Self Organizing Maps Parameterization of Parton Distribution Functions

Simonetta Liuti
University of Virginia

XXIII International DIS Workshop

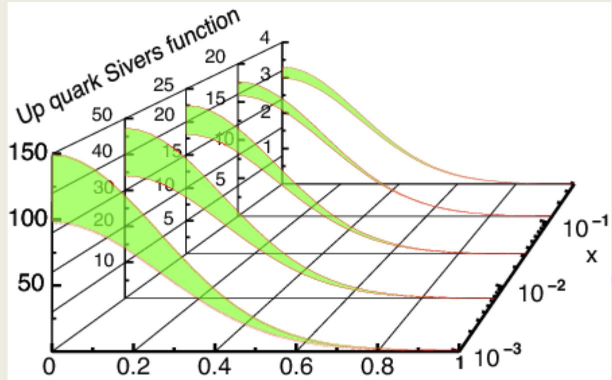
Dallas, Texas, April 26th-May1st 2015

Outline

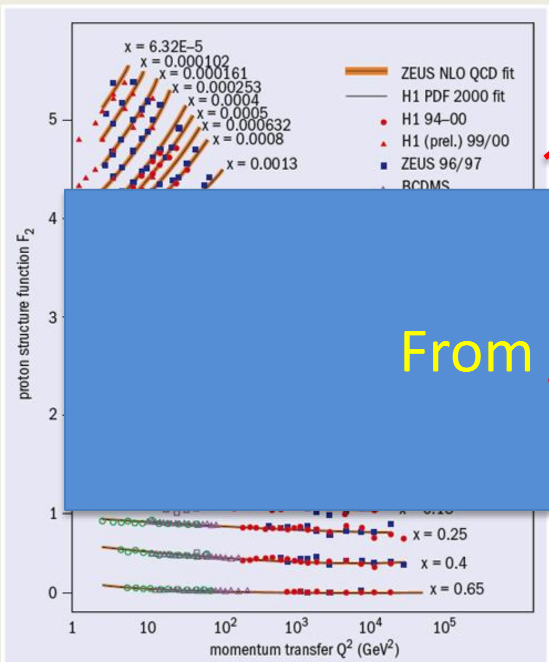
- Introduction
- Artificial Neural Networks in HEP/Nuclear Data Analyses
- Self Organizing Maps (SOMs) Algorithm
- SOMPDFs
- Quantitative example: d/u ratio at large x
- Conclusions/Outlook/Extension to GPDs, TMDs…

Issues in dealing with an increasingly complicated and diverse set of observables

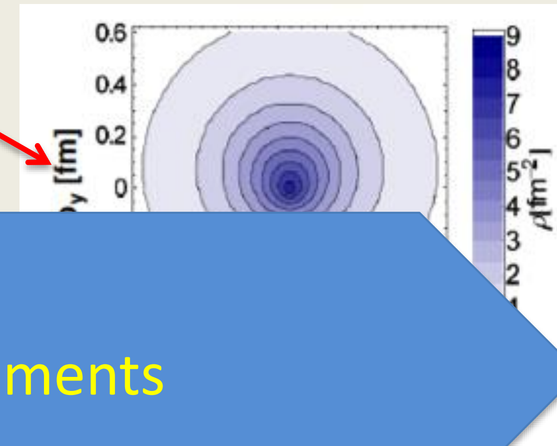
TMDs



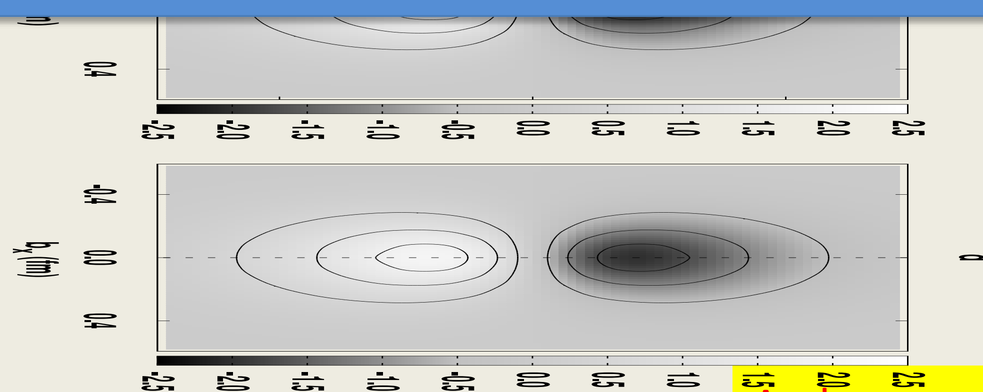
PDFs



GPDs



From inclusive to exclusive measurements



O. Gonzalez-Hernandez, S.L.

And more…
Fragmentation Functions (FFs)
Fracture Functions (FFs)…

✓ Conventional models give interpretations in terms of the microscopic properties of the theory (focus on the behavior of individual particles)

Parameterizations depend on the analytical form of the PDFs

$$\boxed{f_i(x, Q_o^2; A_i, b_i \ldots) = A_i\, x^{b_i}(1 - x)^{c_i}(1 + d_i x + e_i x^2 + \ldots)}$$

In a nutshell:

1) One finds the best-fit values of parameters.

2) The uncertainty is determined in most cases with the Hessian method.

Conventional methods' problem: fits to data depend on the specific functional form

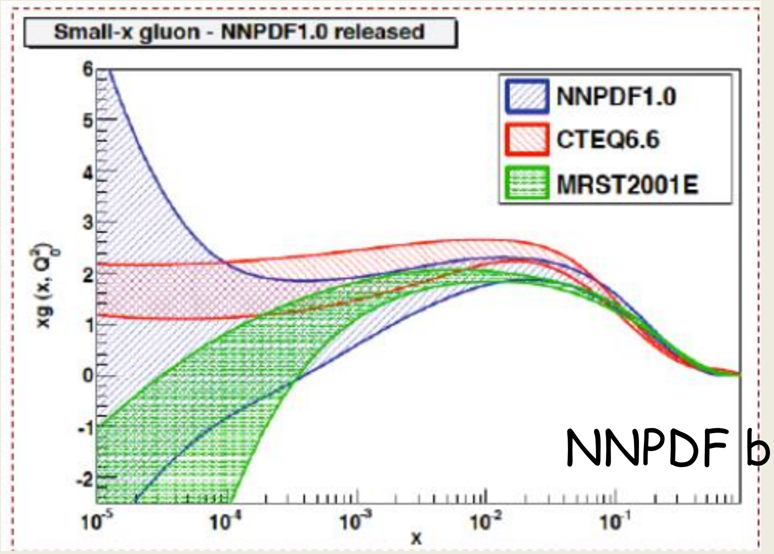
Initial bias!

- To overcome this S. Forte et al. introduced an Artificial Neural Network based approach (NNPDF)

- Attacking the problem from a different perspective: study the behavior of multi-particle systems as they evolve from a large and varied number of initial conditions: this goal is at reach with HPC

- However ANN approach has an inherent problem:

renouncing to a specific form makes extrapolation difficult



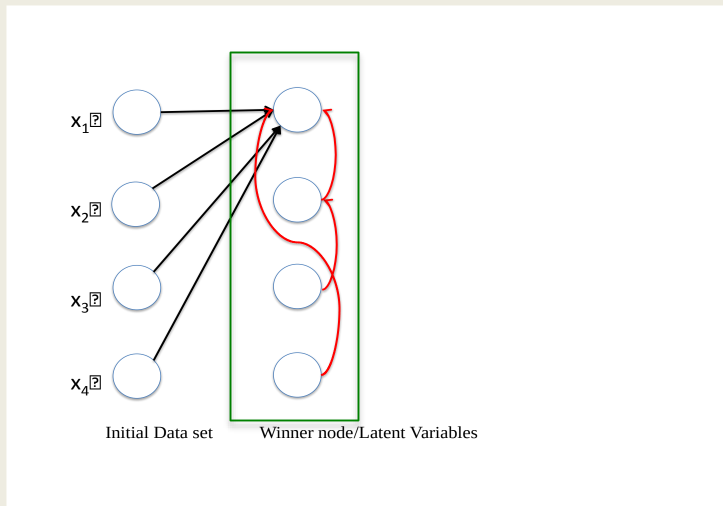
NNPDF before LHC data

Of fundamental importance for TMD,GPD analysis!
If data are missing it is not possible to determine output!

Is there a way of keeping “the best of both worlds”?

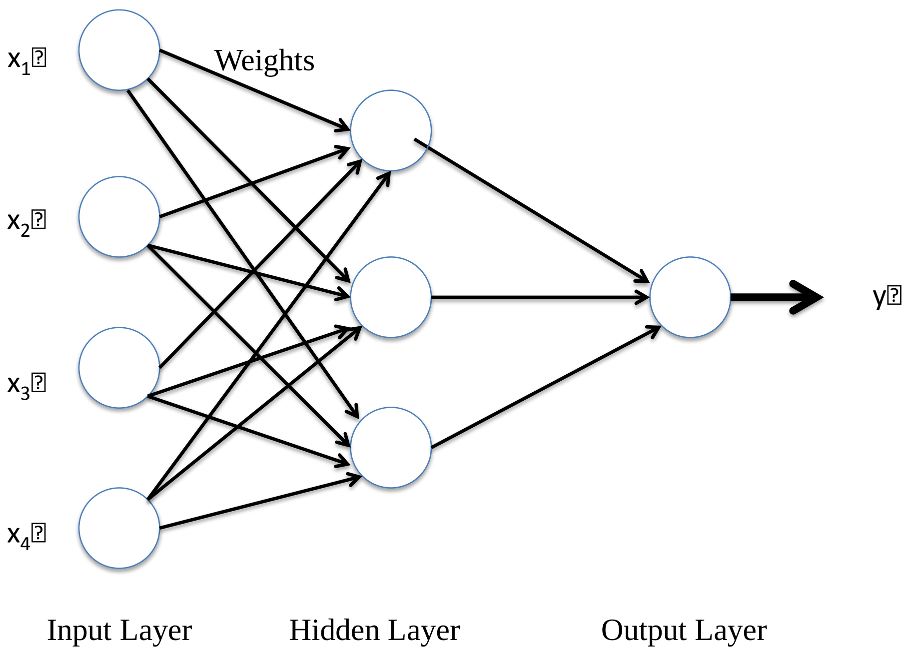
In *J. Carnahan, H. Honkanen, S.Liuti, Y. Loitiere, P. Reynolds, Phys Rev D79, 034022 (2009)* we came to the conclusion that one must improve on the ANN type algorithm!

Self Organizing Maps (SOMs) NN based on ``Unsupervised Learning”



No a priori examples are given.
The NN learns by finding how the data cluster or self-organize

Artificial Neural Networks in HEP/Nuclear Data Analyses



Back propagation/supervised learning

1. Take the output from the network

2. Compare it to the real data values

3. Calculate how wrong the network was (error= how wrong the weights were)

4. Use this information to calculate the partial derivatives in the parameters/weights which are necessary to minimize the cost

NNPDFs…(S.Forte, et al.)

http://nnpdf.hepforge.org/html/GenStr.html

- Monte Carlo generation of data replicas

→ no need for linear propagation of errors
→ possibility to test for non Gaussian behaviour in fitted PDFs

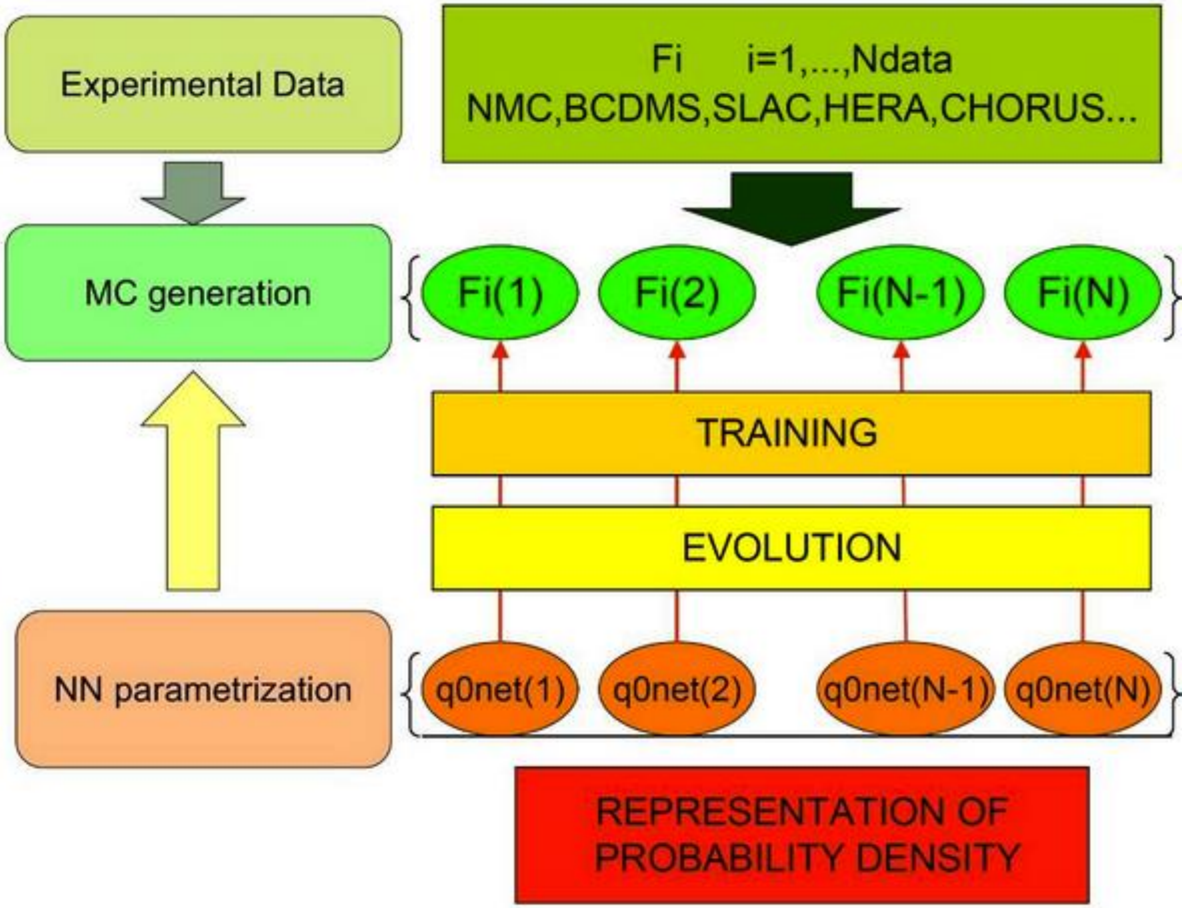
- Neural Networks parametrization of PDFs

→ 7 independent PDFs, 259 parameters
→ unbiased parametrization

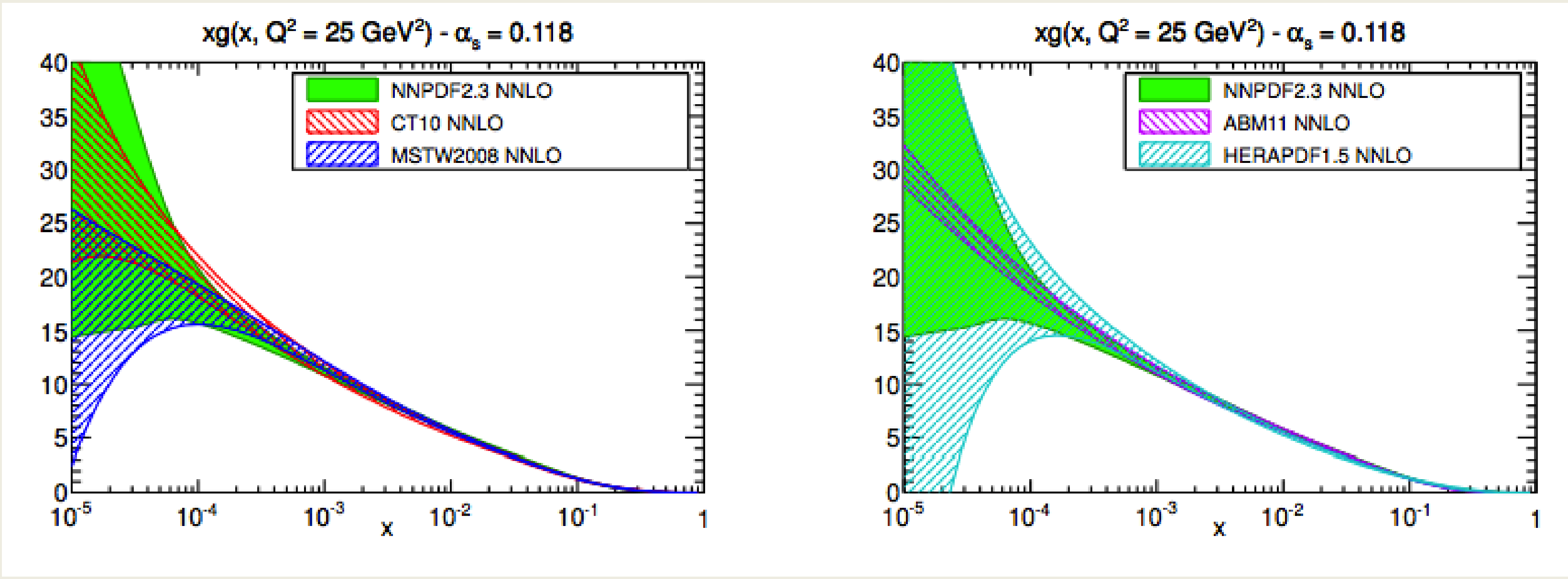
- Evolution using DGLAP equations

- Genetic Algorithm's training of neural networks parameters

- Analysis of χ^2 distributions



NNPDF including LHC data, JHEP(2012)



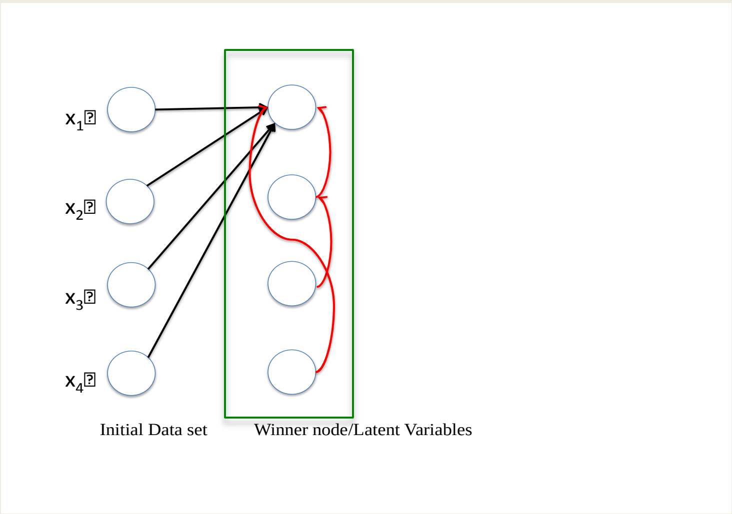
New issues, new benchmarks discussed at this meeting address:

1) Possible non-Gaussian behavior of data; error treatment (H12000,…)
2) Study of variations from using different data sets and different methods (Alekhin,…)
3) Comparison of parameterizations where fits where error treatment is the same but methods are different
4) …

What is the ideal flexibility of the fitting functional forms?
What is the impact of such flexibility on the error determination?

→ SOMs are ideal to study the impact of the different fit variations!

Self Organizing Maps (SOMs)



The various nodes form a topologically ordered map during the learning process.

The learning process is unsupervised ➔ no "correct response" reference vector is needed.

The goal is to minimize the cost function by similarity relations, or by finding how the data cluster or self-organize

The nodes are decoders of the input signals -- can be used for pattern recognition.

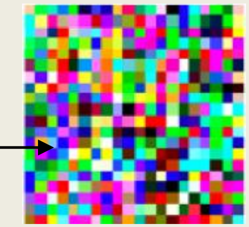
SOMs Algorithm

Each cell (neuron) is sensitized to a different domain of vectors:
cell acts as decoder of domain

V_i=(R,B,G)

Initialization → Input vector of dimension "**n**" associated to cell "**i**":

$$V_i = \left[v_i^{(1)}, \ldots, v_i^{(n)}\right]$$

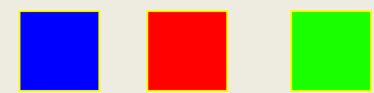


V_i is given spatial coordinates that define the geometry/topology of a 2D map

Training → Input data:

$$x = \left[\xi^{(1)}, \ldots, \xi^{(n)}\right]$$

isomorphic



x compared to V_i 's with "similarity" metric(L1):

$$|| \, x - m_i \, ||$$

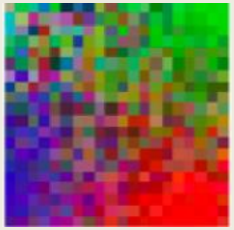
(Aggawal et al., 2000)

Location of best match "winner" gives location of response
(active cell, all others are passive)

Learning (updating) → cells $\mathbf{V_i}$ that are close up to a certain distance activate each other to "learn" from **x**

Learning:

Map cells, V_i, that are close to “winner neuron” activate each other to “learn” from x



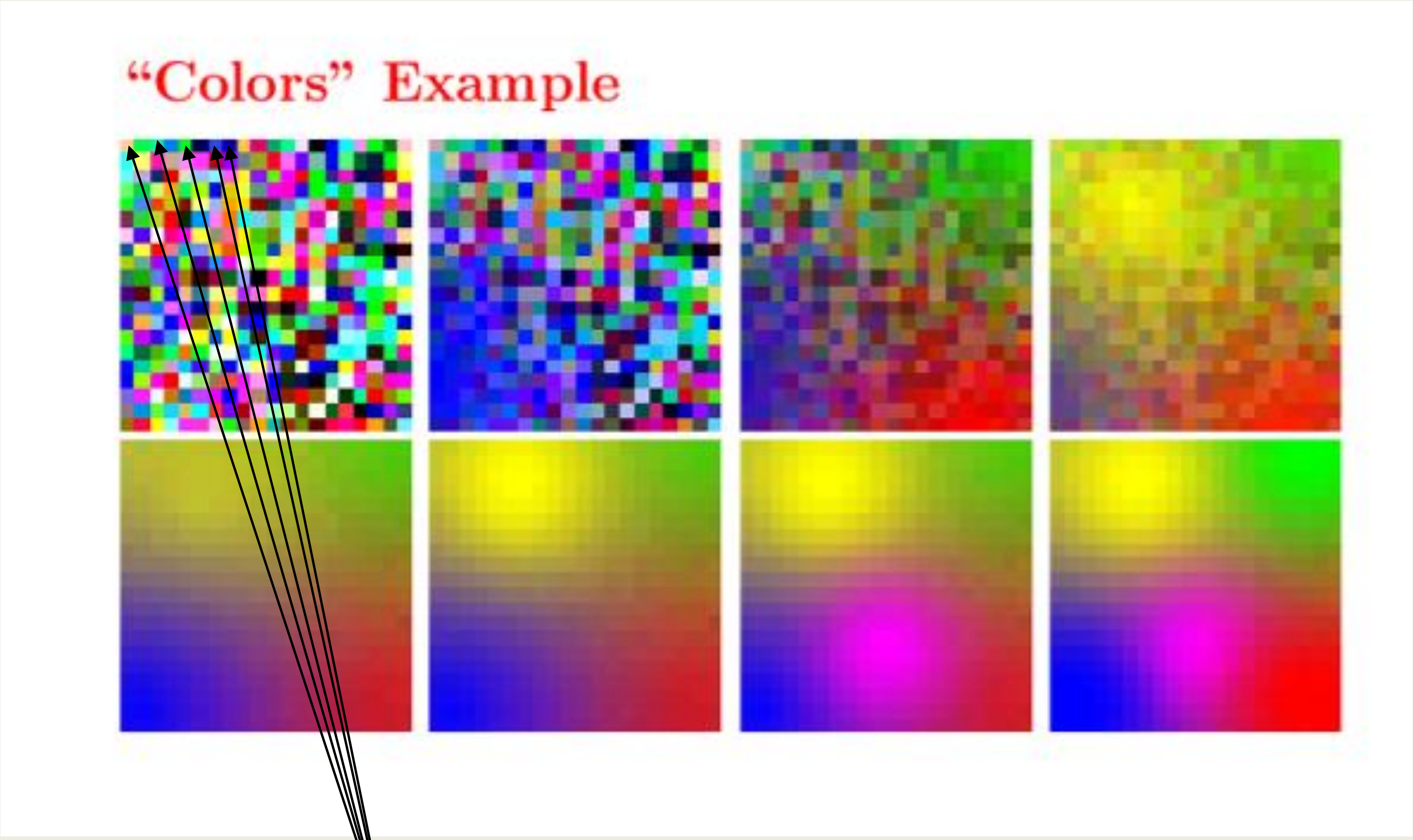
$$\boxed{V_i(n+1) = V_i(n) + h_{ci}(n)\left[x(n) - V_i(n)\right]}$$

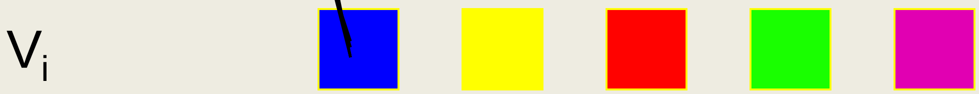
iteration number

$$h_{ci}(n) = f(\|r_c - r_i\|) \equiv \alpha(n)\exp\left(\frac{-\|r_c - r_i\|^2}{2\sigma^2(n)}\right)$$

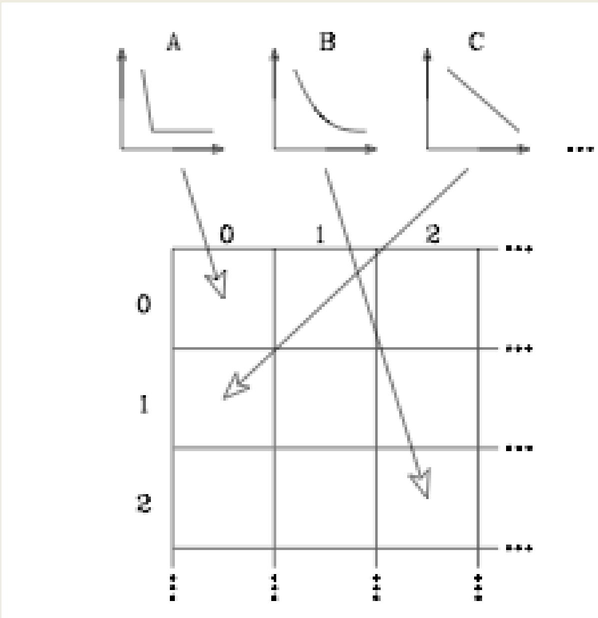
neighborhood function decreases with “n” and “distance”

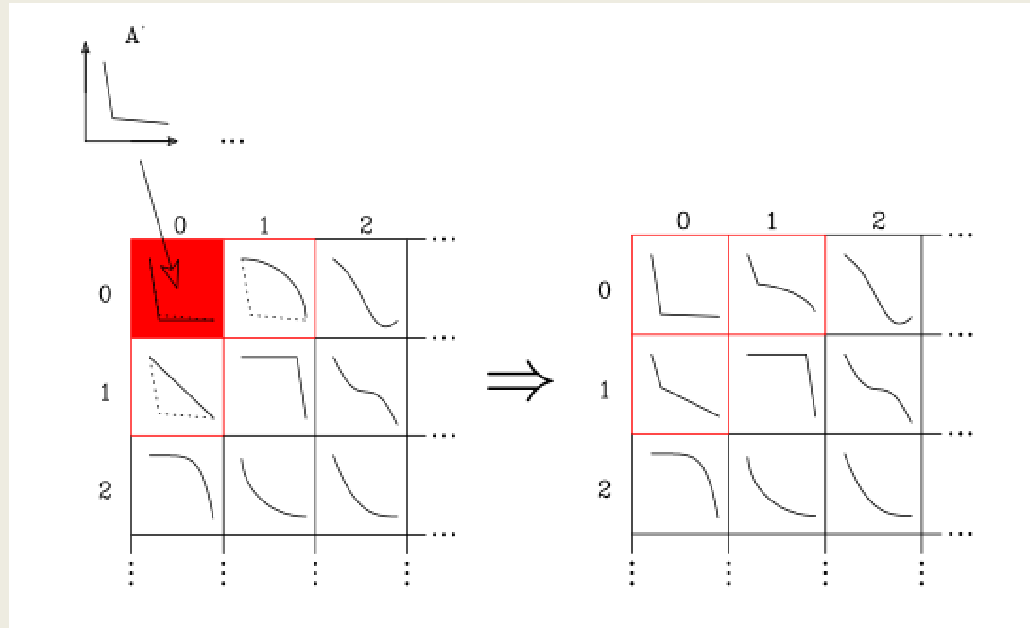
Map representation of 5 initial samples: blue, yellow, red, green, magenta





Simple Functions Example





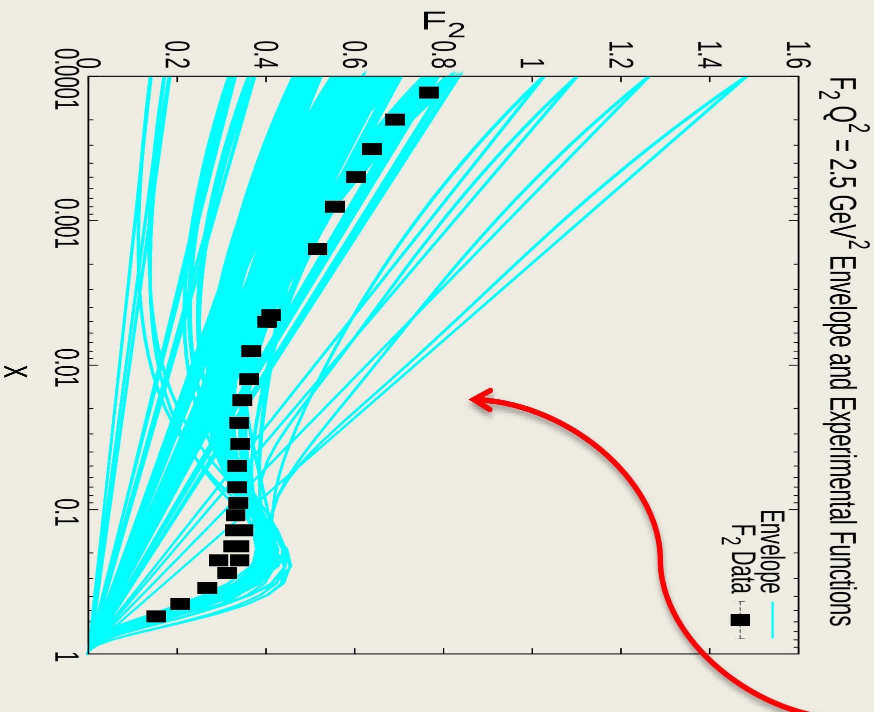
<u>Initialization</u>: functions are placed on map

<u>Training</u>: "winner" node is selected,
<u>Learning</u>: adjacent nodes readjust according to similarity criterion

Final Step : clusters of similar functions from input data get distributed on the map

Now on to PDFs…

Initialization: a set of database/input PDFs is obtained selecting at random from existing PDF sets and varying their parameters according to a pre-defined procedure.



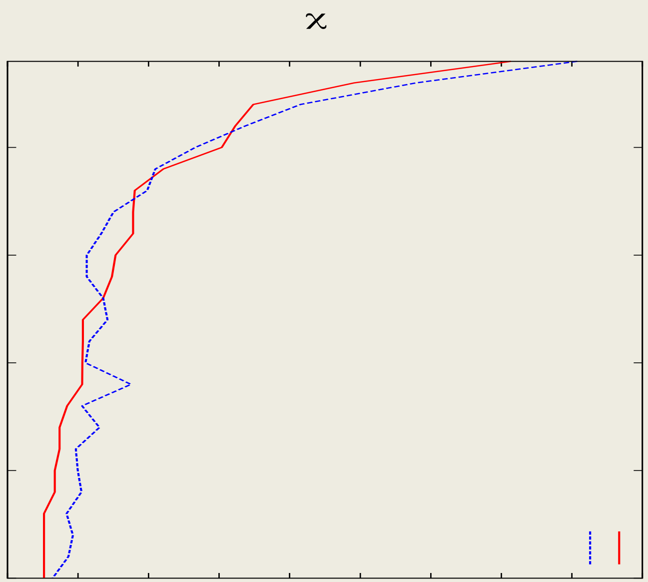
Training: A subset of input PDFs (envelope) is used to train the map.

Learning:The similarity is tested by comparing the PDFs at given (x,Q^2) values. The new

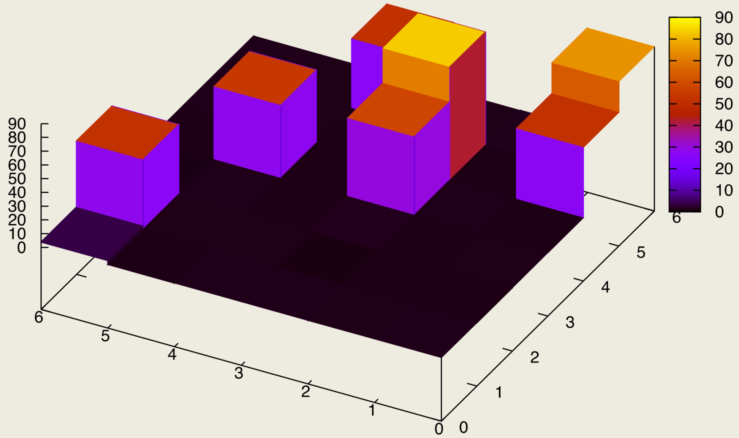
map PDFs are obtained by averaging the neighboring PDFs with the “winner” PDFs.)

χ^2 minimization through genetic algorithm

- Once the first map is trained, the χ^2 per map cell is calculated.
- We take a subset of PDFs that have the best χ^2 from the map and form a new initialization set including them.
- We train a new map, calculate the χ^2 per map cell, and repeat the cycle.
- We iterate until the χ^2 stops varying (stopping criterion).



χ^2



Error Analysis

- Treatment of experimental error is complicated because of incompatibility of various experimental χ^2.
- Treatment of theoretical error is complicated because they are not well known, and their correlations are not well known.
- In our approach we performed the theoretical error evaluation with the Lagrange multiplier method and using the generated PDFs as a statistical ensemble

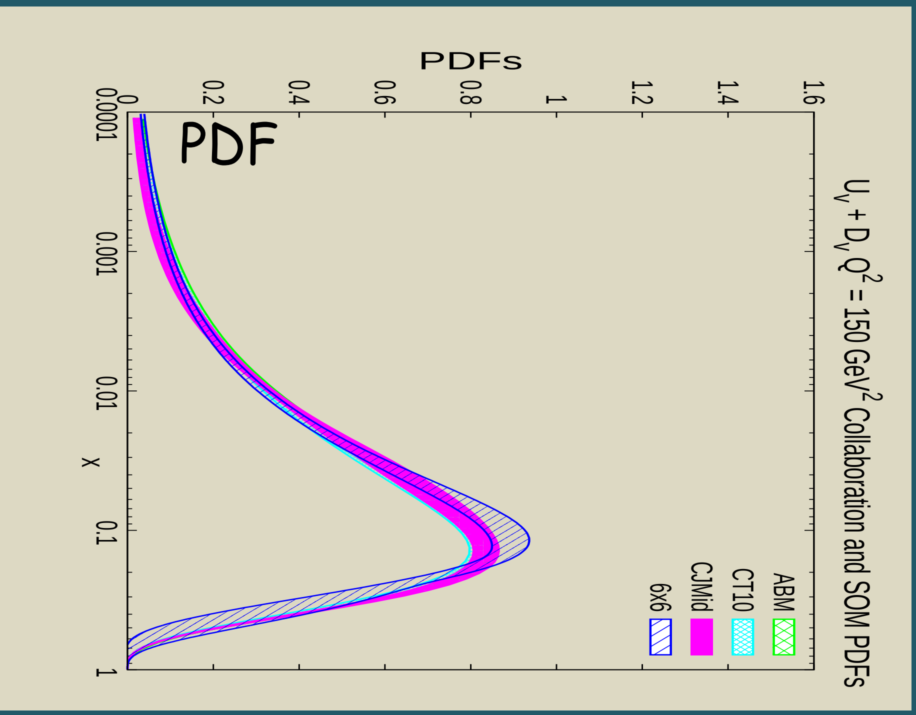
Advantages over NNPDFs

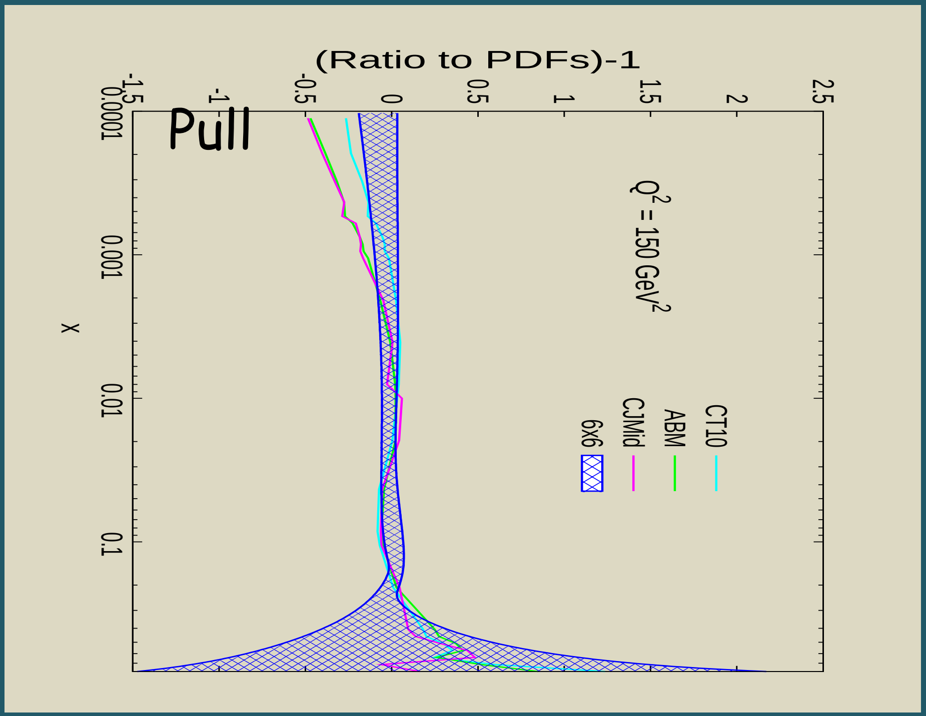
Clustering properties: generic ANNs do not keep track of inter-connections/correlations of data at the various stages of the network training

Advantages over “conventional” PDFs

Similarly to NNPDFs we eliminate the bias due to the initial parametric form

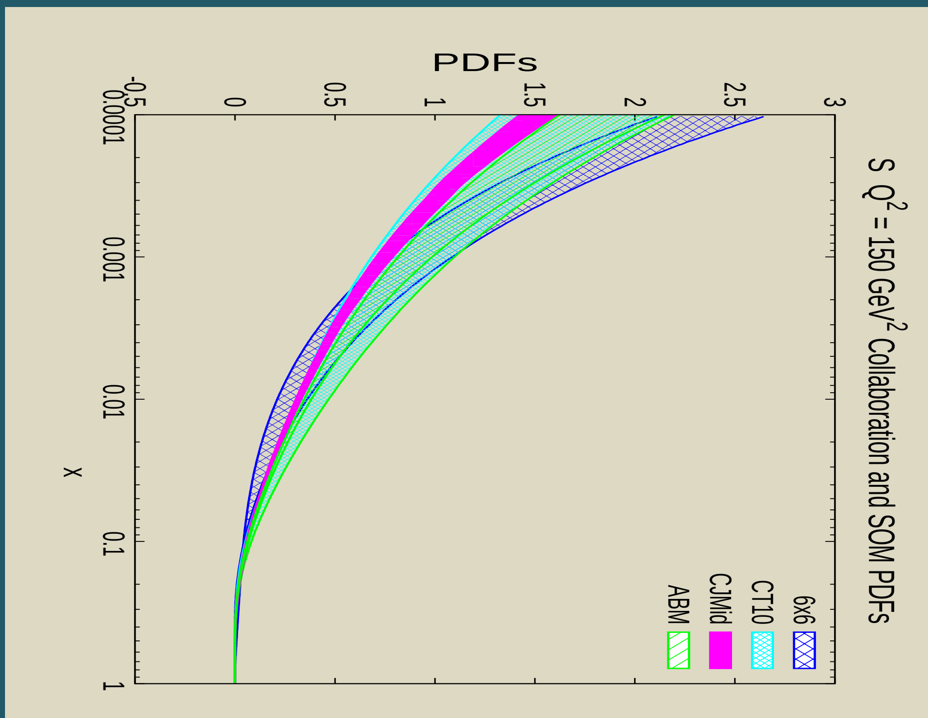
Valence

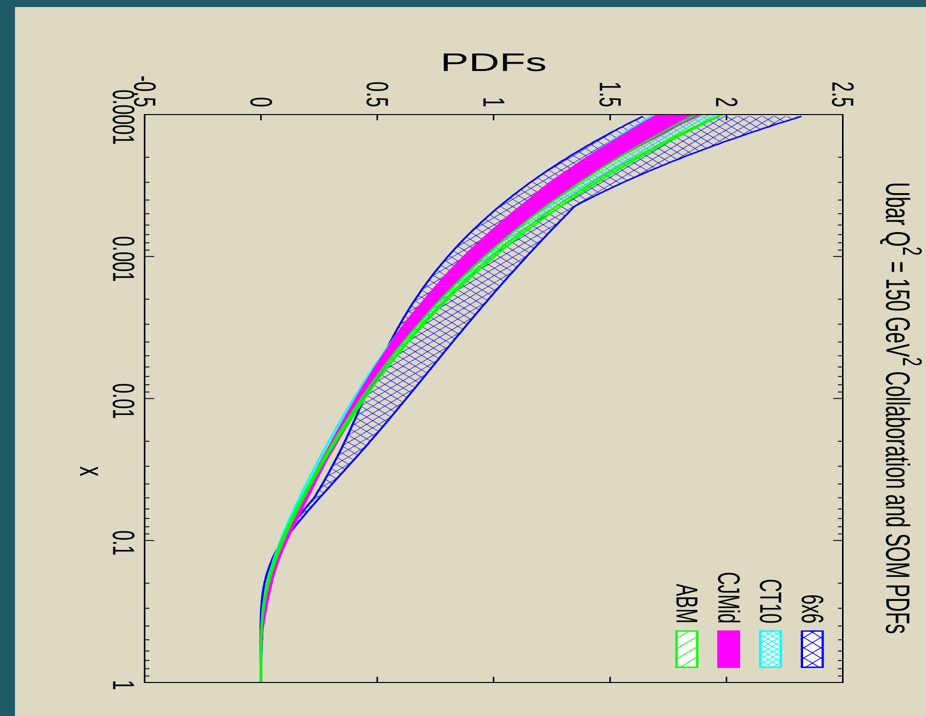




E. Askanazi, K. Holcomb, S. Liuti, J. Phys. G 42, no. 3, 034030 (2015) [arXiv:1411.2487 [hep-ph]].

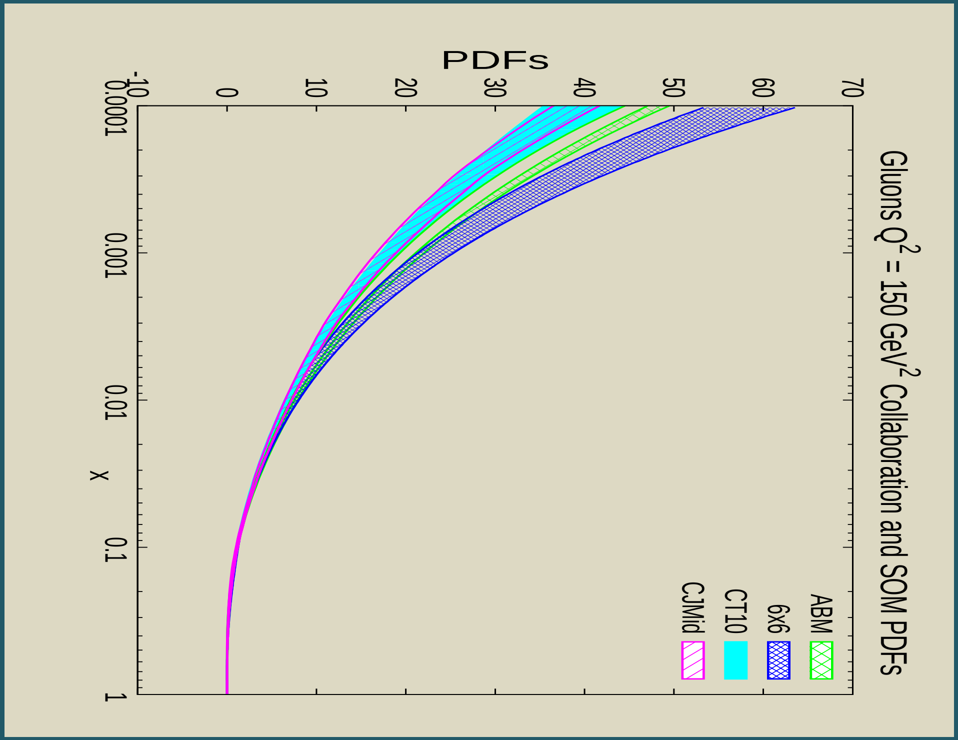
Strange, ubar, dbar

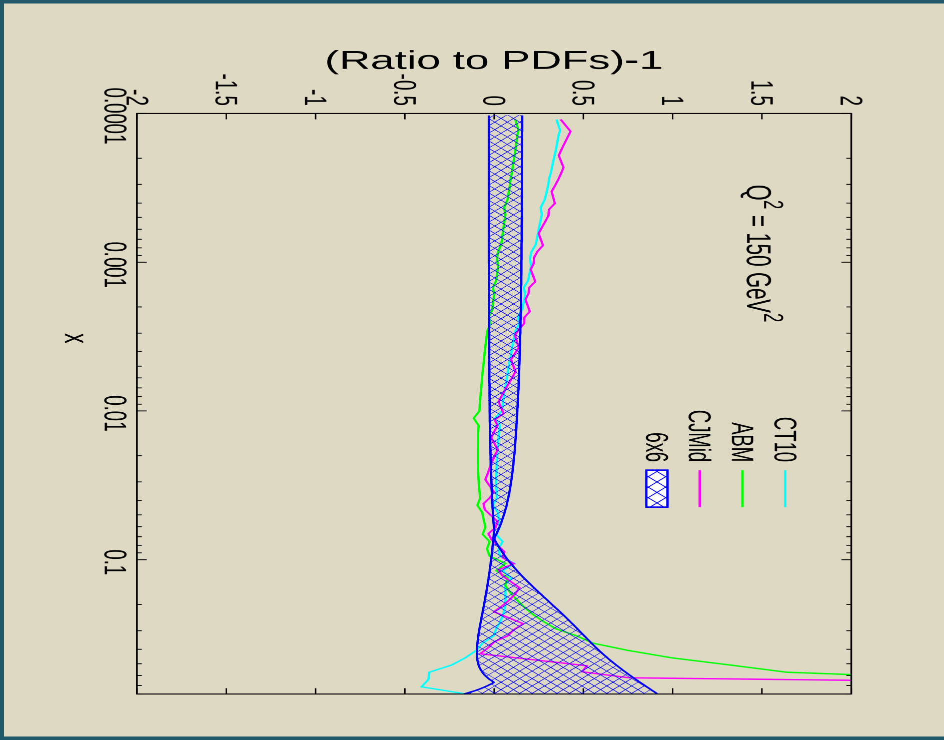




E. Askanazi, K. Holcomb, S. Liuti, J. Phys. G 42, no. 3, 034030 (2015) [arXiv:1411.2487 [hep-ph]].

Gluons



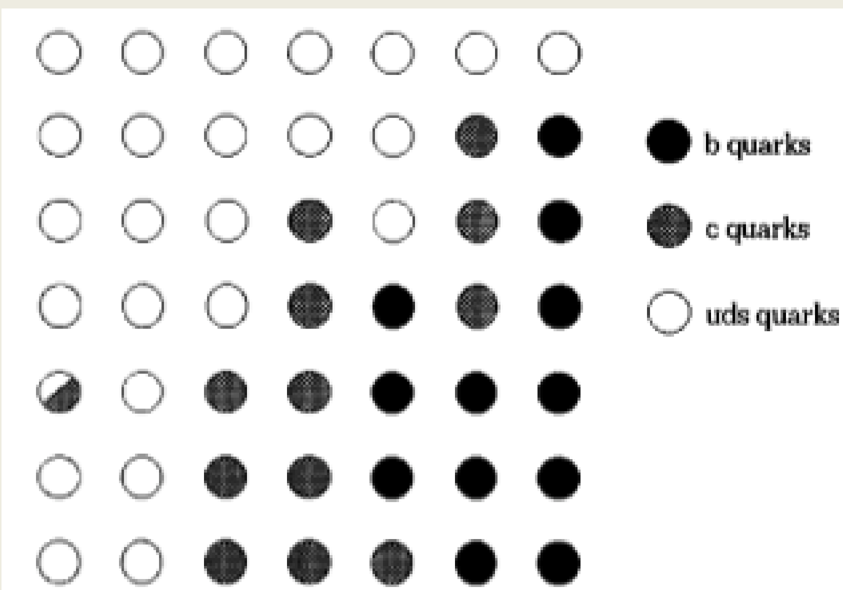


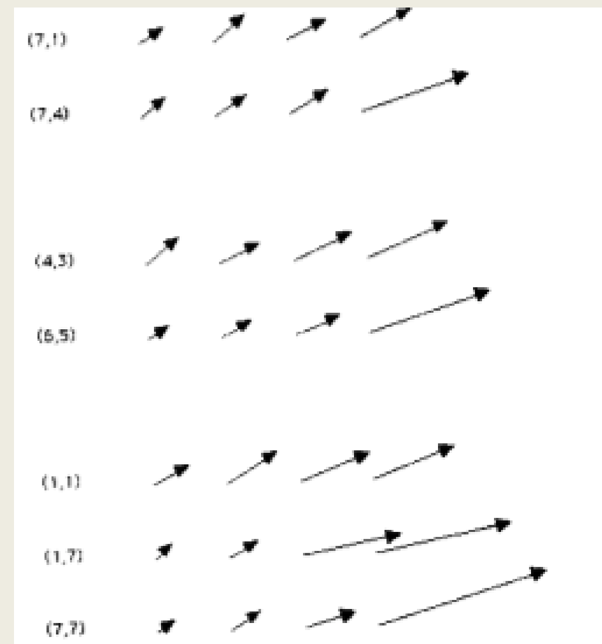
E. Askanazi, K. Holcomb, S. Liuti, J. Phys. G 42, no. 3, 034030 (2015) [arXiv:1411.2487 [hep-ph]].

SOMs can do more than this:

- SOMs differently from standard ANN methods are "unsupervised": <u>they find similarities in the input data without a training target.</u>
- They have been used in theoretical physics approaches to critical phenomena, to the study of complex networks, and in general for the study of high dimensional non-linear data (e.g. Der, Hermann, Phys.Rev.E (1994), Guimera et al., Phys. Rev.E (2003))
- Our final goal: use SOMs to study multidimensional parton distributions/multiparton correlations (GPDs…)

Example





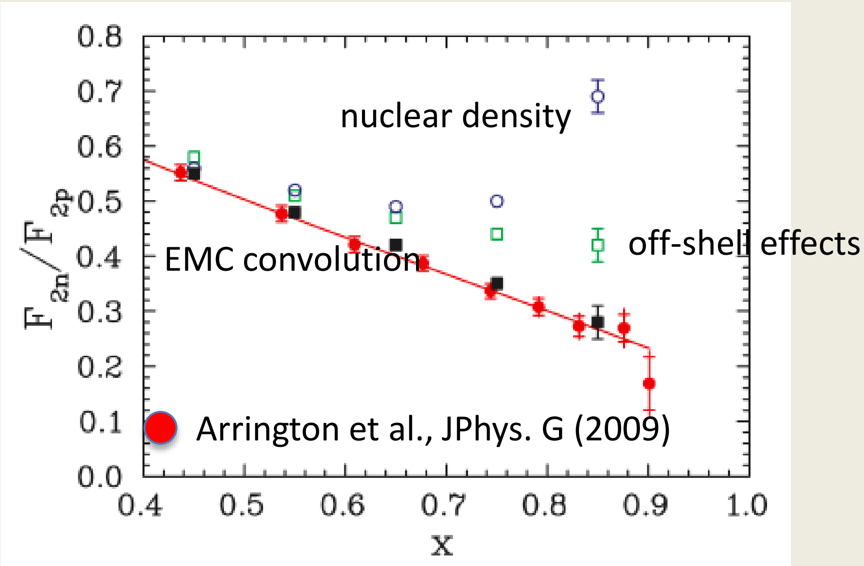
b quarks

c quarks

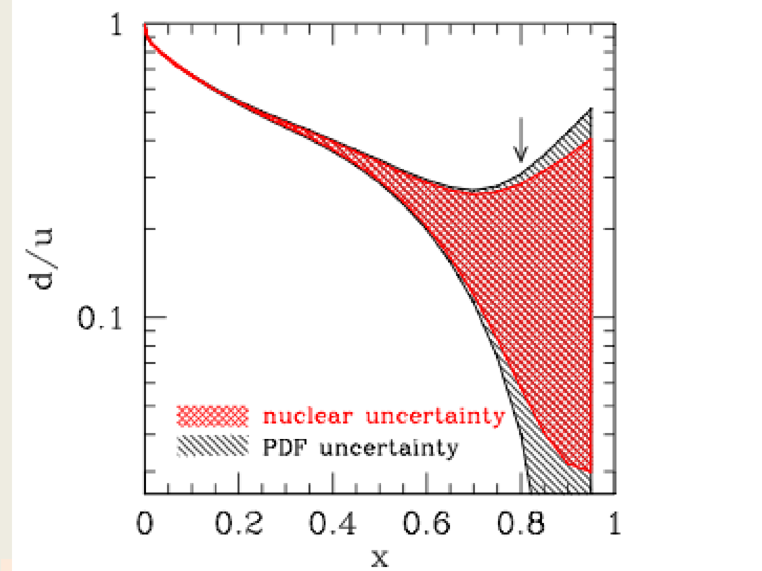
uds quarks

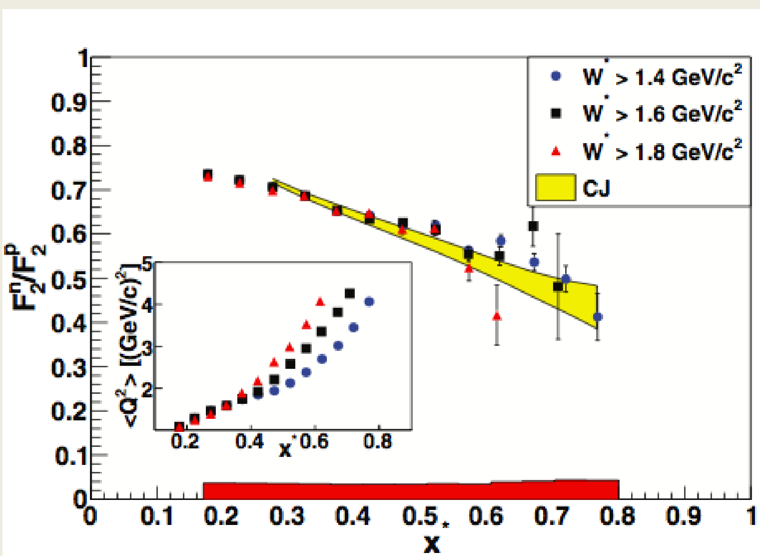
Lonnblad, Peterson, Pi, Computer Physics Comm. 1991

Large x → d/u ratio



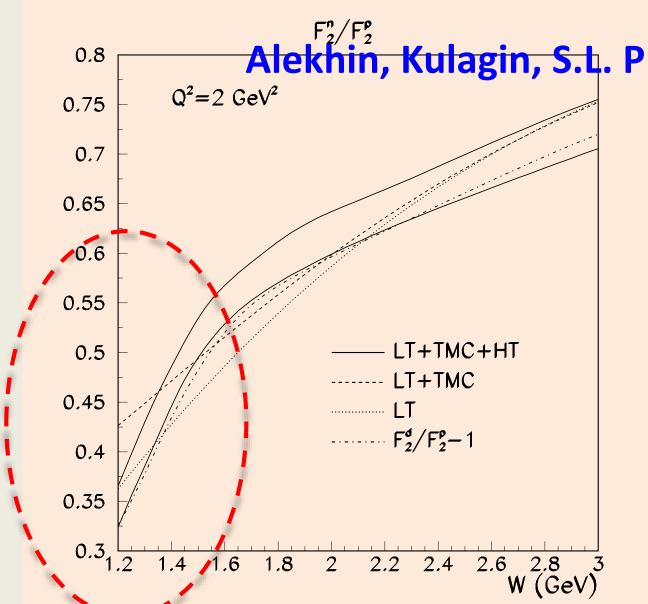
Accardi PHYSICAL REVIEW D 84, 014008 (2011)





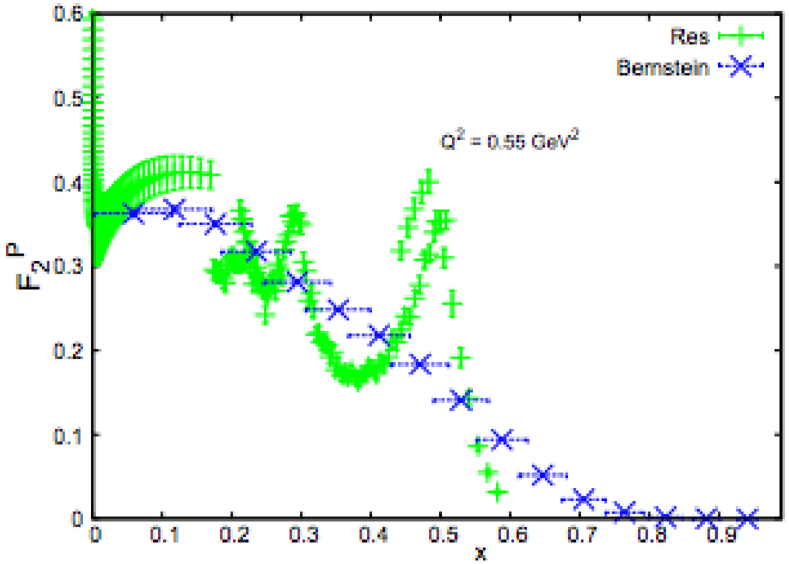
BONUS, Tkachenko et al, PRC(2014)

Alekhin, Kulagin, S.L. PRD(2004)

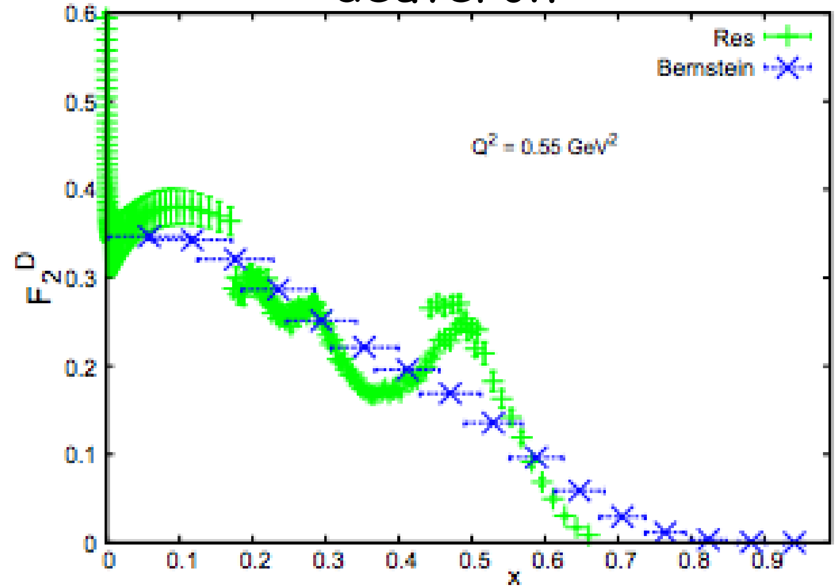


Most of the large x data lie in the resonance region: use Bernstein polynomials to average the data

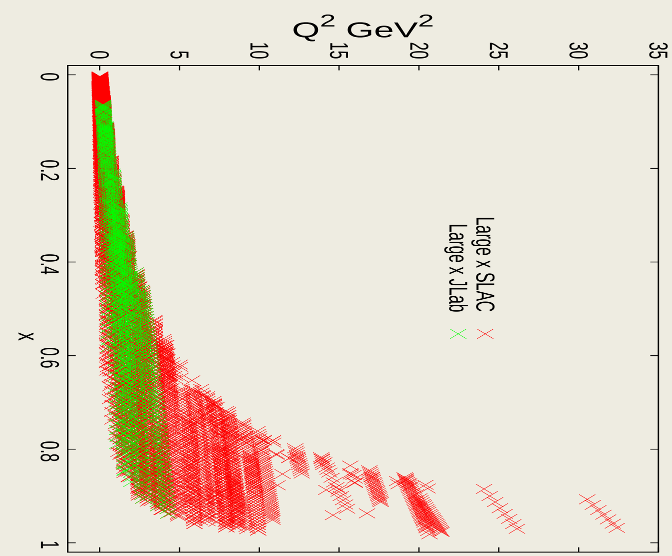
Q^2=0.55 GeV^2 proton



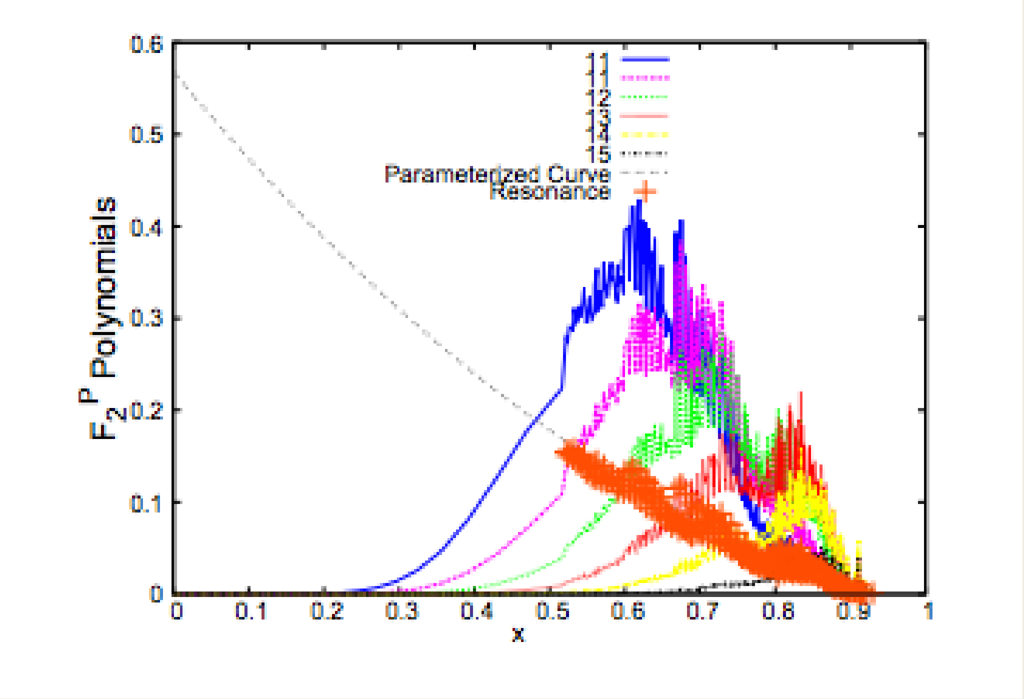
deuteron



Data set from Jefferson Lab +SLAC

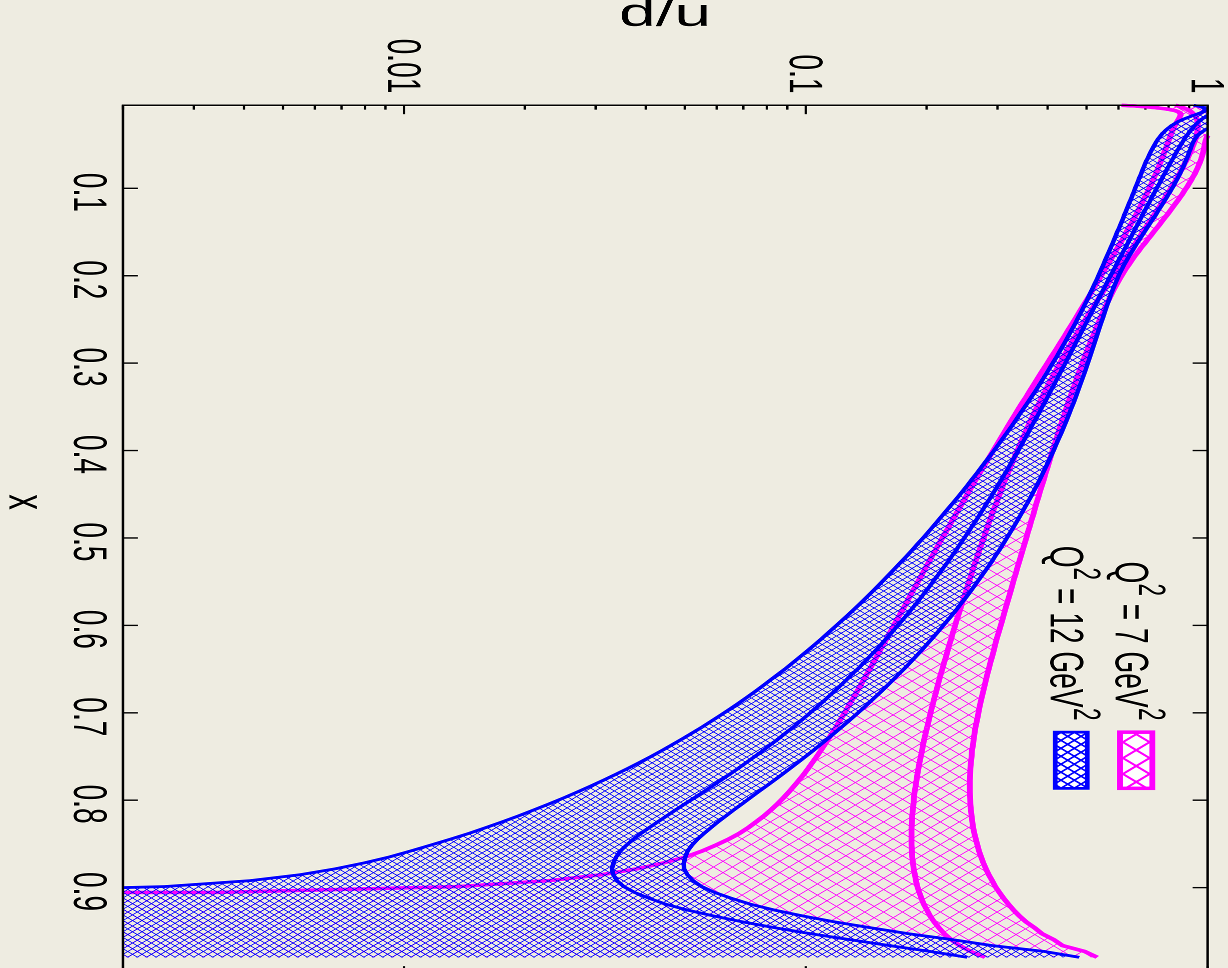


How the Bernstein polynomials work: weighted average with data



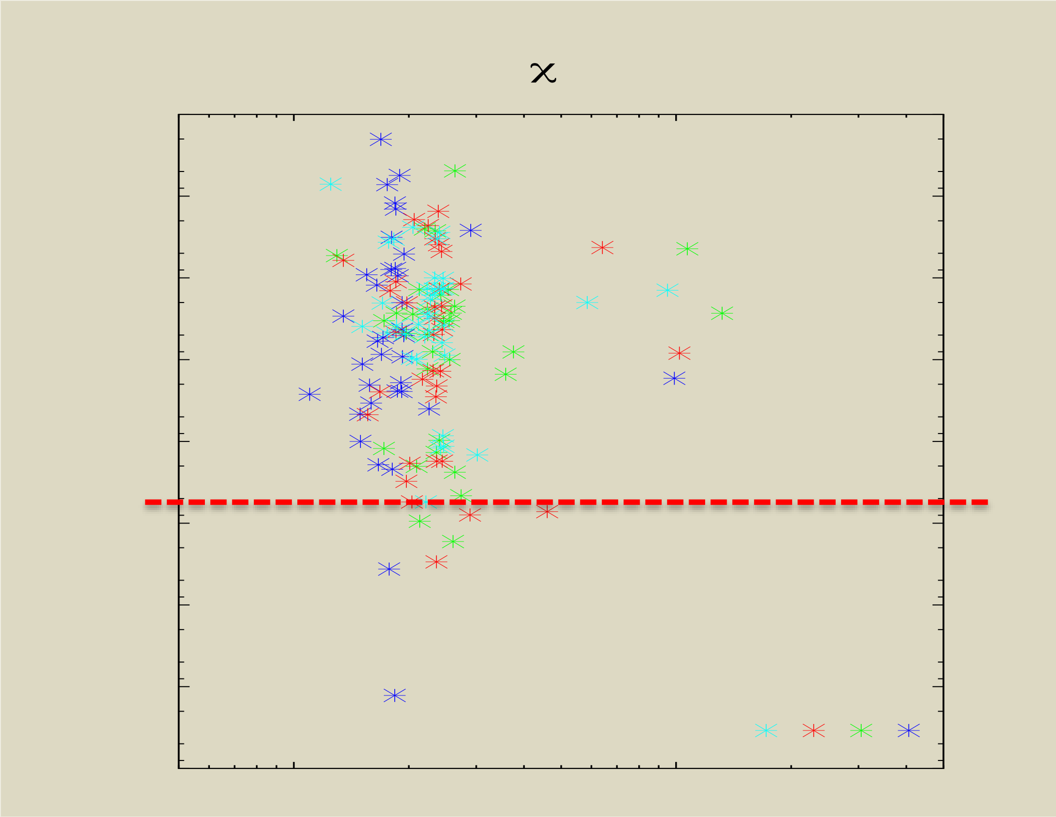
Q^2=2.5 GeV^2

Analysis with no Q^2 dependent corrections (no TMCs etc…)

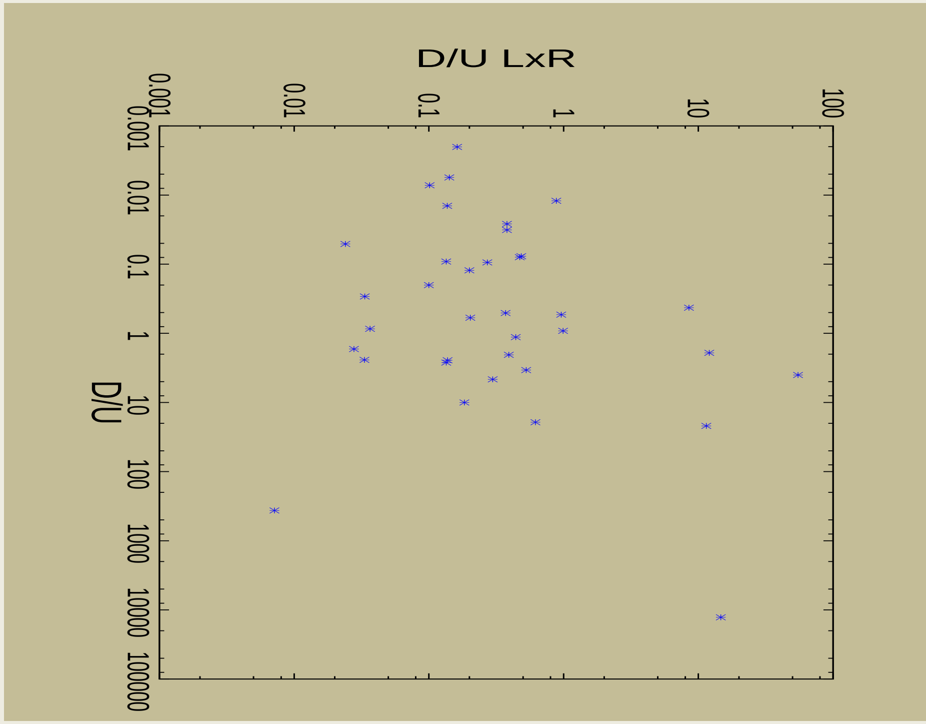


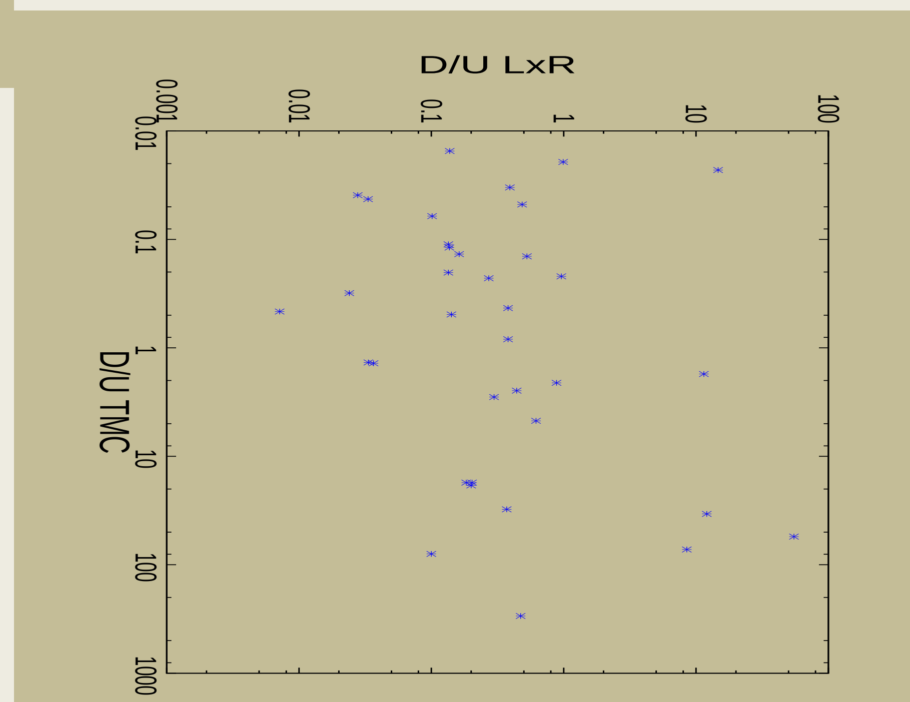
E. Askanazi, K. Holcomb, S. Liuti, J. Phys. G 42, no. 3, 034030 (2015) [arXiv:1411.2487 [hep-ph]].

Study clustering properties of data/correlations of various effects to reduce size of the error



…ongoing





… analysis of various components

We are studying similar characteristics of SOMs to devise a fitting procedure for GPDs: our new code has been made flexible for this use

Main question: Which experiments, observables, and with what precision are they relevant for which GPD components?

From Guidal and Moutarde, and Moutarde analyses (2009)

$$H_{Re} = P\int_0^1 dx\,[H(x,\xi,t) - H(-x,\xi,t)]\,C^+(x,\xi), \quad (1$$

$$E_{Re} = P\int_0^1 dx\,[E(x,\xi,t) - E(-x,\xi,t)]\,C^+(x,\xi), \quad (2$$

$$\tilde{H}_{Re} = P\int_0^1 dx\left[\tilde{H}(x,\xi,t) + \tilde{H}(-x,\xi,t)\right]C^-(x,\xi) \quad (3$$

$$\tilde{E}_{Re} = P\int_0^1 dx\left[\tilde{E}(x,\xi,t) + \tilde{E}(-x,\xi,t)\right]C^-(x,\xi) \quad (4$$

$$H_{Im} = H(\xi,\xi,t) - H(-\xi,\xi,t), \quad (5)$$

$$E_{Im} = E(\xi,\xi,t) - E(-\xi,\xi,t), \quad (6)$$

$$\tilde{H}_{Im} = \tilde{H}(\xi,\xi,t) + \tilde{H}(-\xi,\xi,t) \quad \text{and} \quad (7)$$

$$\tilde{E}_{Im} = \tilde{E}(\xi,\xi,t) + \tilde{E}(-\xi,\xi,t) \quad (8)$$

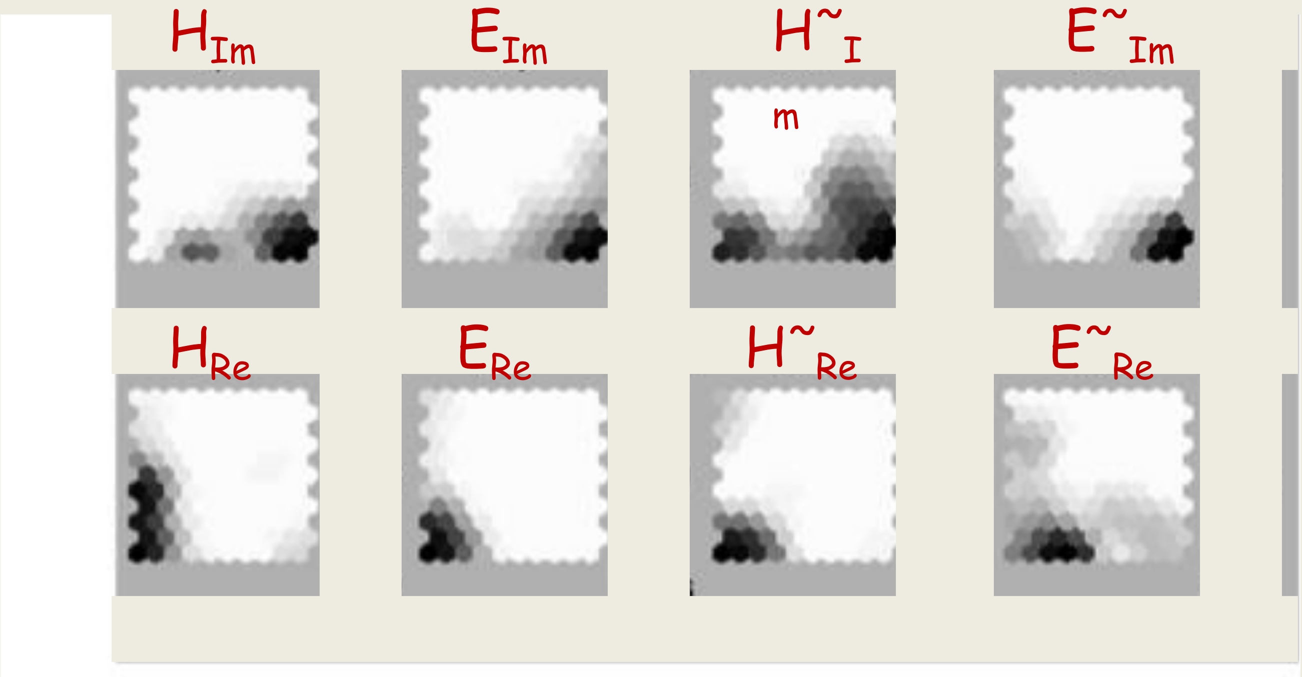
8 GPD-related functions

$$\boldsymbol{A_{\{C\}}}, A^{\sin\phi}_{\{C\}}, \boldsymbol{A^{\cos\phi}_{\{C\}}}, A^{\cos 2\phi}_{\{C\}}, A^{\cos 3\phi}_{\{C\}}$$
$$A_{\{LU,DVCS\}}, A^{\sin\phi}_{\{LU,DVCS\}}, A^{\cos\phi}_{\{LU,DVCS\}}, A^{\sin 2\phi}_{\{LU,DVCS\}}$$
$$A_{\{LU,I\}}, \boldsymbol{A^{\sin\phi}_{\{LU,I\}}}, A^{\cos\phi}_{\{LU,I\}}, A^{\sin 2\phi}_{\{LU,I\}}$$
$$\boldsymbol{A^{\sin\phi}_{\{Ux,I\}}},$$
$$A_{\{Uy,DVCS\}},$$
$$\boldsymbol{A_{\{Uy,I\}}} \quad \text{and} \quad \boldsymbol{A^{\cos\phi}_{\{Uy,I\}}} \quad (13)$$

17 obsvervables (6 LO) from HERMES + Jlab data

“a challenge for phenomenology…” (Moutarde) + “theoretical bias”

The 8 GPDs are the dimensions in our analysis



Conclusions/Outlook

✓ Presented: a new computational method,

Self-Organizing Maps

for parametrizing nucleon PDFs … and beyond…

✓ The method works: we succeeded in minimizing the χ^2 and in performing error analyses for PDFs

E. Askanazi, K. Holcomb, S. Liuti, J. Phys. G 42, no. 3, 034030 (2015) [arXiv:1411.2487 [hep-ph]].

✓ In progress: study more observables from varied sets of data where predictivity/theoretical input is important (d/u at x ➔ 1, …)

✓Future Studies: GPDs, theoretical developments, connection with "similar approaches", complexity theory…

Issues for discussion

- New ingredients for multi-variable analysis

- Theoretical vs. Experimental, Systematic and Statistical Uncertainties (correlations)

- Estimators: χ^2, weighted χ^2, …

- Non-linearity