On the determination of uncertainties in parton densities

N. Hunt-Smith, **Alberto Accardi**, W. Melnitchouk,
N. Sato, A.W. Thomas, M.J. White

Phys. Rev. D 106 (2022) 036003

DIS 2023
March 29, 2023

HAMPTON UNIVERSITY
THE STANDARD OF EXCELLENCE

Jefferson Lab

This work is in part supported by the DOE Office of Science

Overview

- **Uncertainty Quantification: Parametric Methods**
  - Monte Carlo Bayesian estimators
  - Hessian approximation
  - Data resampling

- **Description of Toy Model**
  - Benchmark of Hessian and MC methods

- **Neural Network Comparison**
  - Algorithmic modification of likelihood?

    (see also N. Sato @ DIS 2018)

A whole session devoted to PDF uncertainties:

- ➢ P. Nadolsky – “Epistemic uncert. quant.”
- ➢ L. Kotz – “Bezier curve parametrizations”
- ➢ K. Mohan – A new statistical method”

Uncertainty quantification: parametric methods

Bayesian estimators

- **Bayes theorem** $p(a|m) = \frac{1}{Z}\, p(m|a)\, p(a)$

  with "evidence" $Z = \int \mathrm{d}a\; p(m|a)\, p(a)$

  and "likelihood" $p(m|a) = \mathcal{N} \exp\left[-\frac{1}{2}\chi^2(a,m)\right]$

  Typical choice in PDF analyses

- Algorithms for sampling of likelihood ⟶ {$\boldsymbol{a}_k$}
  - **HMC**: Hamiltonian Monte Carlo (an example of Markov-Chain MC methods)
  - **NS**: Nested Sampling, primarily aimed at estimating the evidence

    ⟶ Samples the likelihood as a byproduct

- Expectation values $E_{\mathrm{Bayes}}\{\mathcal{O}(a)\} = \frac{1}{n}\sum_{k=1}^{n} \mathcal{O}(a_k)\,,$

  and variance $V_{\mathrm{Bayes}}\{\mathcal{O}(a)\} = \frac{1}{n}\sum_{k=1}^{n} \left[\mathcal{O}(a_k) - E_{\mathrm{Bayes}}\{\mathcal{O}(a)\}\right]^2$

Data resampling

- Data Resampling (**DR**) approximates Bayes' posterior using frequentist logic
  - Reshuffle data within data uncertainty (Gaussian distribution)
  - Maximize likelihood
  - Repeat n_{rep} times $\longrightarrow \{\boldsymbol{a}_k\}$

- Estimate

$$E_{\text{freq}}\{\mathcal{O}(a)\} = \frac{1}{n_{\text{rep}}} \sum^{n_{\text{rep}}} \mathcal{O}(a_{\text{rep}}) \,,$$

$$V_{\text{freq}}\{\mathcal{O}(a)\} = \frac{1}{n_{\text{rep}}} \sum^{n_{\text{rep}}} \left[\mathcal{O}(a_{\text{rep}}) - E_{\text{freq}}\{\mathcal{O}(a)\}\right]^2$$

- Good in parameter space region well constrained by data

Generalized Hessian Approximation

Hunt-Smith et al., PRD 106 (2022) 036003

- Start as usual:
  - Find minimum of likelihood
  - Diagonalize Hessian → $\boldsymbol{e}_k$ eigenvectors, w_k eigenvalues
- Change variables: $a(t) = a_0 + \sum_{k=1}^{n_{\rm par}} t_k \frac{e_k}{\sqrt{w_k}}$, then $p(a|m) \to p(t|m)$
- Assume likelihood factorized along Hessian eigendirection, then

$$E_{\rm Hess}\{\mathcal{O}(a)\} = \int d^n t \; p(t|m)\, \mathcal{O}(a(t)) \approx \mathcal{O}(a_0)$$

$$V_{\rm Hess}\{\mathcal{O}(a)\} \approx \sum_k T_k^2 \left(\frac{\partial \mathcal{O}(a(t))}{\partial t_k}\bigg|_{a_0} \right)^2$$

- Here $T_k^2 = \int dt_k \; p_k(t_k|m)\, t_k^2$ is the "tolerance" :
  - T_k = 1 where likelihood is Gaussian;
  - Approximates well the likelihood in non-Gaussian directions
  - Maintains a "68%" or "1σ" kind of meaning also when ≠ 1

CT, MSTW → T=5-10

- Often T_k determined "ad hoc" to account for statistical inconsistency of data

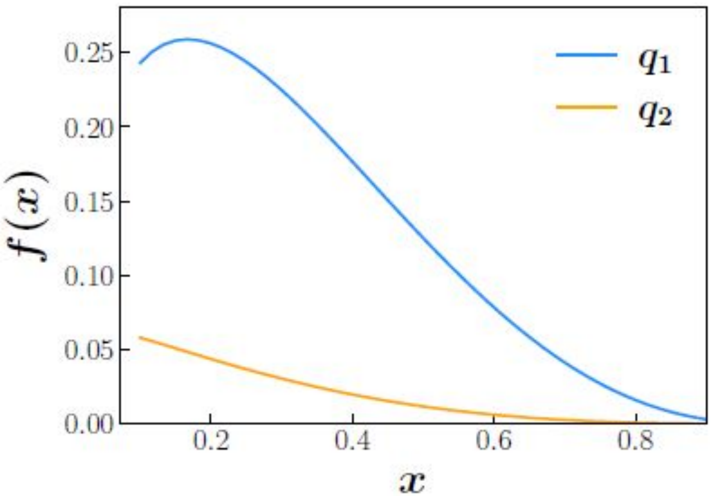
Toy Model

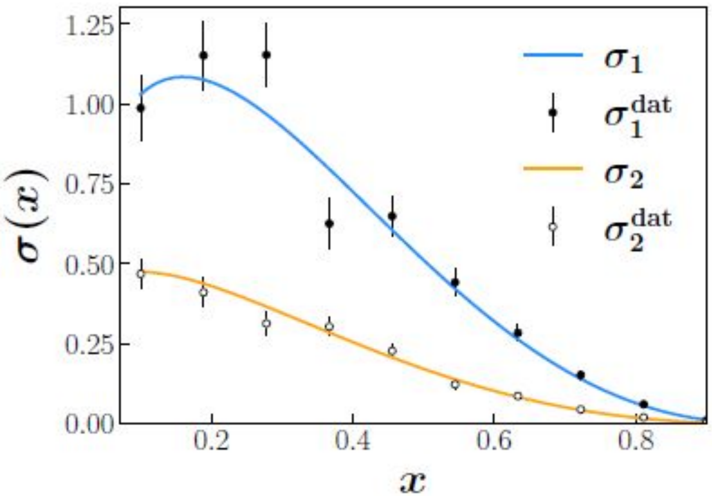
Toy model

- **PDFs *f*** : mimic up and down quarks
- **Observables *σ*** : mimic proton, neutron DIS cross section at fixed Q^2
  - Data randomly generated according to corresponding x distributions

$$q_i(x) = x^{\alpha_i}(1-x)^{\beta_i}, \quad i = 1, 2.$$

$$\sigma_j = \sum_{i=1,2} c_{ji}\, q_i, \quad c_{11} = 4c_{12} = 4c_{21} = c_{22}.$$



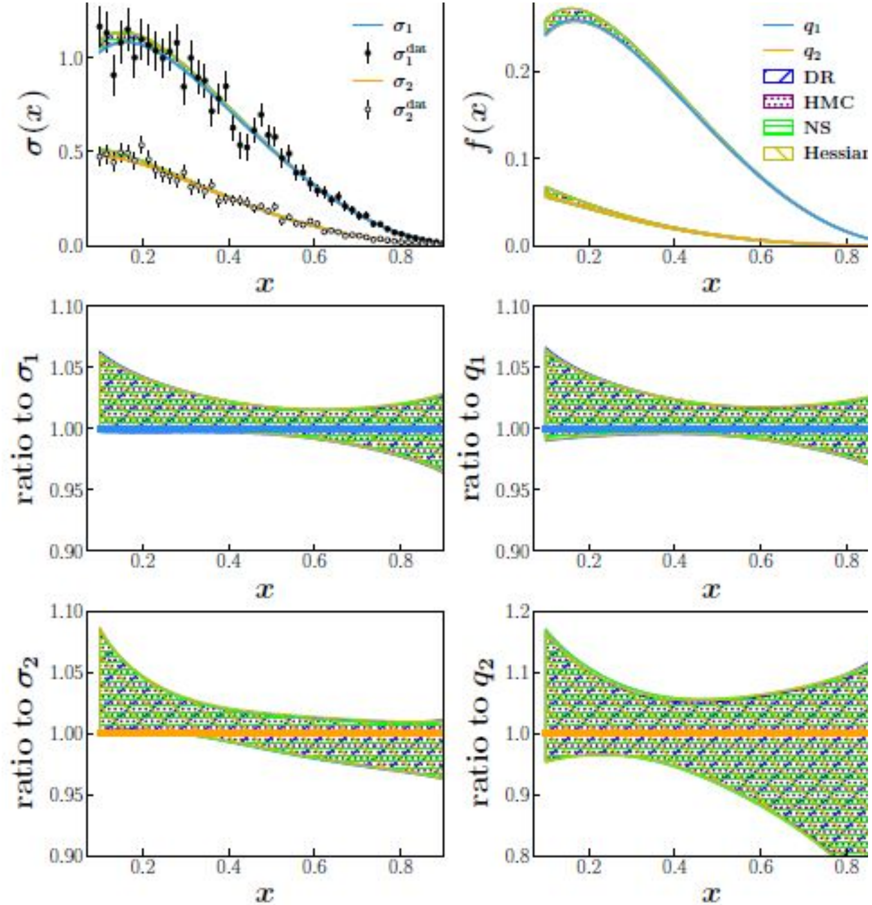


Equivalency of parametric methods

- **Bayesian MC estimators** used as benchmark

- **Hessian approximation is good!**
  - Generalized tolerance marginally needed even in this simplified example

- Crucially, **data resampling provides same likelihood estimation** as Bayesian MC methods



Neural Network Fits

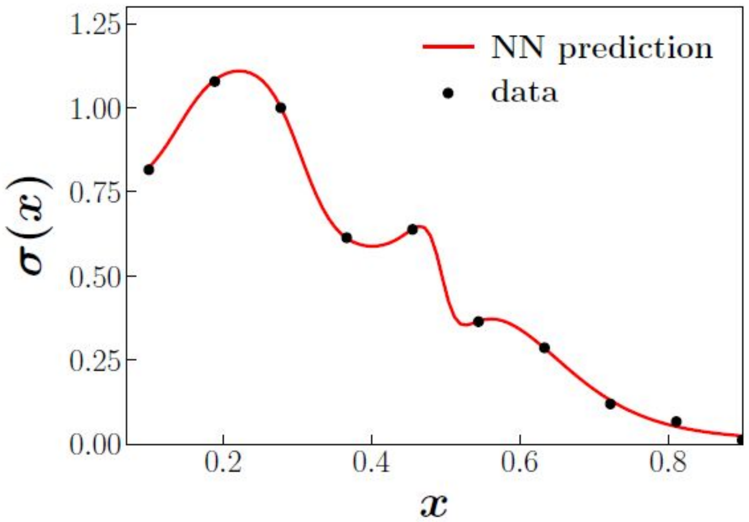
Neural Networks and overfitting

- Neural networks provide:
  - Efficient, very flexible parametrizations
  - Hundreds of parameters
  - Essentially a parameter free functional form

- Aim at maximizing the same likelihood

$$p(m|a) = \mathcal{N} \exp\left[-\frac{1}{2}\chi^2(a,m)\right]$$

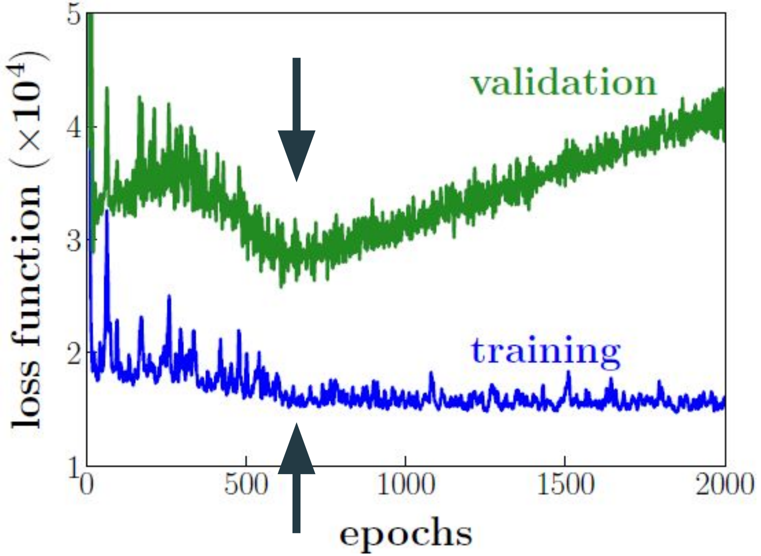
- Without intervention, will overfit the data
  - The plot shows an extreme example



Cross-validation (CV) and stopping

- Needs a “stopping criterion”
  - to avoid fitting statistical noise instead of physics
- Randomly separate the data into 2 groups, say
  - 70% ⟶ training (T)
  - 30 % ⟶ validation (V)
- Fit the training, calculate χ^2(T) and χ^2(V)
- Resample data, repeat
- “Stop” training when χ^2(V) is minimum:

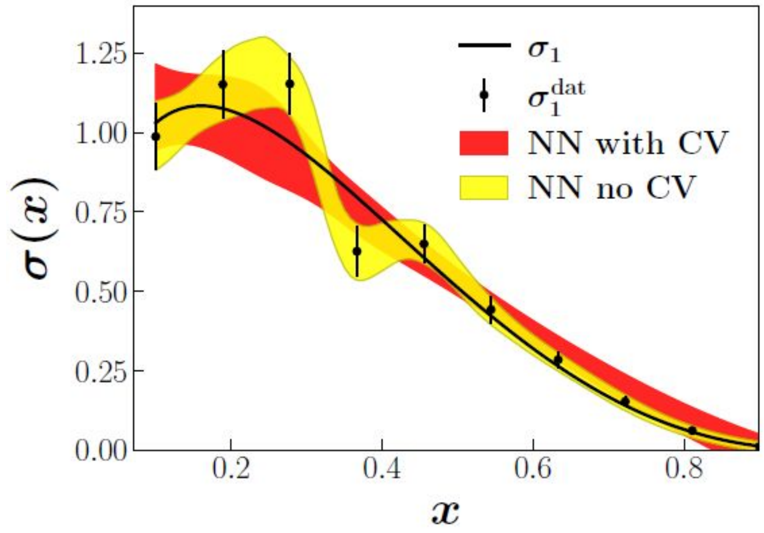
$$\sigma = E[\sigma_{\rm fit}]$$
$$\delta\sigma = V[\sigma_{\rm fit}]$$



Cross-validation (CV) and stopping

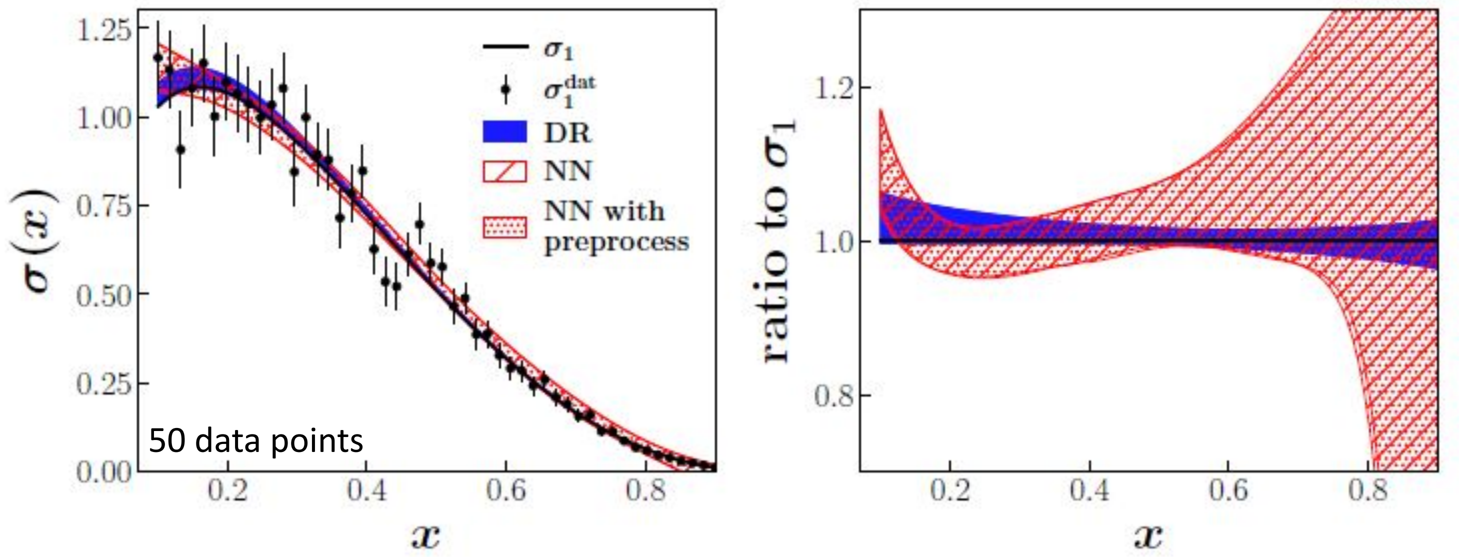
- Needs a "stopping criterion"
  - to avoid fitting statistical noise instead of physics
- Randomly separate the data into 2 groups, say
  - 70% $\longrightarrow$ training (T)
  - 30 % $\longrightarrow$ validation (V)
- Fit the training, calculate χ^2(T) and χ^2(V)
- Resample data, repeat
- "Stop" training when χ^2(V) is minimum:

$$\sigma = E[\sigma_{\rm fit}]$$
$$\delta\sigma = V[\sigma_{\rm fit}]$$



Comparison of NN to parametric methods

- DR as representative of parametric methods
- **Neural Network fits:**
  - Comparable in shape
  - **Quite larger uncertainty!**

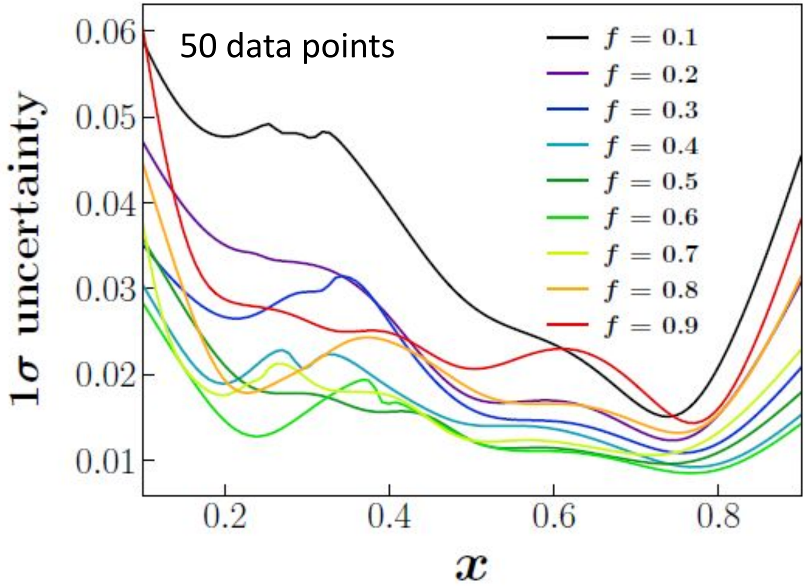


Dependence on training fraction *f*

- **The fit is quite independent of the T/V partitioning**
  - b/c in each replica the training data is randomly chosen
  - So it spans the whole *x* range

BUT

- **The uncertainty strongly depends on the fraction of training data**
  - with f ≈ 0.6 providing
    the smallest uncertainty

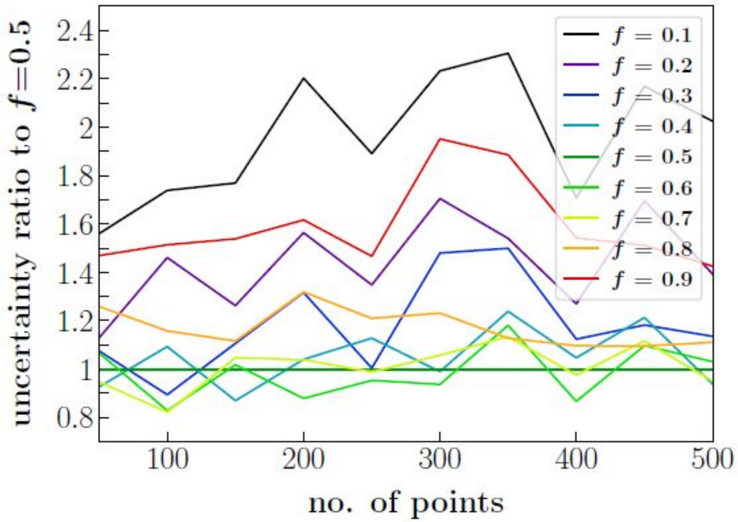


Dependence on training fraction *f*

- **The fit is quite independent of the T/V partitioning**
  - b/c in each replica the training data is randomly chosen
  - So it spans the whole x range

BUT

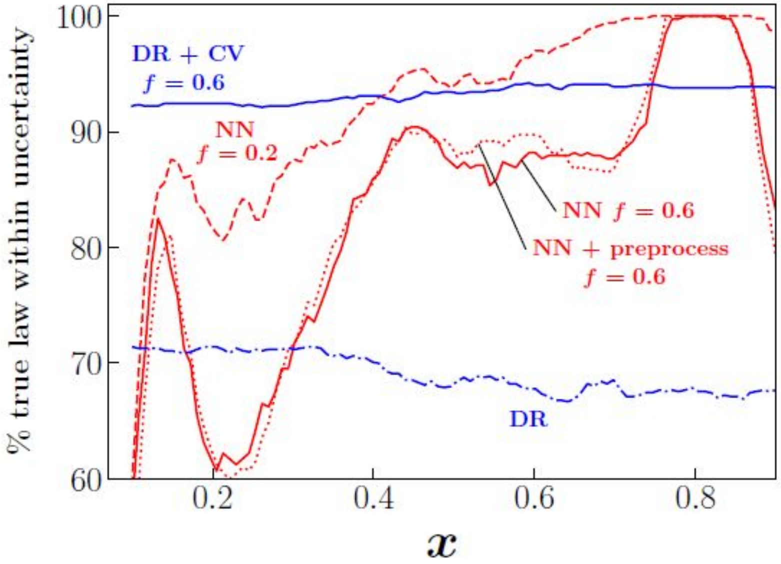
- **The uncertainty strongly depends on the fraction of training data**
  - with f ≈ 0.6 providing the smallest uncertainty
  - Independently of how dense the data is in *x*



Comparison of NN to parametric methods

- **NN fits inflate the uncertainty estimate!**
  - Partly due to cross-validation
  - Structure in x difficult to understand
  - Uncertainty explodes at large x

- Data resampling + Cross Validation also inflates the uncertainty
  - Validation set “pulls” against training set
  - But in the same way across *x*



The algorithms have effectively modified the nominal exp(-χ^2) likelihood!

In conclusion...

Food for thought

- Reliable quantification of PDF uncertainties needed for QCD and HEP applications

- **Parametric methods produce the same likelihood estimates**
  - Bayesian MC methods
  - Hessian approximation
  - Data resampling

- **Neural Network fits**
  - **Algorithmically modify the nominal likelihood**
  - The resulting uncertainties are not directly comparable to parametric estimates

    ⟶ *Enlarged uncertainties do not look like a natural replacement for tolerance criterion to account for tension in the data sets*

- **In what sense can NNPDF be combined with others in, say, PDF4LHC fits?**