A Neural Network Approach to Nuclear PDFs

Jacob Ethier

Work in progress with R. Abdul Khalek and J. Rojo
NNV Subatomic Physics Meeting
Nov. 2nd, 2018

Motivation

- Early measurements of deeply inelastic electron-iron scattering showed significant deviations from predictions (EMC effect)

- In the quark-parton model:

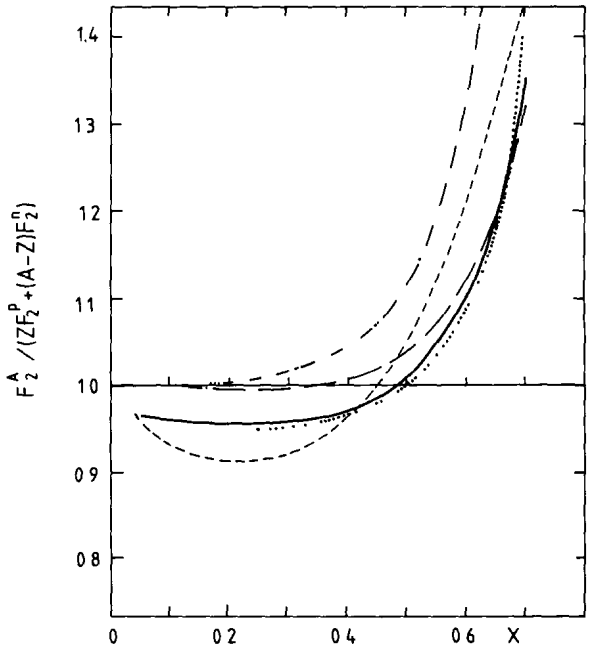
$$F_2^p(x) = x \sum_q e_q^2 \left(q(x) + \bar{q}(x)\right)$$

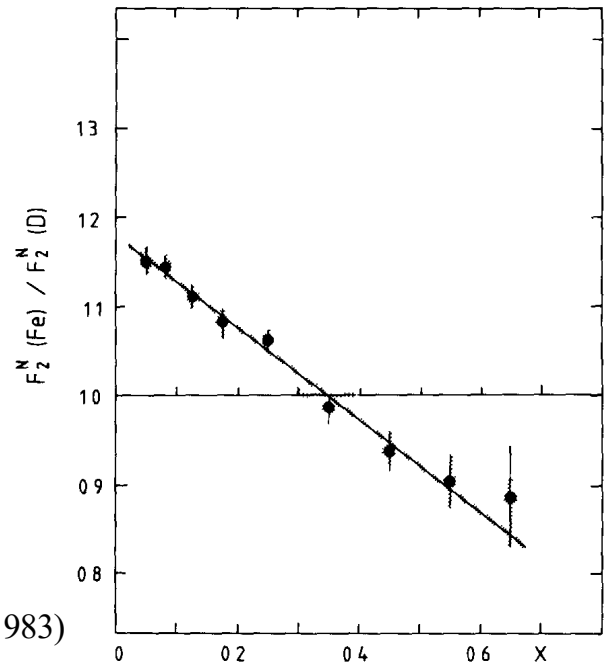
Parton distribution functions (PDFs)

→ Structure functions of the nucleus cannot be described as a sum of the free nucleon structure functions (up to Fermi motion corrections)

- EMC effect is still not well understood

→ Determining the modification of free nucleon PDFs in nuclei can provide crucial insight to origins of nuclear effects





J.J. Aubert et. al. Phys. Lett. B 123B (1983)

Motivation

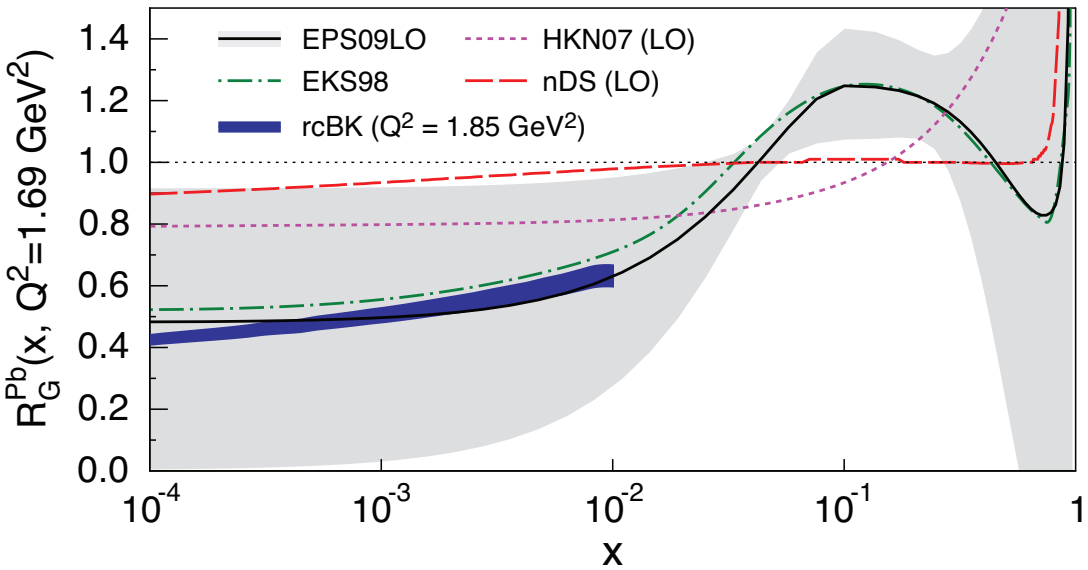
- Nuclear PDFs can also provide evidence of gluon saturation at low x and Q^2 (momentum transfer scale of lepton-nucleus scattering)

  → Enhancement for heavier nuclei – region of saturation begins at larger x

- Effect seen in ratio of nuclear to proton PDFs:

$$R_G^A = \frac{g^{(p/A)}(x, Q^2, A)}{g^{(p)}(x, Q^2)}$$

A. Accardi et. al. EIC White Paper *arXiv:1212.1701*



- Existing nuclear PDF information contains large uncertainty in the saturation region

- Precise determination of nuclear PDFs are relevant for the upcoming electron-ion collider

- How are they determined?

Global QCD Analyses

- We want to obtain reliable information of the nonperturbative dynamics associated with nuclei structure within QCD

Factorization → separation of short and long distance physics in pQCD expressions of experimental observables, e.g. for electron-proton scattering

Unpolarized deep inelastic scattering (DIS) observable

$\ell + (p, d) \to \ell' + X$

$$d\sigma(x, Q^2) \simeq \sum_f \int_x^1 \frac{d\xi}{\xi} \underbrace{f\left(\frac{x}{\xi}, Q^2\right)}_{\text{Parton distribution function (PDF)}} \underbrace{d\hat{\sigma}_f(\xi, Q^2)}_{\text{Hard scattering coefficient}}$$

- Collinear factorization → distributions depend on the fraction of longitudinal proton momentum

- The same formalism can be used to study modification of PDFs in nuclei

$$f^{(p)}(x, Q^2) \to f^{(p/A)}(x, Q^2, A)$$

- Nuclear PDFs are extracted from global data of lepton-nucleus and hadron-nucleus collisions using QCD factorization

Global QCD Analyses

- Standard determination of nonperturbative functions form global QCD analyses:

  → Objects are parameterized $xf(x) = Nx^a(1-x)^b(1+c\sqrt{x}+dx)$

  → Parameters are optimized with a least-squares fit

$$\chi^2 = \sum_e^{N_{exp}} \sum_i^{N_{data}} \frac{(D_i^e - T_i)^2}{(\sigma_i^e)^2}$$

- However, there are many issues with performing single chi-squared minimizations

  → Uncertainties computed by a Hessian method introduce tolerance criteria (uncertainties inflated by arbitrary factor)

  → Parameters difficult to constrain are typically fixed

  → Highly non-linear chi-squared function means many local minima that a single fit can be trapped in

- Nuclear PDFs require proton PDF as boundary condition → typically taken from previous global analyses (which often make different theoretical assumptions!)

- We need a consistent theoretical framework and rigorous fitting procedure to determine nuclear PDFs and, more importantly, estimate their uncertainties

Fitting Methodology

- Based on Bayesian statistical methods – robust determination of "observables" *O* (PDFs,etc.) and their uncertainties

$$E\left[\mathcal{O}\right] = \int d^n a \mathcal{P}(\vec{a}|data)\mathcal{O}(\vec{a})$$

$$V\left[\mathcal{O}\right] = \int d^n a \mathcal{P}(\vec{a}|data)\left[\mathcal{O}(\vec{a}) - E[\mathcal{O}]\right]^2$$

- Bayes' theorem defines probability $\mathcal{P}$ as

$$\mathcal{P}(\vec{a}|data) = \frac{1}{Z}\mathcal{L}(data|\vec{a})\pi(\vec{a})$$

Fitting Methodology

- Based on Bayesian statistical methods – robust determination of "observables" *O* (PDFs,etc.) and their uncertainties

$$E\left[\mathcal{O}\right] = \int d^n a \mathcal{P}(\vec{a}|data)\mathcal{O}(\vec{a})$$

$$V\left[\mathcal{O}\right] = \int d^n a \mathcal{P}(\vec{a}|data)\left[\mathcal{O}(\vec{a}) - E[\mathcal{O}]\right]^2$$

- Bayes' theorem defines probability $\mathcal{P}$ as

$$\mathcal{P}(\vec{a}|data) = \frac{1}{Z}\mathcal{L}(data|\vec{a})\pi(\vec{a})$$

Likelihood function

$$\mathcal{L} = \exp\left(-\frac{1}{2}\chi^2(\vec{a})\right)$$ → Gaussian form in data with $\chi^2 = \sum_e^{N_{exp}} \sum_i^{N_{data}} \frac{(D_i^e - T_i)^2}{(\sigma_i^e)^2}$

Fitting Methodology

- Based on Bayesian statistical methods – robust determination of "observables" *O* (PDFs,etc.) and their uncertainties

$$E\left[\mathcal{O}\right] = \int d^n a \mathcal{P}(\vec{a}|data)\mathcal{O}(\vec{a})$$

$$V\left[\mathcal{O}\right] = \int d^n a \mathcal{P}(\vec{a}|data)\left[\mathcal{O}(\vec{a}) - E[\mathcal{O}]\right]^2$$

- Bayes' theorem defines probability $\mathcal{P}$ as

$$\mathcal{P}(\vec{a}|data) = \frac{1}{Z}\mathcal{L}(data|\vec{a})\pi(\vec{a})$$

Priors

"Evidence" $Z = \int d^n a \mathcal{L}(data|\vec{a})\pi(\vec{a})$

Fitting Methodology

- Based on Bayesian statistical methods – robust determination of "observables" *O* (PDFs,etc.) and their uncertainties

$$E\left[\mathcal{O}\right] = \int d^n a \mathcal{P}(\vec{a}|data)\mathcal{O}(\vec{a})$$

$$V\left[\mathcal{O}\right] = \int d^n a \mathcal{P}(\vec{a}|data)\left[\mathcal{O}(\vec{a}) - E[\mathcal{O}]\right]^2$$

- Bayes' theorem defines probability $\mathcal{P}$ as

$$\mathcal{P}(\vec{a}|data) = \frac{1}{Z}\mathcal{L}(data|\vec{a})\pi(\vec{a})$$

- Monte Carlo technique is used to evaluate expectation value and variance integrals

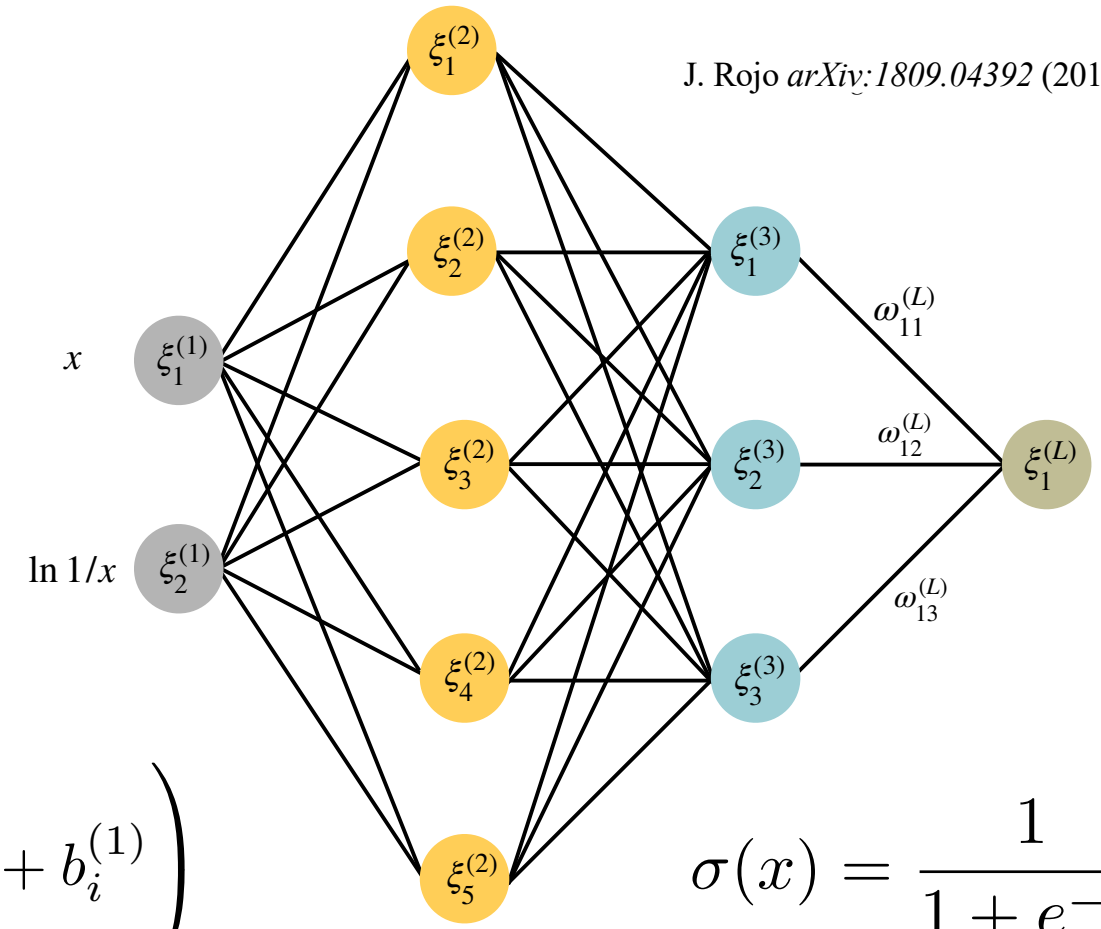
  → assuming uniform sampling of parameter space:

$$E\left[\mathcal{O}(\vec{a})\right] = \frac{1}{N}\sum_k \mathcal{O}(\vec{a}_k) \qquad V\left[\mathcal{O}(\vec{a})\right] = \frac{1}{N}\sum_k \left(\mathcal{O}(\vec{a}_k) - E[\mathcal{O}]\right)^2$$

- We can perform many fits to generate a Monte Carlo representation of the probability distribution

Neural Networks

- Consists of layers of neurons
- Each neuron contains a corresponding weight and bias (fit parameters)
- First layer of neurons take variables of function being parametrized (e.g. longitudinal momentum fraction for PDFs)



$$\xi_i^{(1)} = \sigma_i^{(1)} \left(\sum_j^{N_{\text{inputs}}} w_{ij}^{(1)} x_j + b_i^{(1)} \right)$$

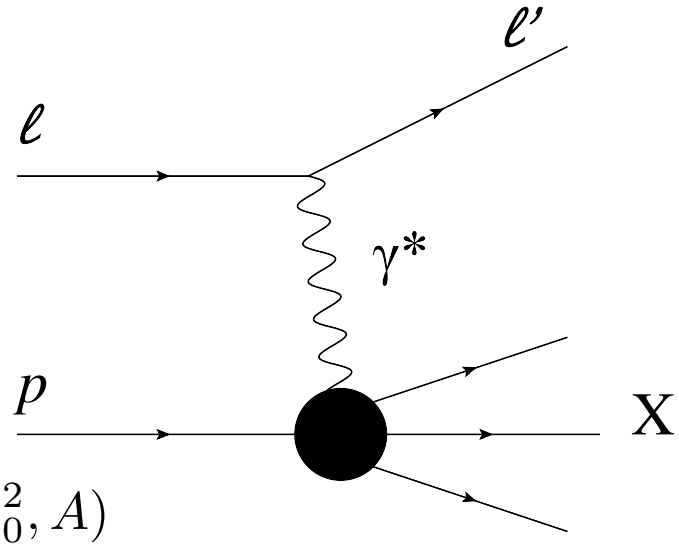
$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

- Input of subsequent layers given by activation of previous layer neurons
- Output layer contains the result of the parameterized function(s)

$$\xi_i^{(l+1)} = \sigma_i^{(l+1)} \left(\sum_j^{N_l} w_{ij}^{(l+1)} \sigma_j^{(l)} + b_i^{(l+1)} \right)$$

$$\xi_i^{(L)} = \sum_j^{N_{L-1}} w_{ij}^{(L)} \sigma_j^{(L-1)} + b_i^{(L)}$$

QCD Analysis of NC DIS



- Observable (differential cross section/structure function) defined by collinear factorization

$$d\sigma^{eA\to eX} \simeq \sum_q d\hat{\sigma}^{eq\to eX}(x,Q^2)\otimes q(x,Q^2,A)$$

$$\simeq \sum_q \underbrace{d\hat{\sigma}^{eq\to eX}(x,Q^2)}_{\text{Hard scattering cross section}} \otimes \underbrace{\Gamma(Q_0^2,Q^2)}_{\text{DGLAP Evolution}} \otimes \underbrace{q(x,Q_0^2,A)}_{\text{Nuclear PDF}}$$

- Strategy

→ Use neural network with three input features (x, ln(x), A) and three outputs (corresponding to singlet, octet, and gluon)

$$x\Sigma = (1-x)^3 \mathrm{NN}_\Sigma$$

$$xT_8 = (1-x)^3 \mathrm{NN}_{T_8}$$

$$xg = A_g(1-x)^4 \mathrm{NN}_g$$

where

$$\Sigma = u+\bar{u}+d+\bar{d}+s+\bar{s}$$

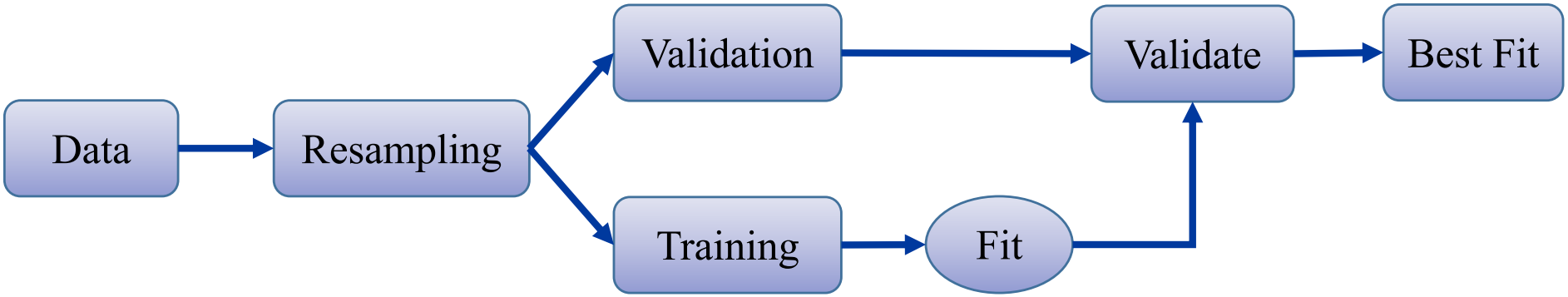
$$T_8 = u+\bar{u}+d+\bar{d}-2s-2\bar{s}$$

$$A_g = \frac{1-\int_0^1 dx x\Sigma(x,A)}{\int_0^1 dx x g(x,A)}$$

$$\int_0^1 dx x\left(\Sigma(x,A)+g(x,A)\right)=1$$

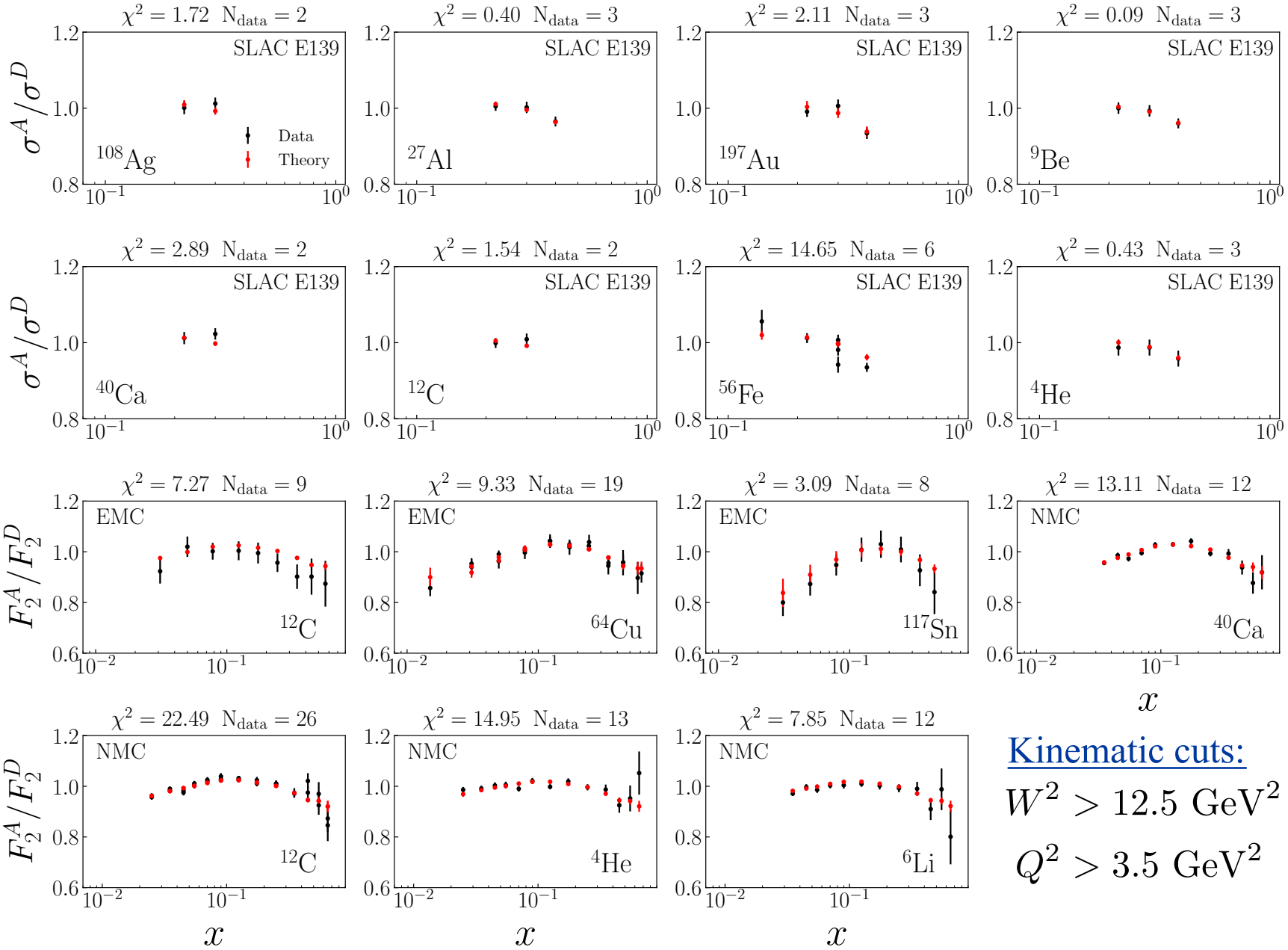
Normalization to satisfy momentum sum rule

Parameter Optimization



- Data resampling to construct pseudo-data sets: $\widetilde{\mathcal{D}}_i^e = \mathcal{D}_i^e + R_i \alpha_i^e$
- Prevent overfitting with cross validation:
  - → Partition pseudo-data into training and validation sets
  - → Train parameters on training set
  - → Early stopping: end optimization at best fit of validation set
- Chi-squared minimization procedure:
  - → Define cost function as chi-squared: $$\chi^2 = \sum_e^{N_{exp}} \sum_i^{N_{data}} \frac{(\widetilde{\mathcal{D}}_i^e - T_i)^2}{(\alpha_i^e)^2}$$
  - → Neural network parameters trained with gradient descent

Results of NC DIS Analysis – Data vs Theory

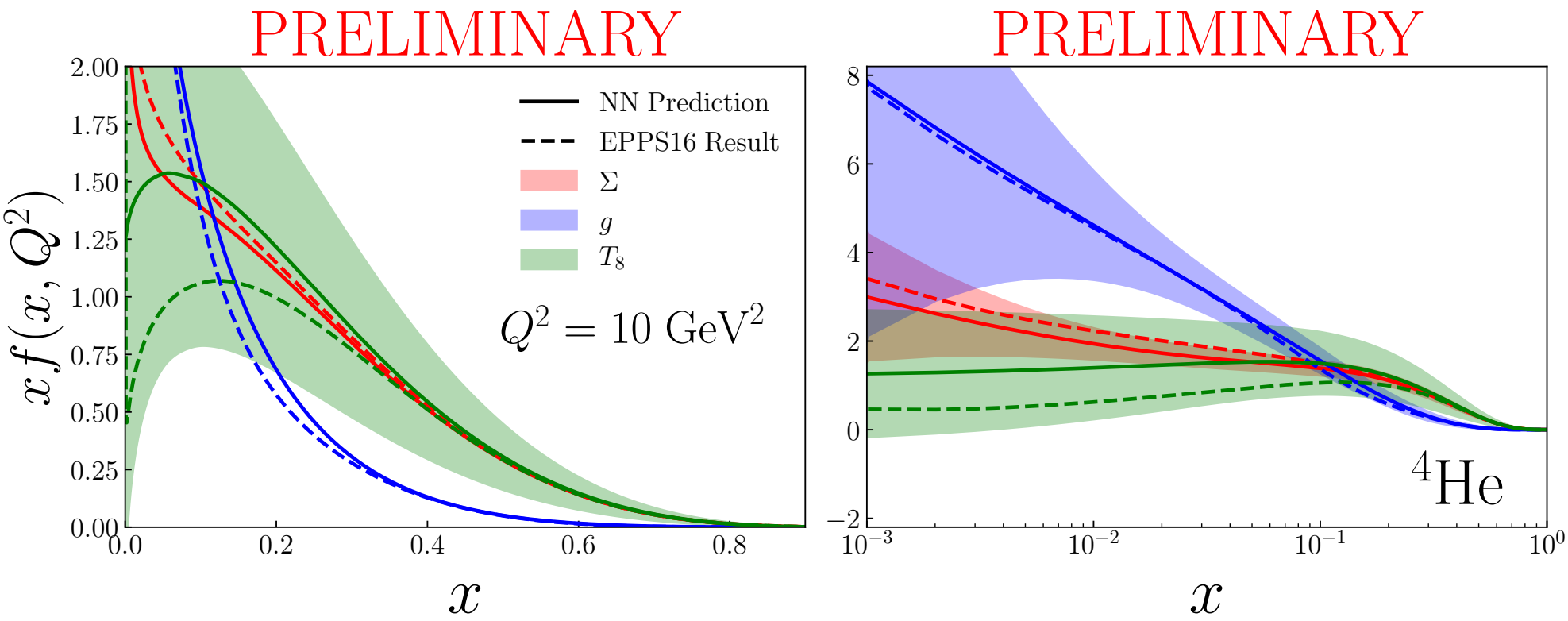


Kinematic cuts:

$$W^2 > 12.5 \text{ GeV}^2$$

$$Q^2 > 3.5 \text{ GeV}^2$$

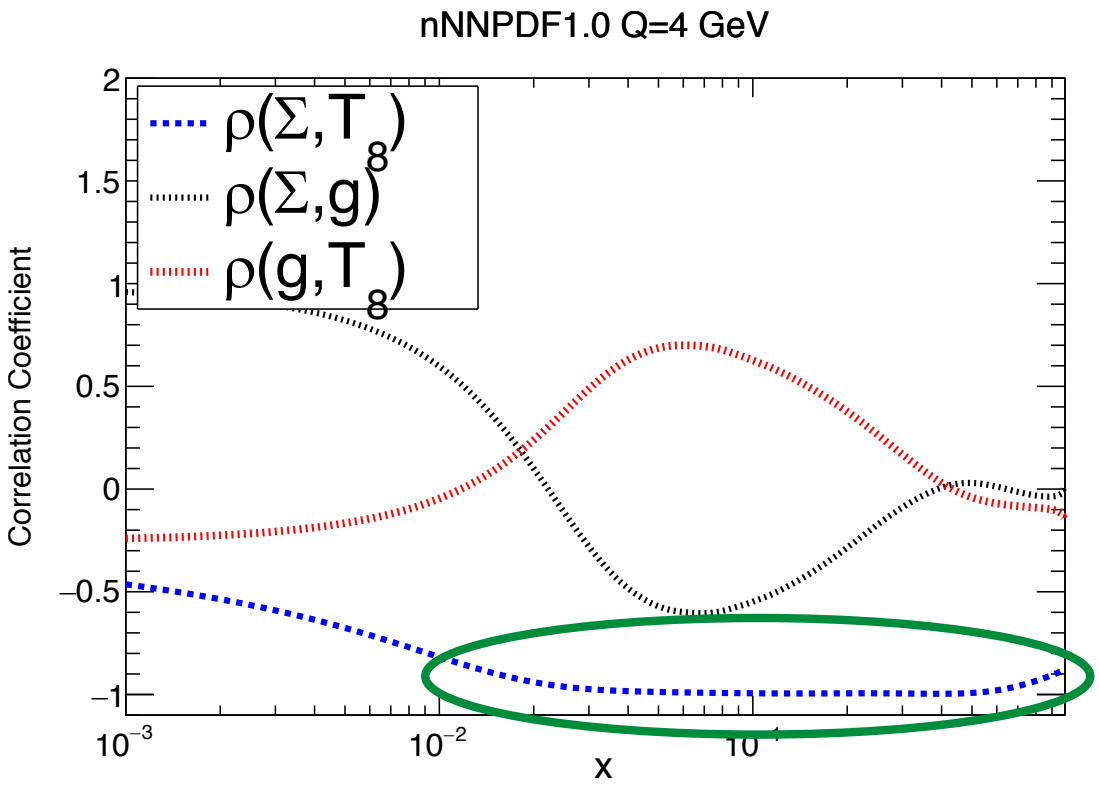
Results of NC DIS Analysis – Nuclear PDFs



- Excellent agreement with EPPS16 results within uncertainties
- Singlet and octet distributions strongly anti-correlated in data region

  → Observables only sensitive to linear combination (sum) of the two distributions
- Mild *A*-dependence

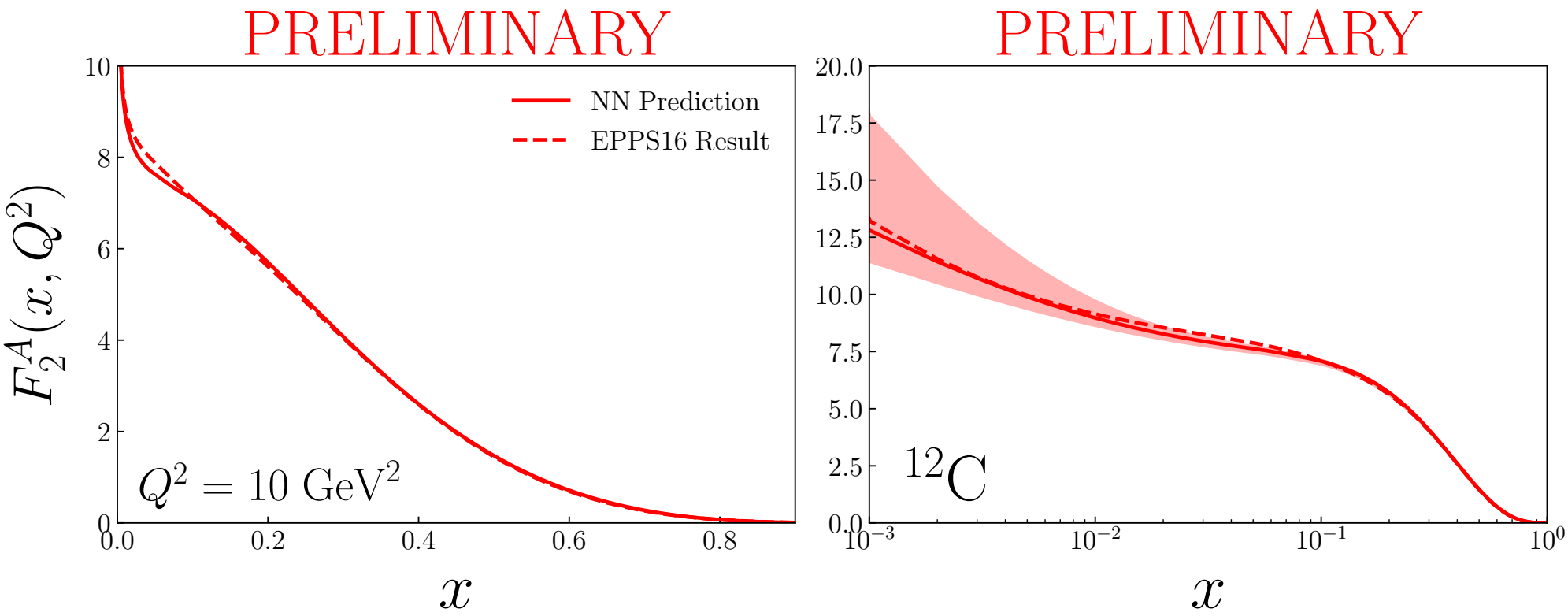
Results of NC DIS Analysis – Nuclear PDFs



- Excellent agreement with EPPS16 results within uncertainties
- Singlet and octet distributions strongly anti-correlated in data region

  → Observables only sensitive to linear combination (sum) of the two distributions
- Mild *A*-dependence

Results of NC DIS Analysis – Nuclear PDFs

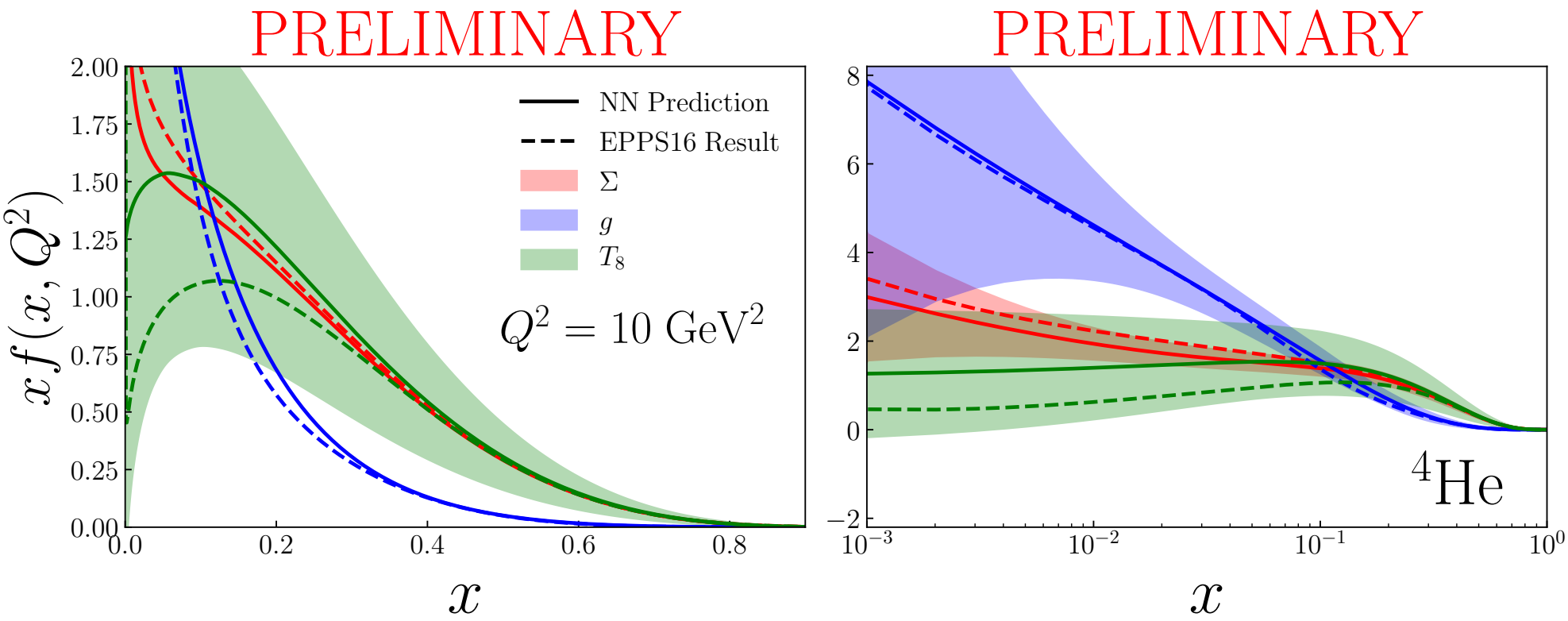


- Excellent agreement with EPPS16 results within uncertainties
- Singlet and octet distributions strongly anti-correlated in data region

→ Observables only sensitive to linear combination (sum) of the two distributions

- Mild *A*-dependence

Results of NC DIS Analysis – Nuclear PDFs

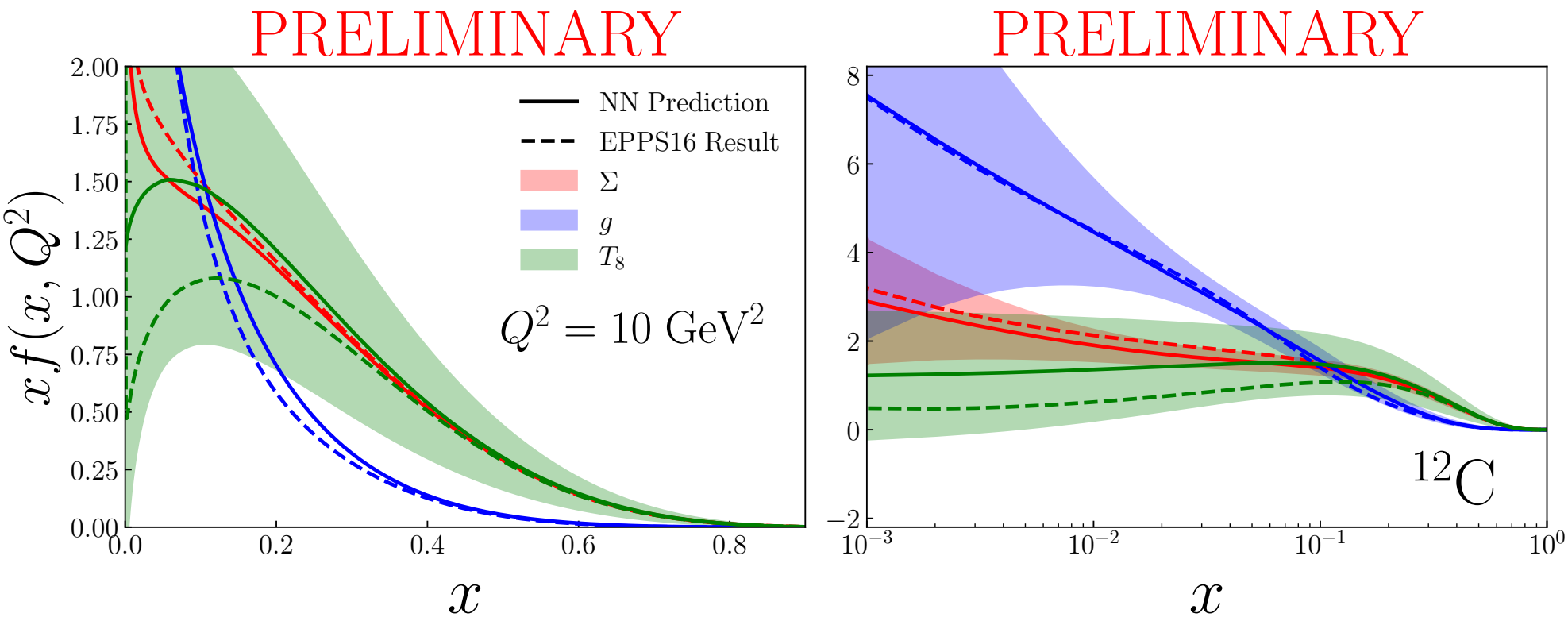


- Excellent agreement with EPPS16 results within uncertainties
- Singlet and octet distributions strongly anti-correlated in data region

  → Observables only sensitive to linear combination (sum) of the two distributions

- Mild A-dependence

Results of NC DIS Analysis – Nuclear PDFs

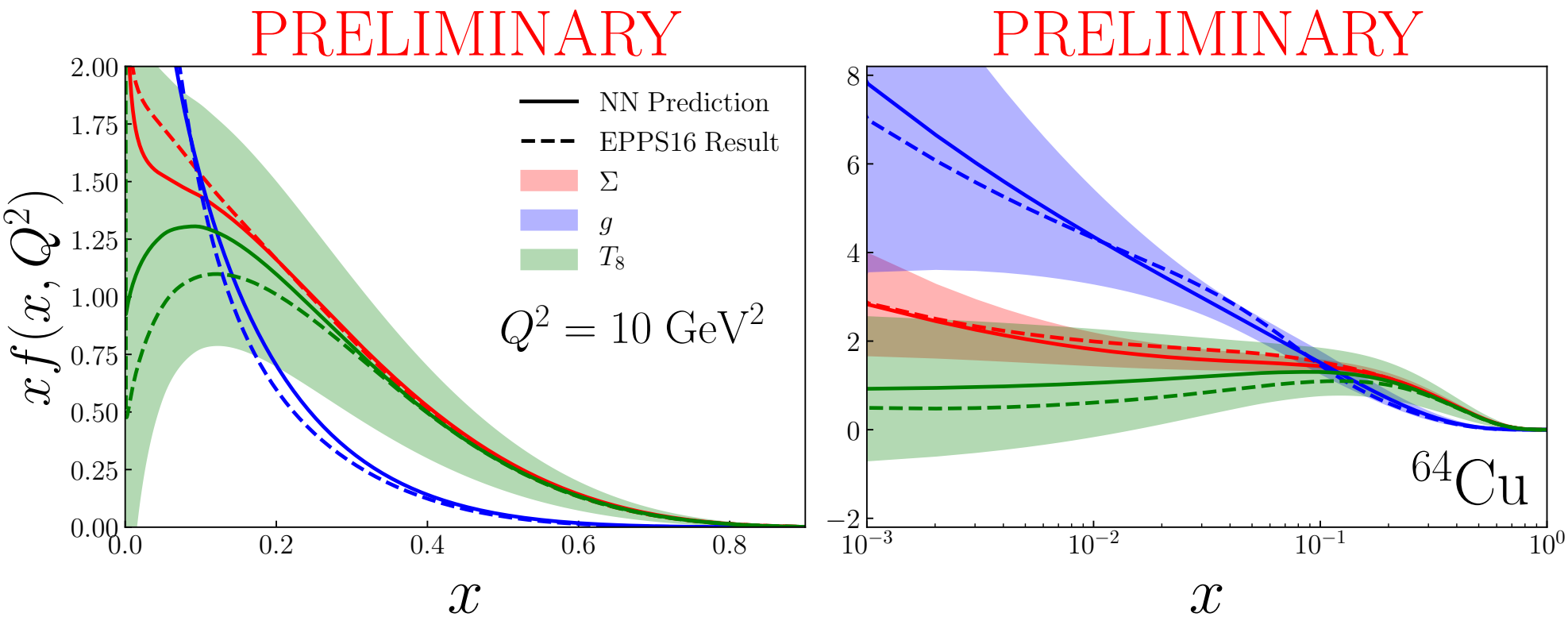


- Excellent agreement with EPPS16 results within uncertainties
- Singlet and octet distributions strongly anti-correlated in data region

  → Observables only sensitive to linear combination (sum) of the two distributions

- Mild A-dependence

Results of NC DIS Analysis – Nuclear PDFs

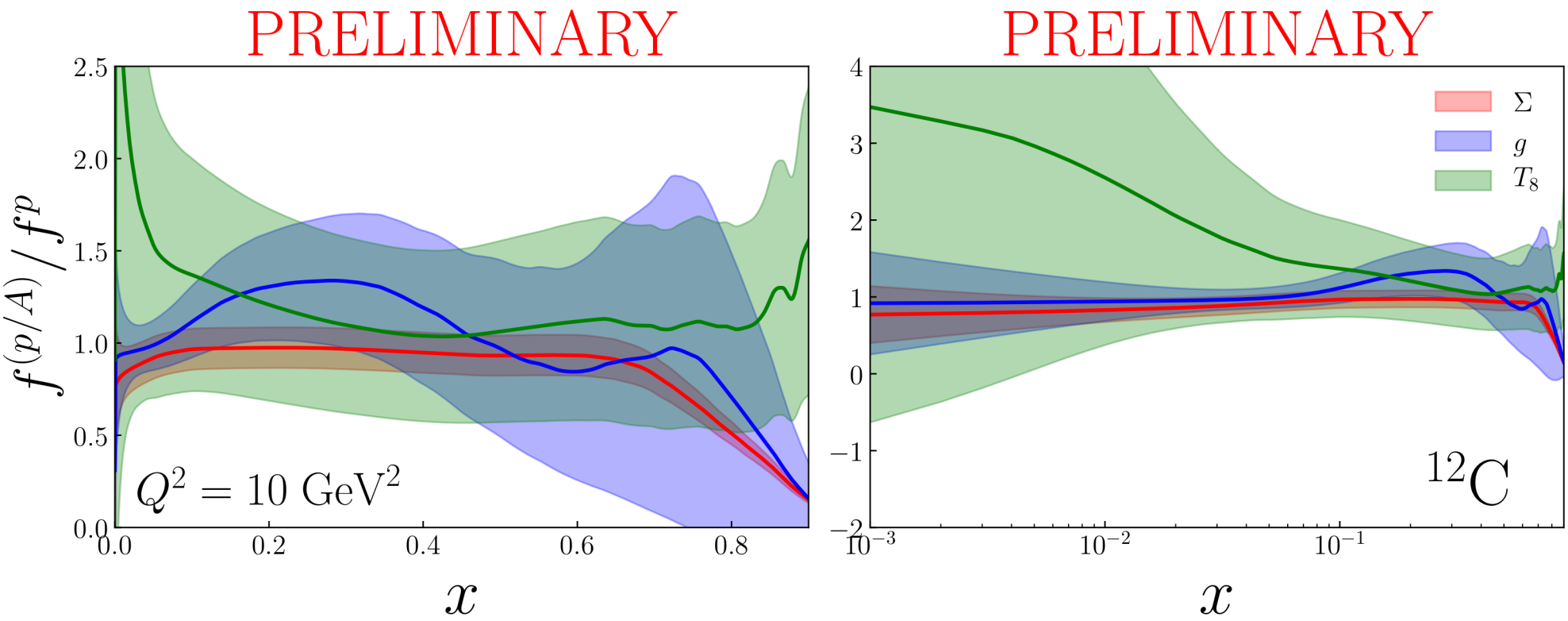


- Excellent agreement with EPPS16 results within uncertainties
- Singlet and octet distributions strongly anti-correlated in data region

  → Observables only sensitive to linear combination (sum) of the two distributions

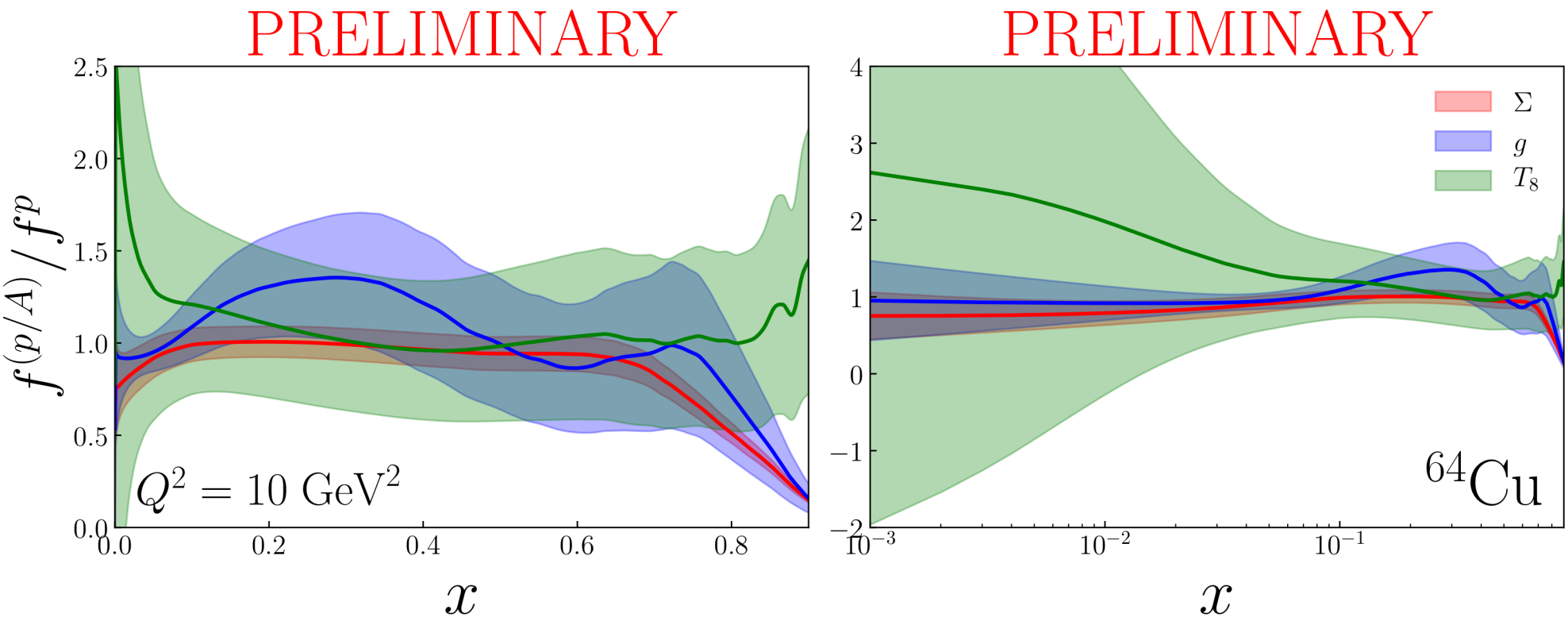
- Mild *A*-dependence

Results of NC DIS Analysis – Nuclear PDFs



- Ratio of nuclear PDFs with NNPDF3.1 (positive-definite) proton set
  → Both extracted using similar theoretical assumptions
- Suggests significant impact on gluon distribution in EMC region
  → Need better constraints from gluon sensitive observables

Results of NC DIS Analysis – Nuclear PDFs



- Ratio of nuclear PDFs with NNPDF3.1 (positive-definite) proton set
  → Both extracted using similar theoretical assumptions
- Suggests significant impact on gluon distribution in EMC region
  → Need better constraints from gluon sensitive observables

Summary and Outlook

- Machine learning + Monte Carlo methods are important for robust extractions of nonperturbative functions and their uncertainties

  → Necessary for future global QCD studies that will contain large data sets and many fit parameters

- New approach to global nuclear PDF studies are being developed:

  → Use of neural networks and machine learning tools to minimize bias

- Preliminary extractions of PDFs from NC DIS show good agreement with previous analyses

  → Still need additional DIS data sets not currently implemented (HERMES, Fermilab)

- Available NC DIS data not sensitive to separation of singlet and octet distributions

  → Inclusion of additional observables (pA collisions, CC DIS, Drell-Yan) for flavor separation and uncertainty reduction is needed