ESTIMATING UNCERTAINTIES:
MONTE CARLO vs HESSIAN

STEFANO FORTE
MILAN UNIVERSITY & INFN

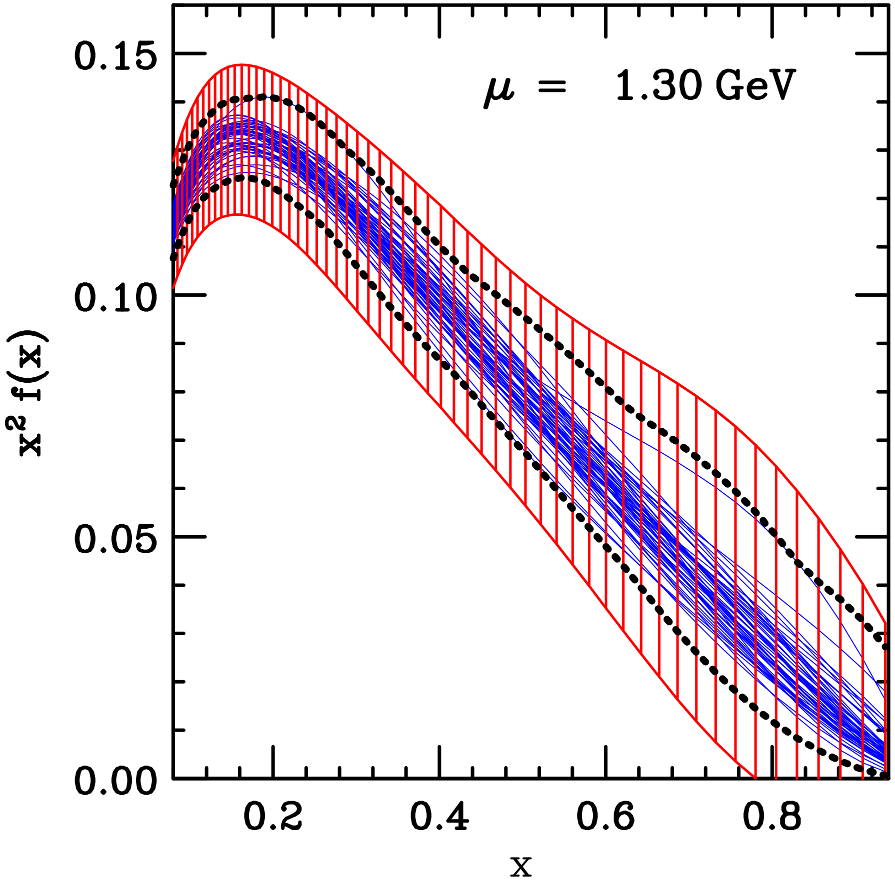
PDF4LHC WORKSHOP

CERN, MAY 29, 2009

A PUZZLE?

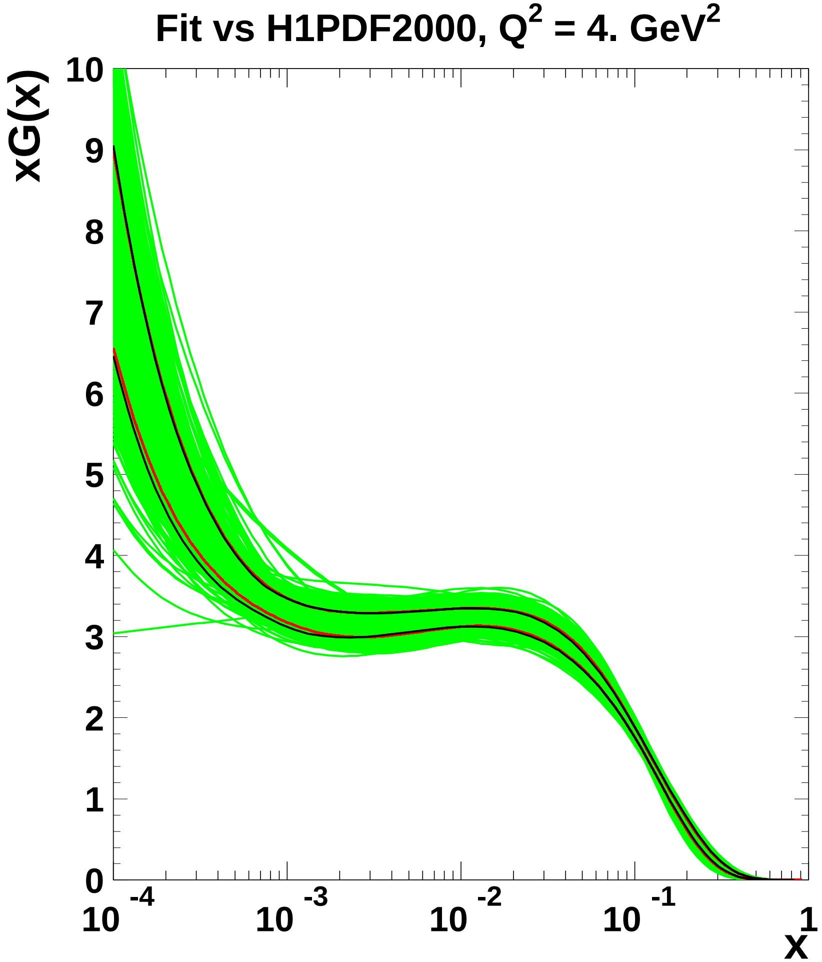
GLUON DISTRIBUTION

CTEQ Hessian result compared to envelope of 500 MC replicas



Pumplin et al., 2009

H1 Hessian result compared to st. dev. from 100 MC replicas



Feltesse, Glazov, Radescu, 2008

HESSIAN ERROR ESTIMATES

THE STANDARD METHOD

OBSERVABLE X DEPENDING ON PARAMETERS $\vec{z}$: (LINEAR ERROR PROPAGATION)

$$X(\vec{z}) \approx X_0 + z_i \partial_i X(\vec{z}) \qquad \text{ASSUMING MOST LIKELY VALUE AT } \vec{z} = 0$$

VARIANCE: $\sigma_X^2 = \sigma_{ij} \partial_i X \partial_j X,$

$\sigma_{ij} \Rightarrow$ COVARIANCE MATRIX IN PARAMETER SPACE

MAXIMUM LIKELIHOOD: COVARIANCE $\Leftrightarrow$ HESSIAN $\sigma_{ij} = \partial_i \partial_j \chi^2$ EVALUATED AT MIN. OF χ^2

DIAGONALIZATION: CHOOSE z_i AS EIGENVECTORS OF σ_{ij} WITH UNIT EIGENVALUES

$$\sigma_X^2 = |\vec{\nabla} X|^2 \text{ (LENGTH OF GRADIENT)}$$

THE ENVELOPE METHOD

ONE SIGMA CONTOUR IN PARAMETER SPACE:

TWO PARAMETERS $X - X_0 = \partial_1 X \cos\theta + \partial_2 X \sin\theta$

n PARAMETERS $X - X_0 = \vec{n} \cdot \vec{\nabla} X$; ($\vec{n}$ UNIT VECTOR IN PARAMETER SPACE)

HALF-WIDTH OF THE ENVELOPE: $\text{MAX}(X) = |\vec{\nabla} X| = \sigma_X$ (MAX IF $\vec{n}$& $\vec{\nabla} X$ PARAL.)

$\Rightarrow$ ENVELOPE & STANDARD DEVIATION COINCIDE

THE "HESSIAN MONTE CARLO"

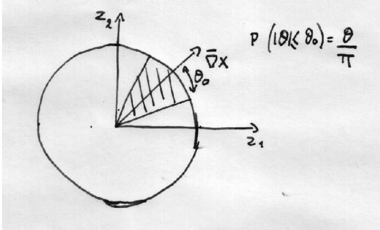
Q: IF ONE PICKS REPLICAS AT RANDOM ON THE ONE-SIGMA CONTOUR WHAT IS THE CHANCE OF "FILLING" THE ENVELOPE?

A: DETERMINE THE PROBABILITY FOR AT LEAST ONE REPLICA TO BE WITHIN ANGLE θ OF DIRECTION $\vec{\nabla} X$ OF MAX

TWO PARAMETERS: ONE REPLICA WITH $\theta < \theta_0 \Rightarrow P(2, 1 : \theta_0) = \frac{\theta_0}{\pi}$

PROBABILITY OF MAX(ENVELOPE)=$\sigma_X \cos\theta_0$

$\Rightarrow$ ALL n REPLICAS HAVE $\theta > \theta_0 \Rightarrow P(2, n; \theta_0) = \left(1 - \frac{\theta_0}{\pi}\right)^n$

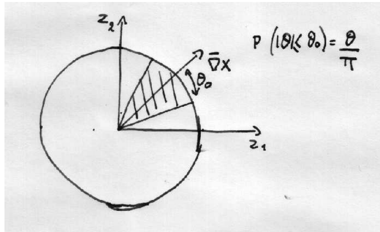


THE "HESSIAN MONTE CARLO"

Q: IF ONE PICKS REPLICAS AT RANDOM ON THE ONE-SIGMA CONTOUR WHAT IS THE CHANCE OF "FILLING" THE ENVELOPE?

A: DETERMINE THE PROBABILITY FOR AT LEAST ONE REPLICA TO BE WITHIN ANGLE θ OF DIRECTION $\vec{\nabla} X$ OF MAX

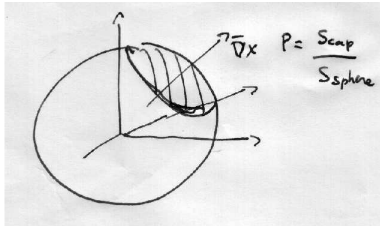
TWO PARAMETERS: ONE REPLICA WITH $\theta < \theta_0 \Rightarrow P(2,1:\theta_0) = \frac{\theta_0}{\pi}$



PROBABILITY OF MAX(ENVELOPE)=$\sigma_X \cos\theta_0$

$\Rightarrow$ ALL n REPLICAS HAVE $\theta > \theta_0 \Rightarrow P(2,n;\theta_0) = \left(1 - \frac{\theta_0}{\pi}\right)^n$

d PARAMETERS: ONE REPLICA WITH $\theta < \theta_0$



$$\Rightarrow P(d,1:\theta_0) = \frac{\Gamma\left(\frac{d}{2}\right)}{(d-1)\sqrt{\pi}\,\Gamma\left(\frac{d-1}{2}\right)}\theta_0^{d-1}(1+O(\theta_0)) \approx \frac{\theta_0^{d-1}}{\sqrt{2\pi d}}$$

PROBABILITY OF MAX(ENVELOPE)=$\sigma_X \cos\theta_0$

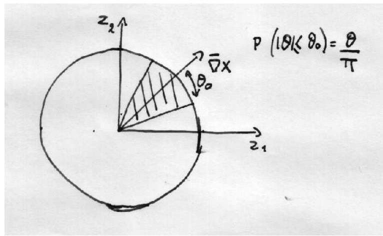
$$\Rightarrow P(d,n;\theta_0) = \left(1 - \frac{\theta_0^{d-1}}{\sqrt{2\pi d}}\right)^n$$

THE "HESSIAN MONTE CARLO"

Q: IF ONE PICKS REPLICAS AT RANDOM ON THE ONE-SIGMA CONTOUR WHAT IS THE CHANCE OF "FILLING" THE ENVELOPE?

A: DETERMINE THE PROBABILITY FOR AT LEAST ONE REPLICA TO BE WITHIN ANGLE θ OF DIRECTION $\vec{\nabla} X$ OF MAX

TWO PARAMETERS: ONE REPLICA WITH $\theta < \theta_0 \Rightarrow P(2,1:\theta_0) = \frac{\theta_0}{\pi}$

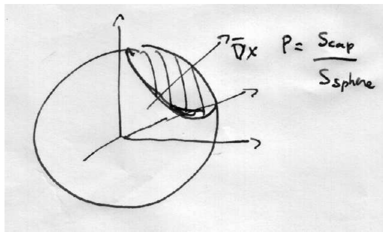


PROBABILITY OF MAX(ENVELOPE)=$\sigma_X \cos\theta_0$

$\Rightarrow$ ALL n REPLICAS HAVE $\theta > \theta_0 \Rightarrow P(2,n;\theta_0) = \left(1 - \frac{\theta_0}{\pi}\right)^n$

d PARAMETERS: ONE REPLICA WITH $\theta < \theta_0$

$$\Rightarrow P(d,1:\theta_0) = \frac{\Gamma\left(\frac{d}{2}\right)}{(d-1)\sqrt{\pi}\,\Gamma\left(\frac{d-1}{2}\right)} \theta_0^{d-1}(1+O(\theta_0)) \approx \frac{\theta_0^{d-1}}{\sqrt{2\pi d}}$$



PROBABILITY OF MAX(ENVELOPE)=$\sigma_X \cos\theta_0$

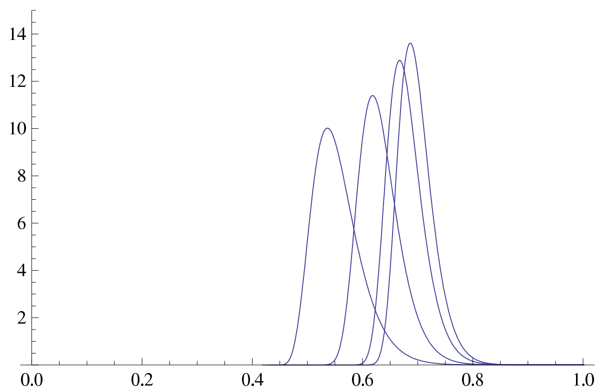
$$\Rightarrow P(d,n;\theta_0) = \left(1 - \frac{\theta_0^{d-1}}{\sqrt{2\pi d}}\right)^n$$

PROBABILITY FOR THE WIDTH OF THE ENVELOPE

TO BE SMALLER BY A FACTOR R THAN THE STANDARD DEVIATION σ_X

PLOTTED VS R FOR

$d = 23$ PARAMETERS AND $n = 10,\ 100,\ 500,\ 1000$ REPLICAS



MONTE CARLO ERROR ESTIMATES

PARAMETER SPACE:

OBSERVABLE X DEPENDS ON PARAMETERS $\vec{z}$

VARIANCE: $\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$

AVERAGES: $\langle X \rangle = \int d^d z X(\vec{z}) P(\vec{z})$, WITH

$P(\vec{z}) \Rightarrow$ PROBABILITY DISTN. OF PARAMETER VALUES

& INTEGRAL PERFORMED BY MONTE CARLO SAMPLING

HOW MANY REPLICAS DOES ONE NEED?

MONTE CARLO ERROR ESTIMATES

PARAMETER SPACE: NOT ADVISABLE

OBSERVABLE X DEPENDS ON PARAMETERS $\vec{z}$

VARIANCE: $\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$

AVERAGES: $\langle X \rangle = \int d^d z X(\vec{z}) P(\vec{z})$, WITH

$P(\vec{z}) \Rightarrow$ PROBABILITY DISTN. OF PARAMETER VALUES

& INTEGRAL PERFORMED BY MONTE CARLO SAMPLING

HOW MANY REPLICAS DOES ONE NEED? THREE BINS PER PARM $\Rightarrow 3^d$ BINS

FOR 23 PARMS., NEED $> \; 10^{11}$ REPLICAS

MONTE CARLO ERROR ESTIMATES

PARAMETER SPACE: NOT ADVISABLE

OBSERVABLE X DEPENDS ON PARAMETERS $\vec{z}$

VARIANCE: $\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$

AVERAGES: $\langle X \rangle = \int d^d z X(\vec{z}) P(\vec{z})$, WITH

$P(\vec{z}) \Rightarrow$ PROBABILITY DISTN. OF PARAMETER VALUES

& INTEGRAL PERFORMED BY MONTE CARLO SAMPLING

HOW MANY REPLICAS DOES ONE NEED? THREE BINS PER PARM $\Rightarrow 3^d$ BINS

FOR 23 PARMS., NEED $> 10^{11}$ REPLICAS

DATA SPACE

DIAGONALIZATION: CHOOSE PARM z_1 ALONG $\vec{\nabla} X$

ALL OTHER PARMS $\Rightarrow$ FLAT DIRECTIONS

AVERAGES: $\langle X \rangle = \int dz_1 X(\vec{z}) P(z_1)$

HOW MANY REPLICAS DOES ONE NEED?

MONTE CARLO ERROR ESTIMATES

PARAMETER SPACE: NOT ADVISABLE

OBSERVABLE X DEPENDS ON PARAMETERS $\vec{z}$

VARIANCE: $\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$

AVERAGES: $\langle X \rangle = \int d^d z X(\vec{z}) P(\vec{z})$, WITH

$P(\vec{z}) \Rightarrow$ PROBABILITY DISTN. OF PARAMETER VALUES

& INTEGRAL PERFORMED BY MONTE CARLO SAMPLING

HOW MANY REPLICAS DOES ONE NEED? THREE BINS PER PARM $\Rightarrow 3^d$ BINS

FOR 23 PARMS., NEED $> 10^{11}$ REPLICAS

DATA SPACE

DIAGONALIZATION: CHOOSE PARM z_1 ALONG $\vec{\nabla} X$

ALL OTHER PARMS $\Rightarrow$ FLAT DIRECTIONS

AVERAGES: $\langle X \rangle = \int dz_1 X(\vec{z}) P(z_1)$

HOW MANY REPLICAS DOES ONE NEED? ONE–DIMENSIONAL AVERAGE OF n REPLICAS CONVERGES TO TRUE AVERAGE WITH STANDARD DEV. $\frac{\sigma}{\sqrt{n}}$

10 REPLICAS ENOUGH FOR $\frac{\sigma}{3}$ ACCURACY

MONTE CARLO ERROR ESTIMATES

PARAMETER SPACE: NOT ADVISABLE

OBSERVABLE X DEPENDS ON PARAMETERS $\vec{z}$

VARIANCE: $\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$

AVERAGES: $\langle X \rangle = \int d^d z X(\vec{z}) P(\vec{z})$, WITH

$P(\vec{z}) \Rightarrow$ PROBABILITY DISTN. OF PARAMETER VALUES

& INTEGRAL PERFORMED BY MONTE CARLO SAMPLING

HOW MANY REPLICAS DOES ONE NEED? THREE BINS PER PARM $\Rightarrow 3^d$ BINS

FOR 23 PARMS., NEED $> \; 10^{11}$ REPLICAS

DATA SPACE

DIAGONALIZATION: CHOOSE PARM z_1 ALONG $\vec{\nabla} X$

ALL OTHER PARMS $\Rightarrow$ FLAT DIRECTIONS

AVERAGES: $\langle X \rangle = \int dz_1 X(\vec{z}) P(z_1)$

HOW MANY REPLICAS DOES ONE NEED? ONE–DIMENSIONAL AVERAGE OF n REPLICAS CONVERGES TO TRUE AVERAGE WITH STANDARD DEV. $\frac{\sigma}{\sqrt{n}}$

10 REPLICAS ENOUGH FOR $\frac{\sigma}{3}$ ACCURACY

Q: HOW IS IT DONE IN PRACTICE?

A: CHOOSE REPLICAS OF THE DATA, DISTRIBUTED AS THE DATA

MONTE CARLO ERROR ESTIMATES

MANY OBSERVABLES

IN GLOBAL PARTON FITS, THE NUMBER OF OBSERVABLES (DATA) IS LARGE

BUT THEY ARE NOT COMPLETELY INDEPENDENT!

Q: HOW MANY REPLICAS DOES ONE NEED? AT WORST, COULD BE VERY LARGE $> N_{\rm dat}$

A: JUST TRY. COMPUTE AVERAGES, VARIANCES, COVARIANCES FROM MC SAMPLE

& ENLARGE SAMPLE UNTIL THEY AGREE WITH THOSE OF DATA

MONTE CARLO ERROR ESTIMATES

MANY OBSERVABLES

IN GLOBAL PARTON FITS, THE NUMBER OF OBSERVABLES (DATA) IS LARGE

BUT THEY ARE NOT COMPLETELY INDEPENDENT!

Q: HOW MANY REPLICAS DOES ONE NEED? AT WORST, COULD BE VERY LARGE $> N_{\rm dat}$

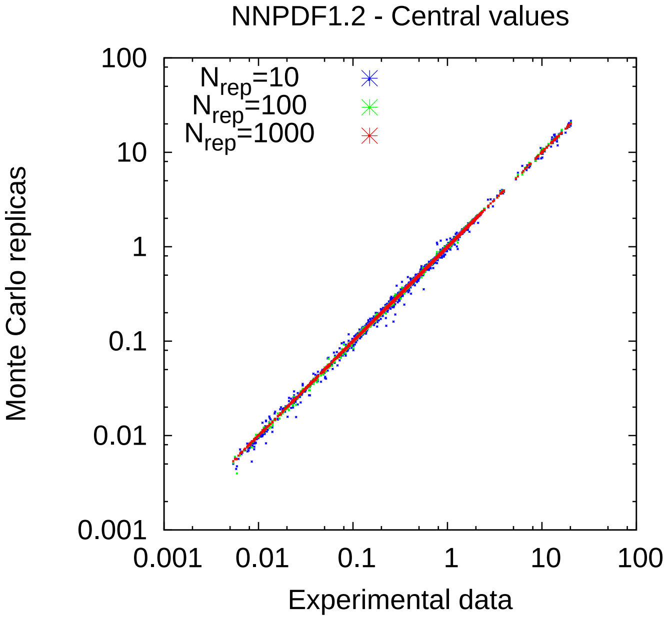
A: JUST TRY. COMPUTE AVERAGES, VARIANCES, COVARIANCES FROM MC SAMPLE

& ENLARGE SAMPLE UNTIL THEY AGREE WITH THOSE OF DATA

NNPDF1.2 DATASET, $N_{\rm dat} \sim 3500$

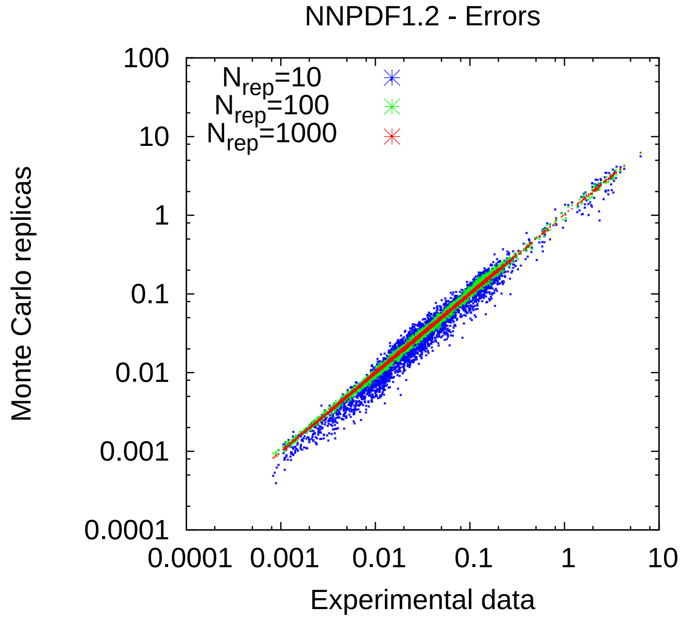
REPLICA AVERAGES

VS. CENTRAL VALUES



REPLICA STANDARD DEV.

VS. UNCERTAINTIES



10 REPLICAS FOR CENTRAL VALS, 100 FOR UNCERTAINTIES, 1000 FOR CORRELNS

NORMALIZATION UNCERTAINTIES

HESSIAN

- FIRST OPTION (MSTW08) RESCALE ALL DATA VALUES & ST. DEVNS. BY FACTOR f & TREAT f AS A FIT PARM WITH A PENALTY TERM SUCH AS
$$\chi_f^2 = \frac{(f-1)^2}{\sigma_f^2}$$
- SECOND OPTION TREAT NORMALIZATION AS AN OFFSET UNCERTAINTY $x = x_{true} + \sigma_0$
ADD OFFSET UNCERTAINTY AS A CORRELATED UNCERTAINTY TO COVARIANCE MATRIX
$$\mathrm{cov}_{ij} = \left(\sum_{k=1}^{N_{\mathrm{sys}}} \sigma_{i,k}\sigma_{j,k} + \delta_{ij}\sigma_{i,t}\right)$$
- WRONG OPTIONS RESCALE DATA VALUES BUT NOT STANDARD DEVIATIONS & FIT WITH PENALTY;
ADD NORMALIZATION UNCERTAINTY TO COVARIANCE MATRIX
$$\mathrm{cov}_{ij} = \left(\sum_{k=1}^{N_{\mathrm{sys}}} \sigma_{i,k}\sigma_{j,k} + F_i F_j \sigma_N^2\right) + \delta_{ij}\sigma_{i,t}.$$

MONTE CARLO

- FIRST OPTION (NNPDF) INCLUDE NORMALIZATION UNCERTAINTY ALONG WITH OTHER SYSTEMATICS IN DATA GENERATION
$$F_i^{(\mathrm{art})(k)} = \left(1 + r_N^{(k)}\sigma_N\right)\left(F_i^{(\mathrm{exp})} + \sum_{p=1}^{N_{\mathrm{sys}}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,t}\right)$$
- SECOND OPTION NORMALIZATION IS NOT INCLUDED IN DATA GENERATION, & TREATED AS A FREE PARAMETER, TO BE FITTED REPLICA BY REPLICA

INCOMPATIBLE DATA

WHEN USED TOGETHER WITH NEURAL NETWORK & CROSS–VALIDATION STOPPING

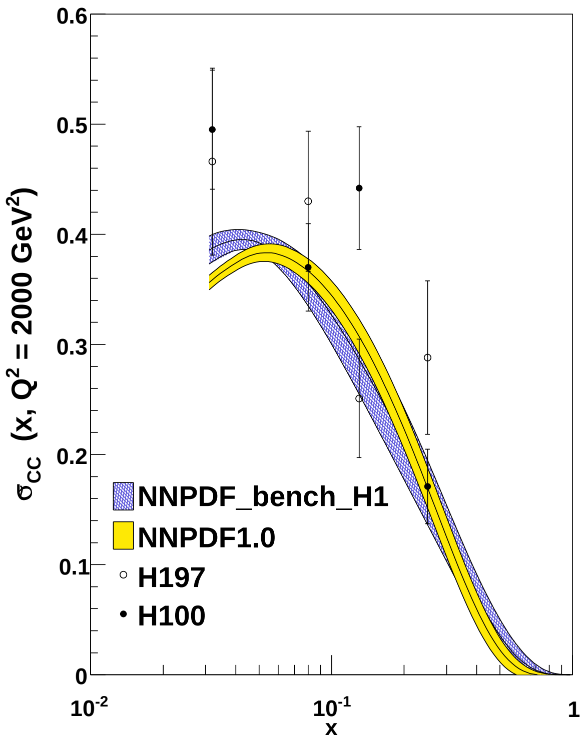
MONTE CARLO LEADS TO INFLATION OF UNCERTAINTY

WHEN COMBINING INCONSISTENT DATA ⇒ SIMILAR TO PDG (SCALE FACTOR) METHOD

EXAMPLE BENCHMARK FIT (HERALHC WSHOP)

LARGER COMPATIBLE DATASET

⇒ ERROR REDUCTION



INCOMPATIBLE DATA

WHEN USED TOGETHER WITH NEURAL NETWORK & CROSS–VALIDATION STOPPING

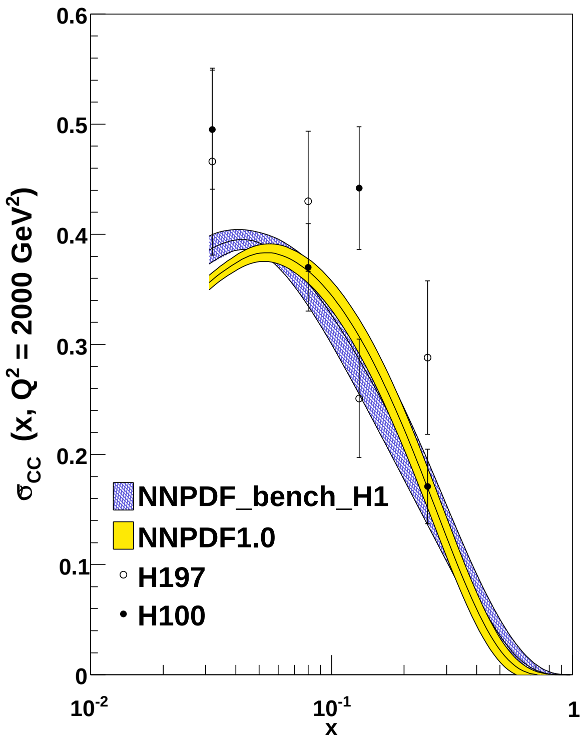
MONTE CARLO LEADS TO INFLATION OF UNCERTAINTY

WHEN COMBINING INCONSISTENT DATA ⇒ SIMILAR TO PDG (SCALE FACTOR) METHOD

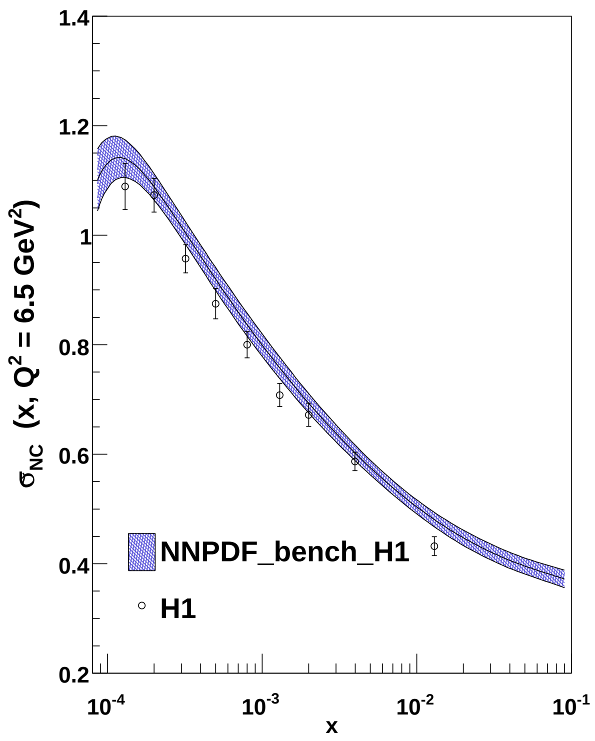
EXAMPLE BENCHMARK FIT (HERALHC WSHOP)

LARGER COMPATIBLE DATASET

⇒ ERROR REDUCTION



LARGER DATASET



INCOMPATIBLE DATA

WHEN USED TOGETHER WITH NEURAL NETWORK & CROSS–VALIDATION STOPPING

MONTE CARLO LEADS TO INFLATION OF UNCERTAINTY

WHEN COMBINING INCONSISTENT DATA ⇒ SIMILAR TO PDG (SCALE FACTOR) METHOD

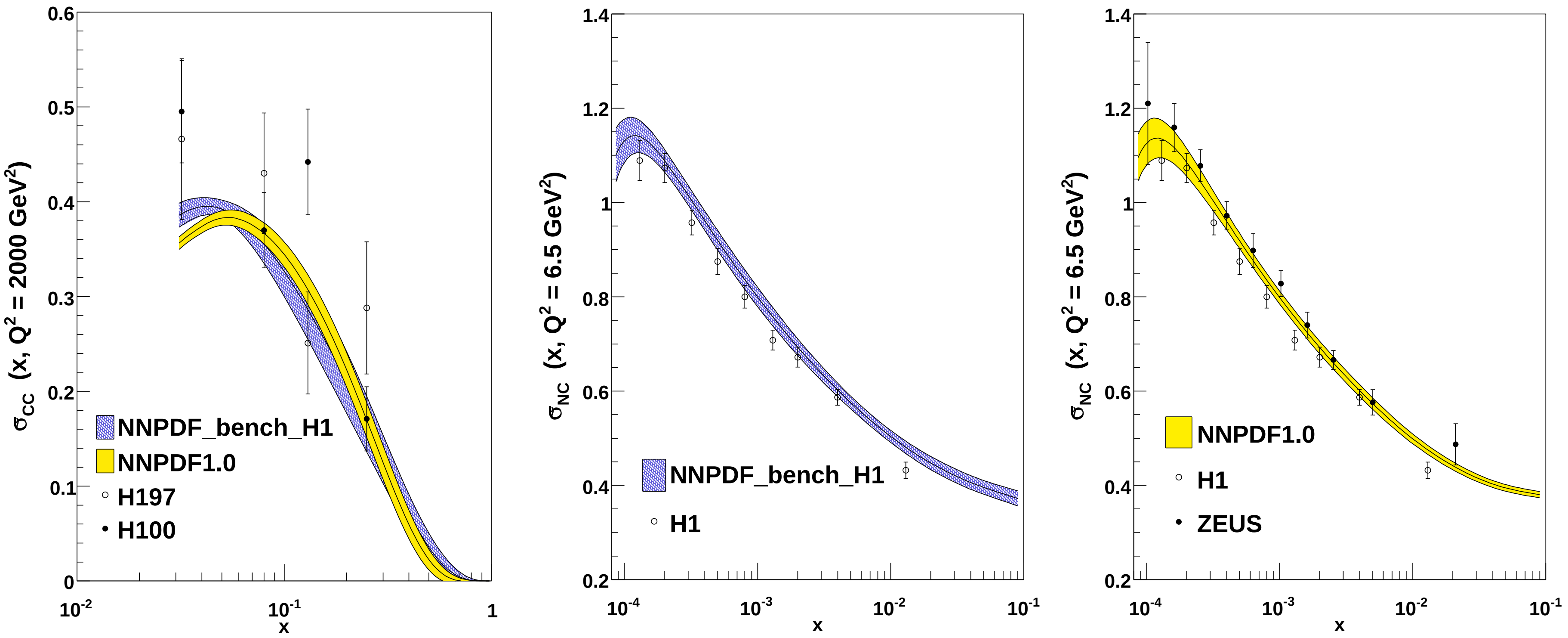
EXAMPLE BENCHMARK FIT (HERALHC WSHOP)

LARGER COMPATIBLE DATASET

⇒ ERROR REDUCTION

LARGER DATASET INCOMPATIBLE

⇒ NO ERROR REDUCTION



CONCLUSION

- HESSIAN & MONTE CARLO METHOD GIVE THE SAME RESULTS IN THE LINEAR (ERROR PROPAGATION) APPROXIMATION
- MONTE CARLO MORE FLEXIBLE IN HANDLING NON GAUSSIAN OR NON LINEAR BEHAVIOUR
- MONTE CARLO MUST BE DONE IN THE SPACE OF DATA, NOT IN THE SPACE OF PARAMETERS
- MONTE CARLO COUPLED WITH NEURAL NETWORKS AND CROSS-VALIDATION STOPPING ALLOWS FOR AN IMROVED TREATMENT OF INCONSISTENT DATA