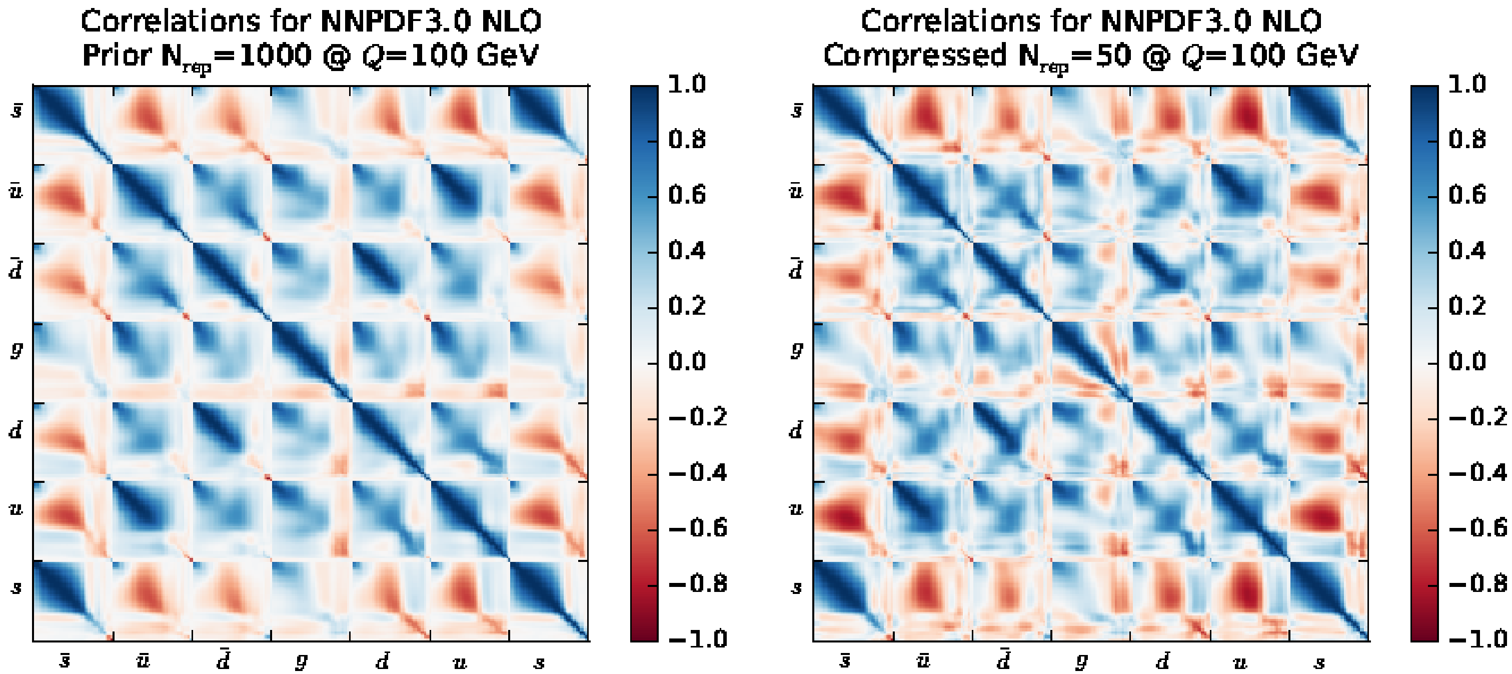
2.3 BASIS VS 3.0 BASIS



300 VS 37 PARMS



OPTIMIZATION I
MONTECARLO COMPRESSION

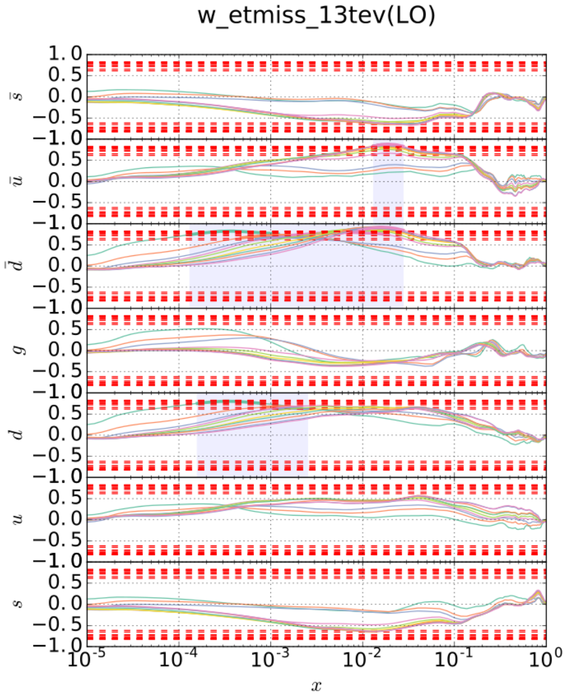


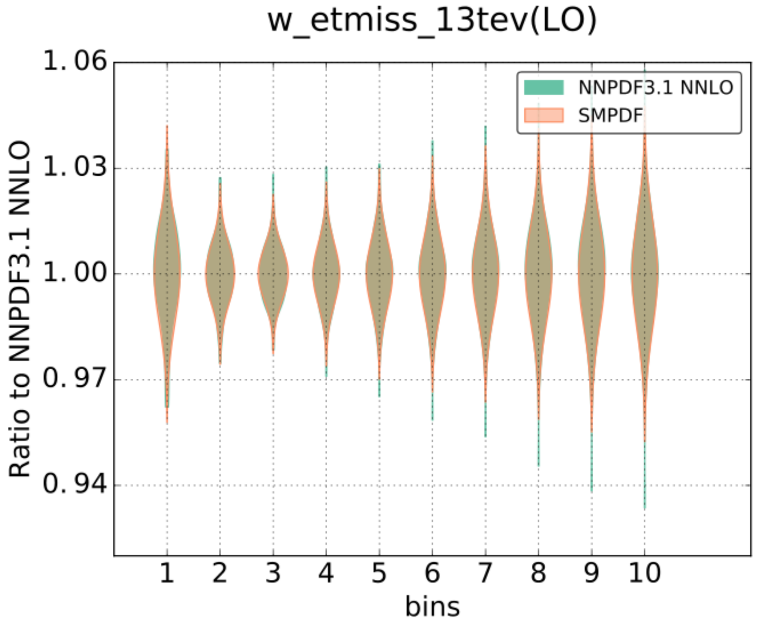
(Carrazza, Latorre, Kassabov, Rojo, 2015)

- 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

OPTIMIZATION II
SMPDF COMPRESSION

- 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
- COMPRESSED HESSIAN REPRESENTATION OF PROBABILITY DISTN.
- 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!

ALL IS WELL?

CAN WE DO BETTER?

- ARCHITECTURE: DO WE NEED SEVEN NNs?
- PREPROCESSING: ARE RESULTS TRULY INDEPENDENT OF IT?
- MINIMIZATION: IS THE GA OPTIMIZED?
- STOPPING: OVER/UNDERLEARNING?

UNCERTAINTIES ARE FAITHFUL, BUT

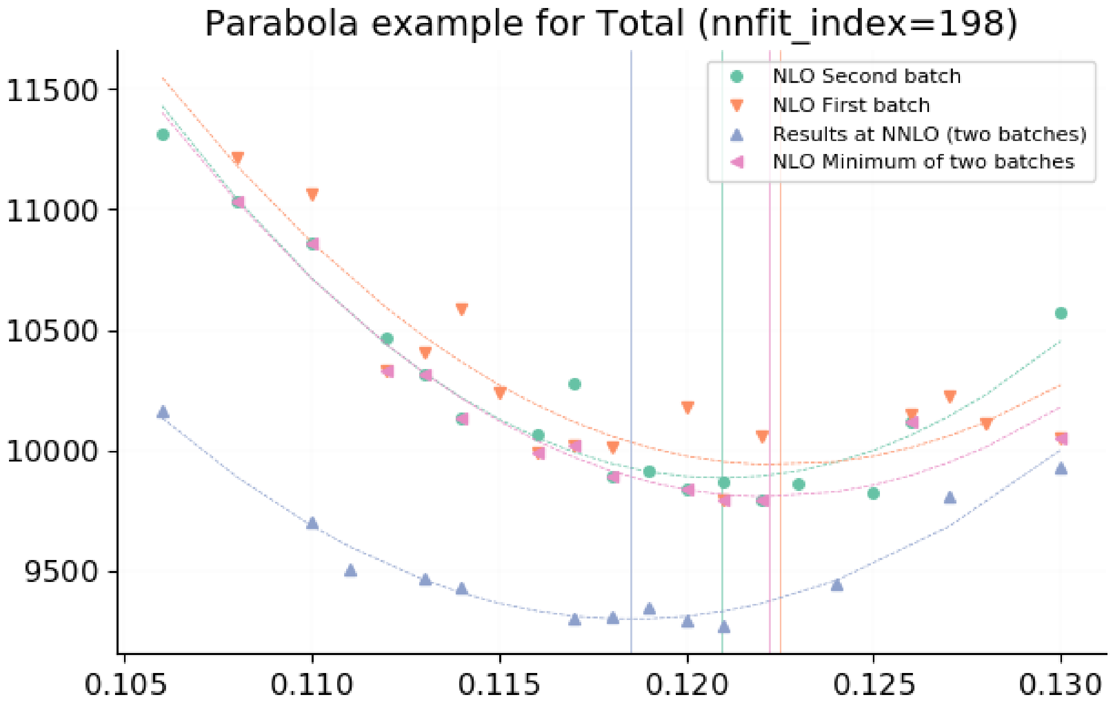
ARE THEY THE SMALLEST WITH GIVEN DATA?

IS THERE NO INFORMATION LOSS?

EXTRAS

MORE EFFICIENT MINIMIZATION?

- LOOK AT α_s DEPENDENCE (CORRELATED REPLICAS)
- SIGNIFICANT FLUCTUATIONS ABOUT PARABOLIC SHAPE
  NOT DUE TO FINITE-SIZE MONTE CARLO SAMPLE



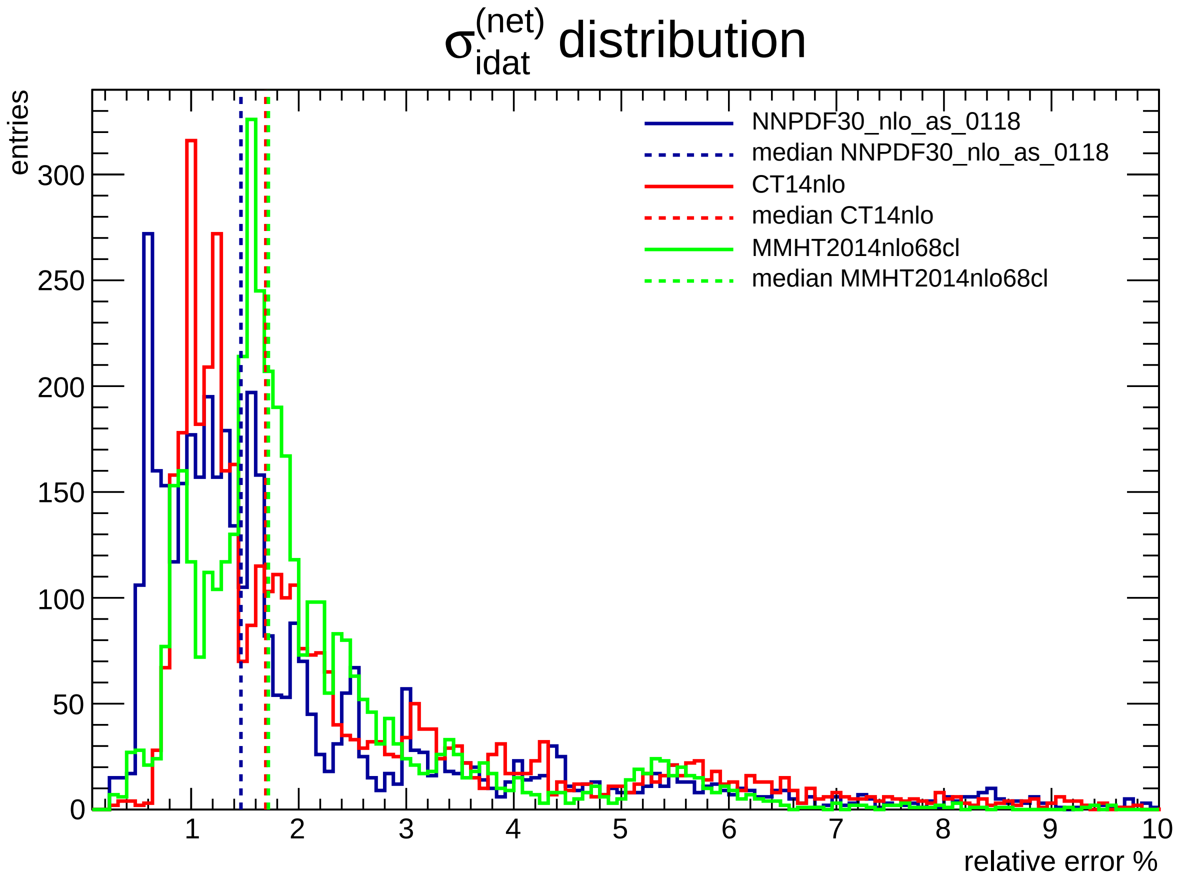
BATCH MINIMIZATION

- MINIMIZE EACH REPLICA MORE THEN ONCE & KEEP BEST RESULTS
- SIGNIFICANT STABILIZATION

PDF UNCERTAINTIES: HOW MUCH DO THEY VARY?

- COMPUTE PERCENTAGE PDF UNCERTAINTY ON ALL DATA INCLUDED IN GLOBAL FIT
- COMPARE GLOBAL FITS

PERCENTAGE PDF UNCERTAINTY ON PREDICTIONS



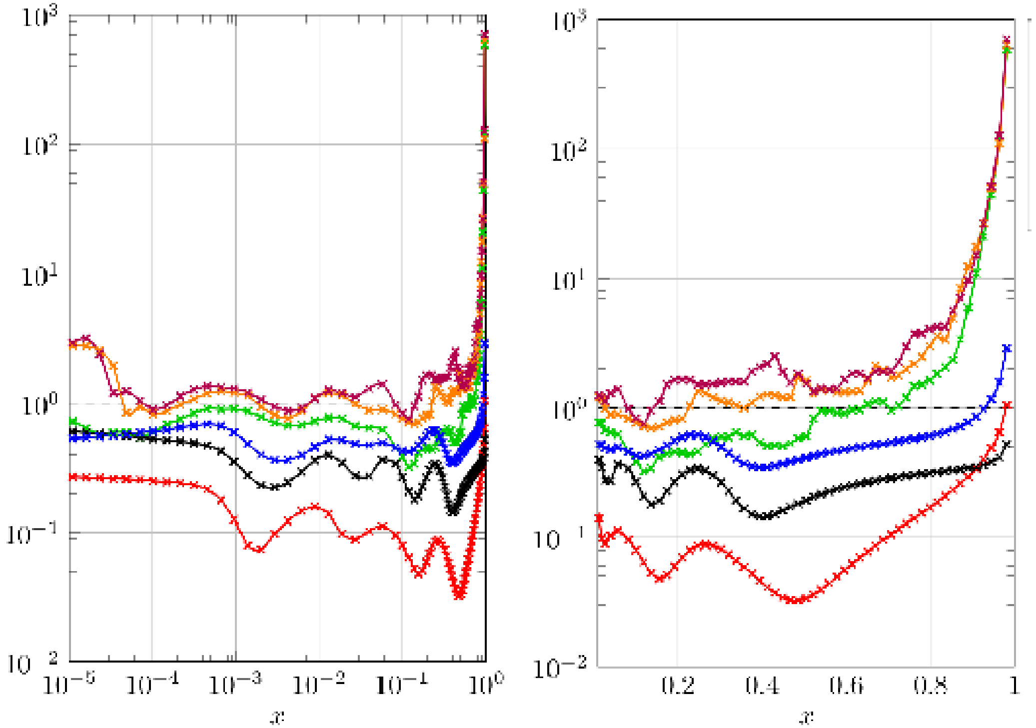
- MEDIAN SIMILAR
- DISTRIBUTION VERY DIFFERENT!
- NNPDF: SMALLER MODE, BUT FAT TAIL ⇔ GREATER FLEXIBILITY

THE $\Delta\chi^2$ PROBLEM

- TOLERANCE MIGHT COMPENSATE FOR MISSING FUNCTIONAL UNCERTAINTY
- BUT WHAT IS $\Delta\chi^2$ FOR AN NNPDF FIT?
- CAN ANSWER USING HESSIAN CONVERSION! $\Delta\chi^2 = 16 \pm 15$
  - NON-PARABOLIC BEHAVIOUR NEAR MINIMUM ON SCALE OF UNCERTAINTIES?
  - INEFFICIENCY OF THE MINIMIZATION PROCEDURE?

CLOSURE-TESTING: THE PARAMETRIZATION DEPENDENCE

GLUON PDF UNCERTAINTY NORMALIZED TO MSTW08



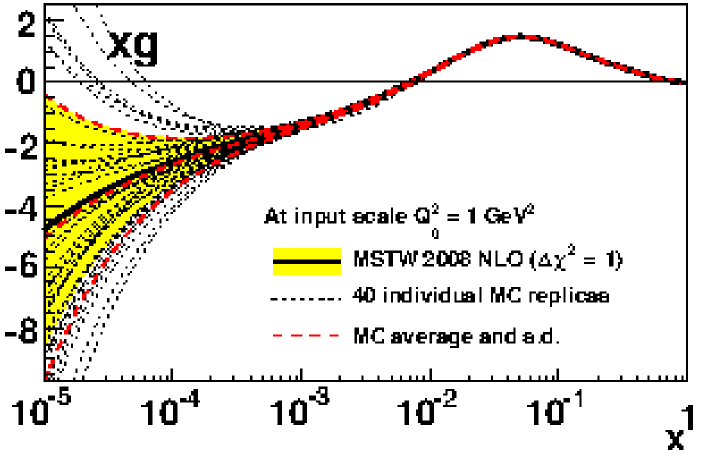
(C. Mascaretti, 2016)

- CLOSURE TEST PERFORMED WITH DATA GENERATED BASED ON MST08 FUNCTIONAL FORM
- REFITTED EITHER WITH NNPDF OR MSTW-CT FUNCTIONAL FORM
- LEVEL 0: VANISHING DATA UNCERTAINTY
  - MSTW-CT: FIT HAS ZERO UNCERTAINTY
  - NNPDF: ABOUT HALF OF TOTAL UNCERTAINTY
- LEVEL 1: NOMINAL DATA UNCERTAINTY, BUT REPLICAS FITTED W/O PSEUDODATA
  - MSTW-CT: FIT HAS SMALL UNCERTAINTY
  - NNPDF: ABOUT 2/3 OF FINAL UNCERTAINTY
- LEVEL 2
  - NNPDF UNCERTAINTY LARGER THAN MSTW-CT
  - NNPDF UNCERTAINTY SIMILAR TO MSTW WITH TOLERANCE

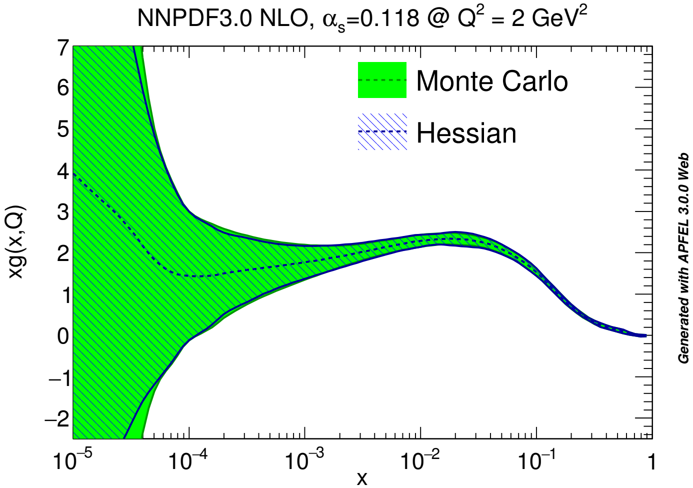
"STANDARD" PARAMETRIZATION MISSES INTERPOLATION & FUNCTIONAL UNCERTAINTY?

MC ⇔ HESSIAN

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



(Thorne, Watt, 2012)



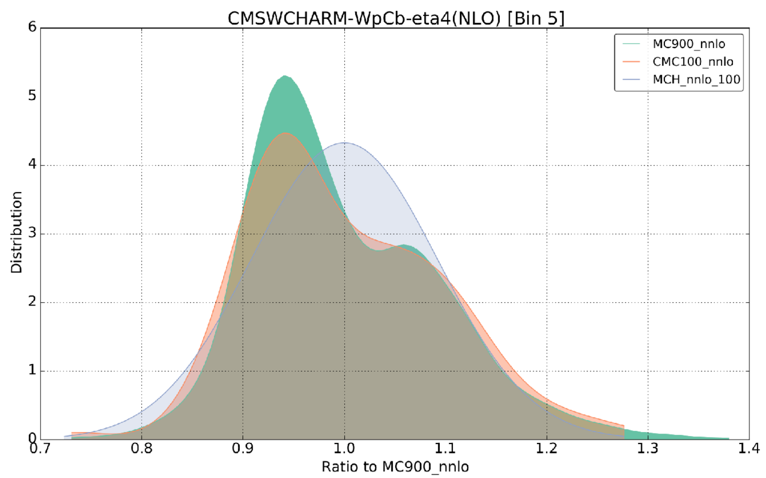
(Carrazza, SF, Kassabov, Rojo, 2015)

- 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 ⇒ ACCURATE REPRESENTATION

NONGAUSSIAN BEHAVIOUR

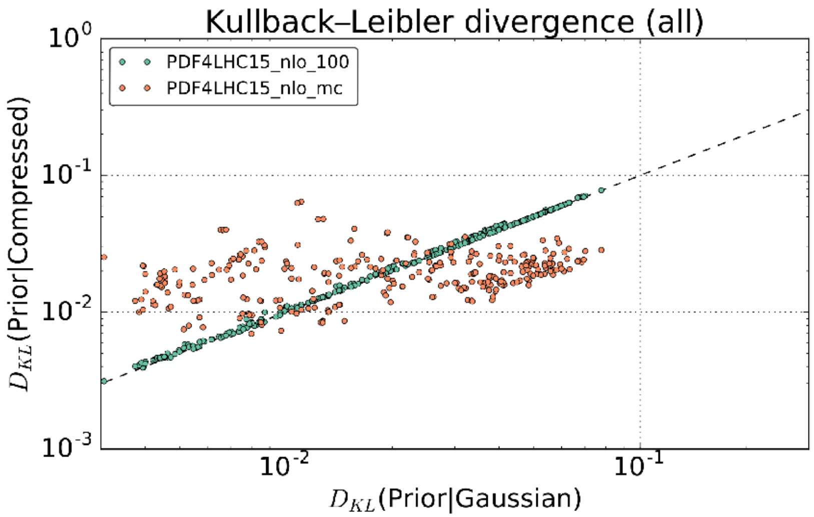
MONTE CARLO COMPARED TO HESSIAN

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



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