Self Organizing Maps Parameterization of Parton Distribution Functions

Simonetta Liuti
University of Virginia

4th HiX Workshop

INFN - Laboratori Nazionali di Frascati
Italy

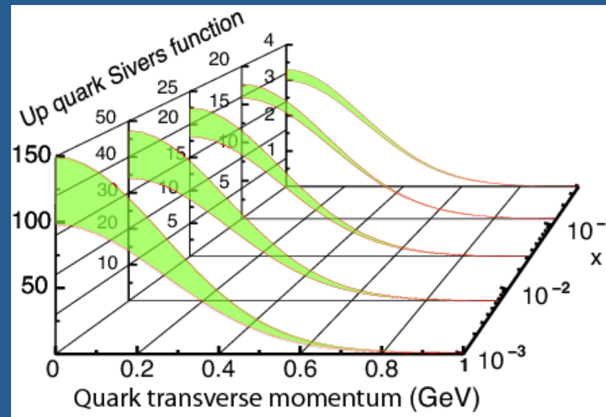
November, 17th-21st 2014

Outline

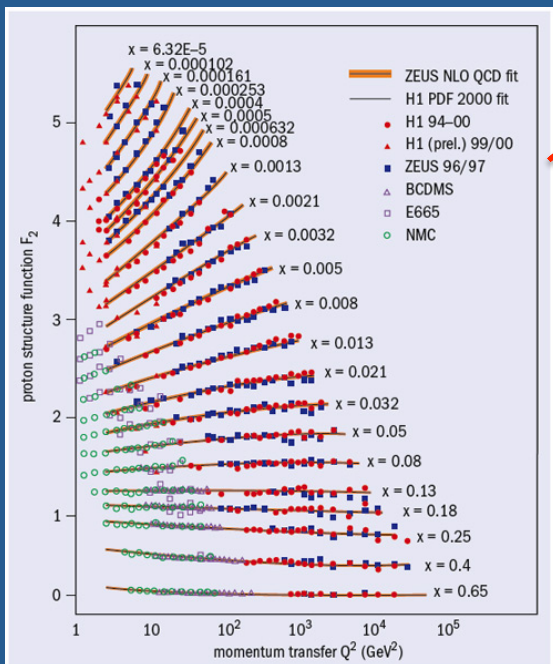
- *Introduction*
- *Artificial Neural Networks in HEP/Nuclear Data Analyses*
- *Self Organizing Maps (SOMs) Algorithm*
- *SOMPDFs*
- *Comparison with NNPDFs*
- *Future Work: Extension to GPDs*
- *Conclusions/Outlook*

Dealing with an increasingly complicated and diverse set of observables

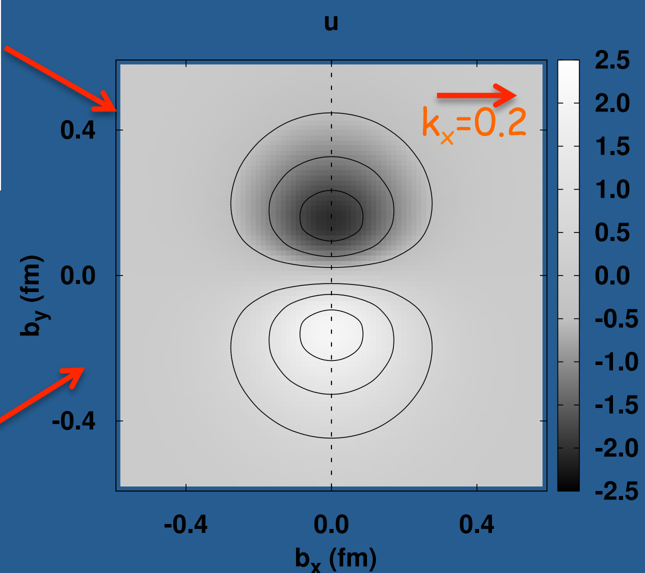
TMDs



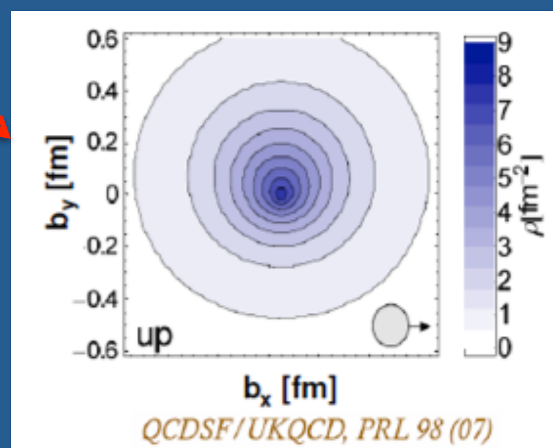
PDFs



GTMDs



GPDs



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

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

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

2) The uncertainty is determined with the Hessian method.

- We can attack 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 (see also the Statistical Parton Model, F. Buccella)
- This goal is at reach with HPC

Artificial Neural Networks in HEP/Nuclear Data Analyses

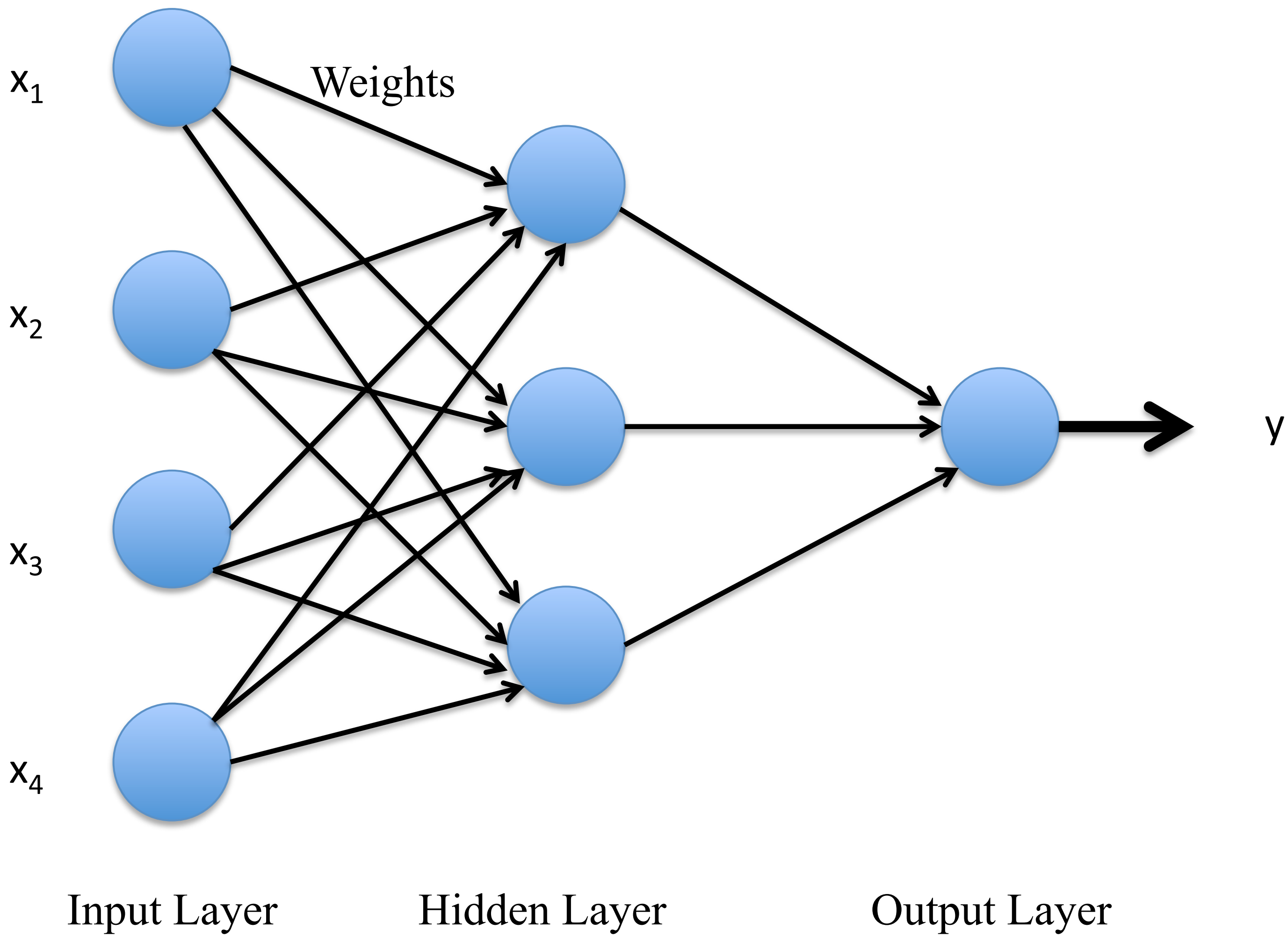
- ✓ Neural Networks (NN) have been widely applied for the analysis of HEP data and PDF parameterizations *(E. Nocera's talk)*
- ✓ When applied to data modeling, NNs are a non-linear statistical tool
- ✓ The network makes changes to its connections upon being informed of the "correct" result.
- ✓ We quantify this through cost/object function.

Cost function

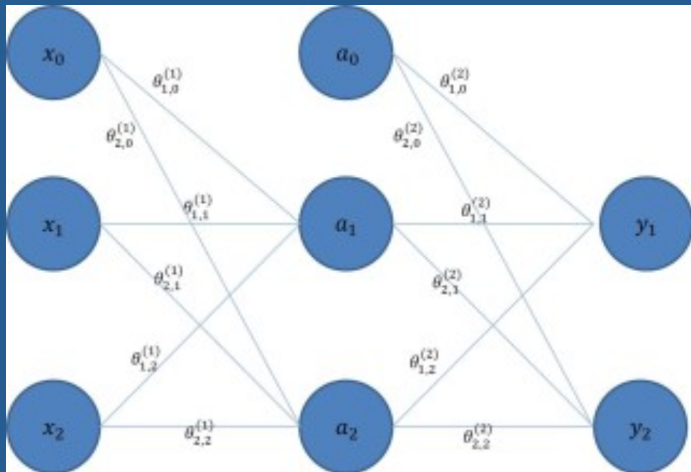
$$C(w) = \frac{1}{2n} \sum_x \| y(x) - a \|^2$$

n= number of training inputs
x = training input
Y(x) = desired output, a=output when x is the input
w= weights

- ✓ C measures the importance to detect or miss a particular occurrence
- ✓ The aim is to minimize the cost!



x_1
x_2
x_3
x_4
Weights
y
Input Layer
Hidden Layer
Output Layer

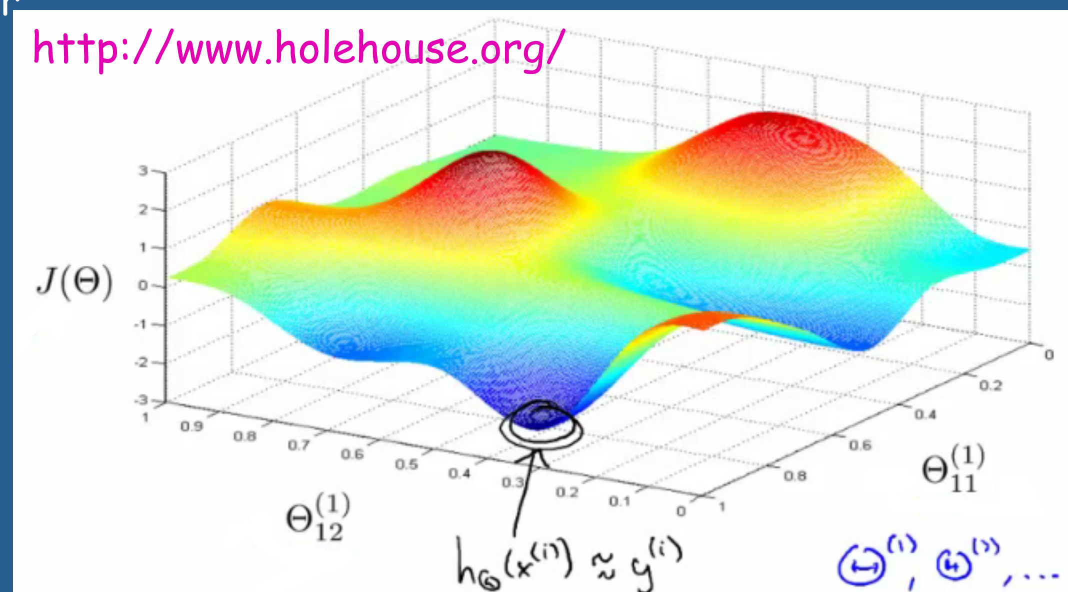


Input layer Hidden layer Output layer

Back propagation

1) Take the output from the network
2) Compare it to the real data values
3) Calculate how wrong the network was (define error: how wrong the parameters/weights were)

4) The errors are used to calculate the partial derivatives in the parameters which are necessary to minimize the cost function!



Supervised learning

1) A set of examples is given
2) The goal is to force the data to match the examples as closely as possible.

The cost function includes information about the data domain!

Application to PDFs…

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

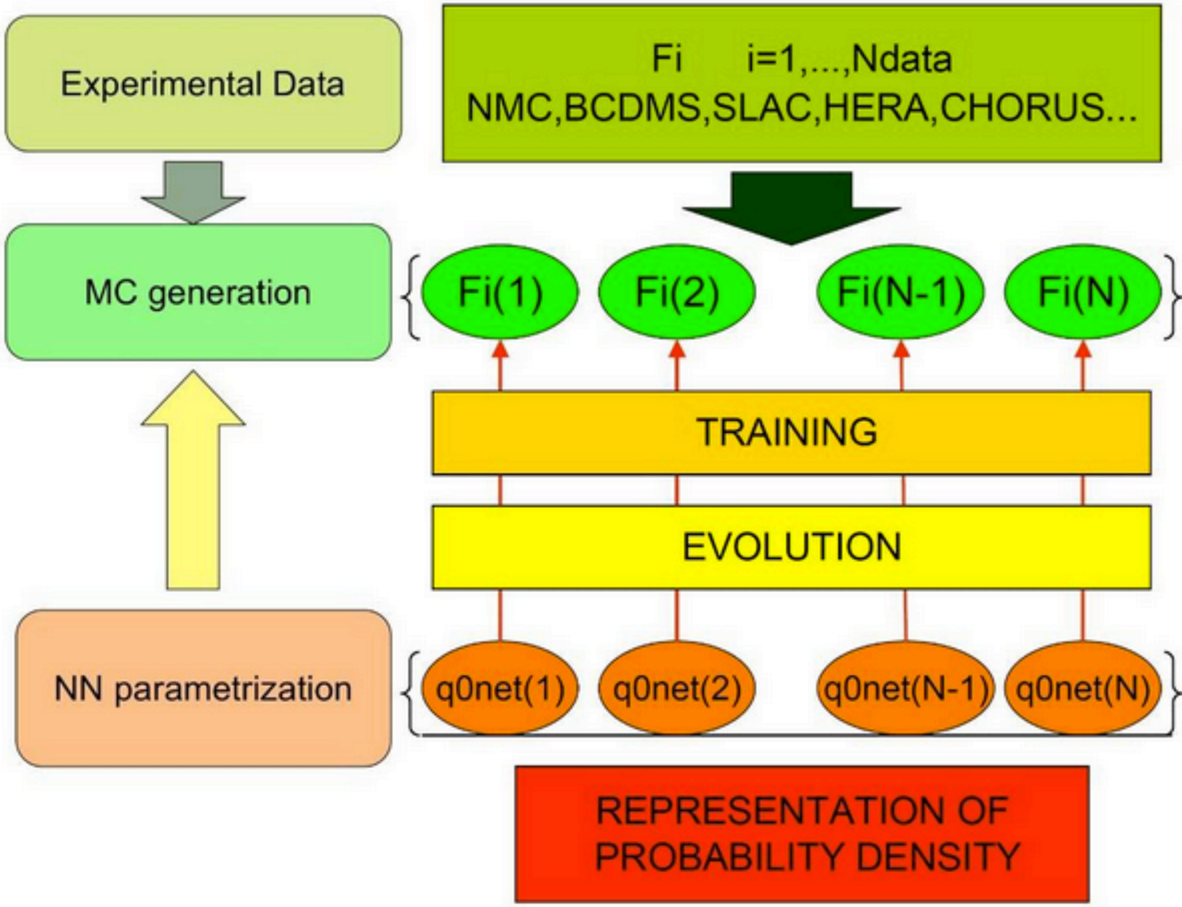
- Monte Carlo generation of data replicas

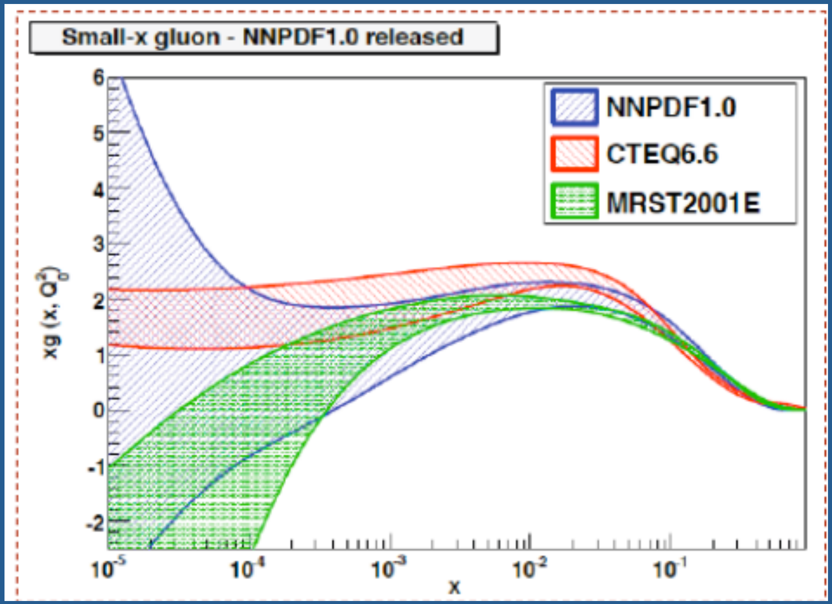
→ no need for linear propagation of errors
→ possibility to test for non Gaussian behaviour in fitted PDFs

- Neural Networks parametrization of PDFs

→ 7 independent PDFs, 259 parameters
→ unbiased parametrization

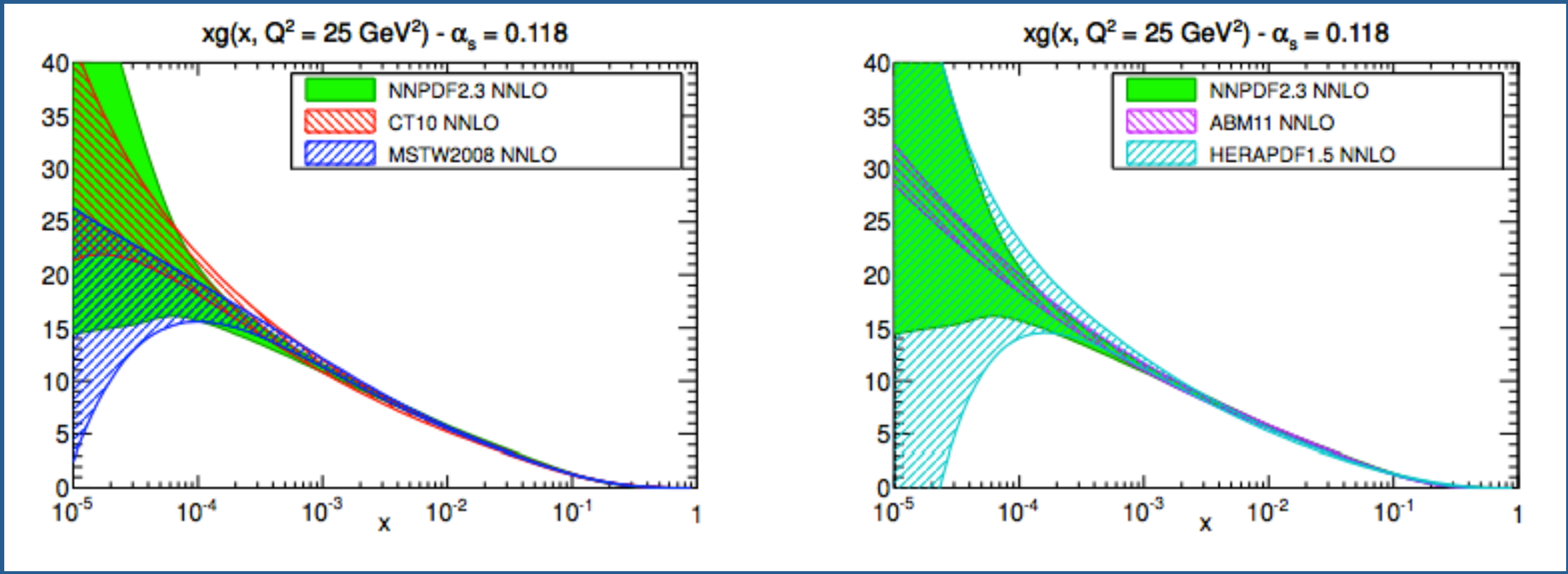
- Evolution using DGLAP equations
- Genetic Algorithm's training of neural networks parameters
- Analysis of χ^2 distributions





NNPDF before LHC data

NNPDF including LHC data, JHEP(2012)



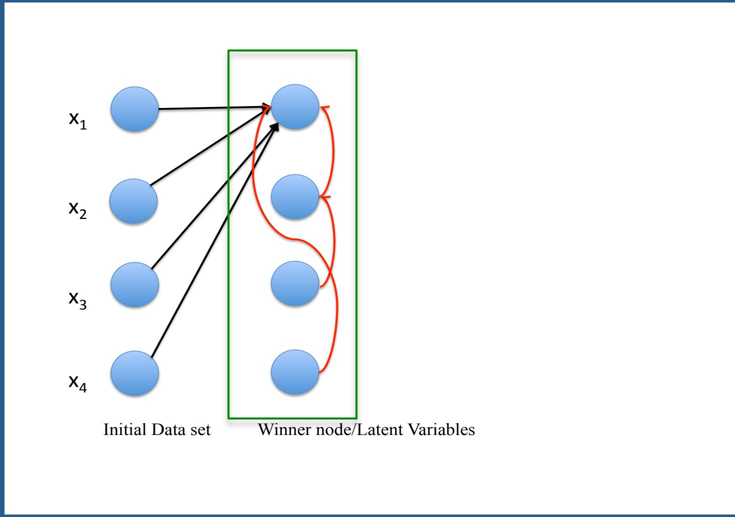
- Conventional methods' problem: Dependence on initial bias
- To overcome this introduce Neural Networks
- NNs' problem: Extrapolation is difficult

Important 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"?

One must improve on the ANN type algorithm!

Unsupervised Learning



No a priori examples are given.
The goal is to minimize the cost function by similarity relations, or by finding how the data cluster or self-organize

J. Carnahan, H. Honkanen, S.Liuti, Y. Loitiere, P. Reynolds, Phys Rev D79, 034022 (2009)

3 of the benchmarks discussed at this meeting aimed at establishing:

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

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.

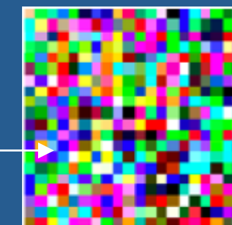
The nodes are decoders of the input signals -- can be used for pattern recognition.

Two dimensional maps are used to cluster/visualize high-dimensional data.

SOMs Algorithm

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

V_i=(R,B,G)



<u>Initialization</u> → Input vector of dimension "**n**" associated to cell "**i**":

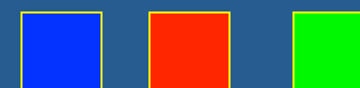
$$V_i = \left[v_i^{(1)}, ..., v_i^{(n)}\right]$$

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

<u>Training</u> → Input data:

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

isomorphic

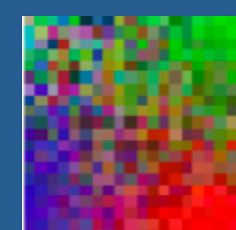


x compared to V_i 's with "similarity" metric(L1):

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

(Aggawal et al., 2000)

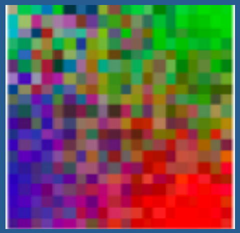
Location of best match "winner" gives location of response
(active cell, all others are passive)



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



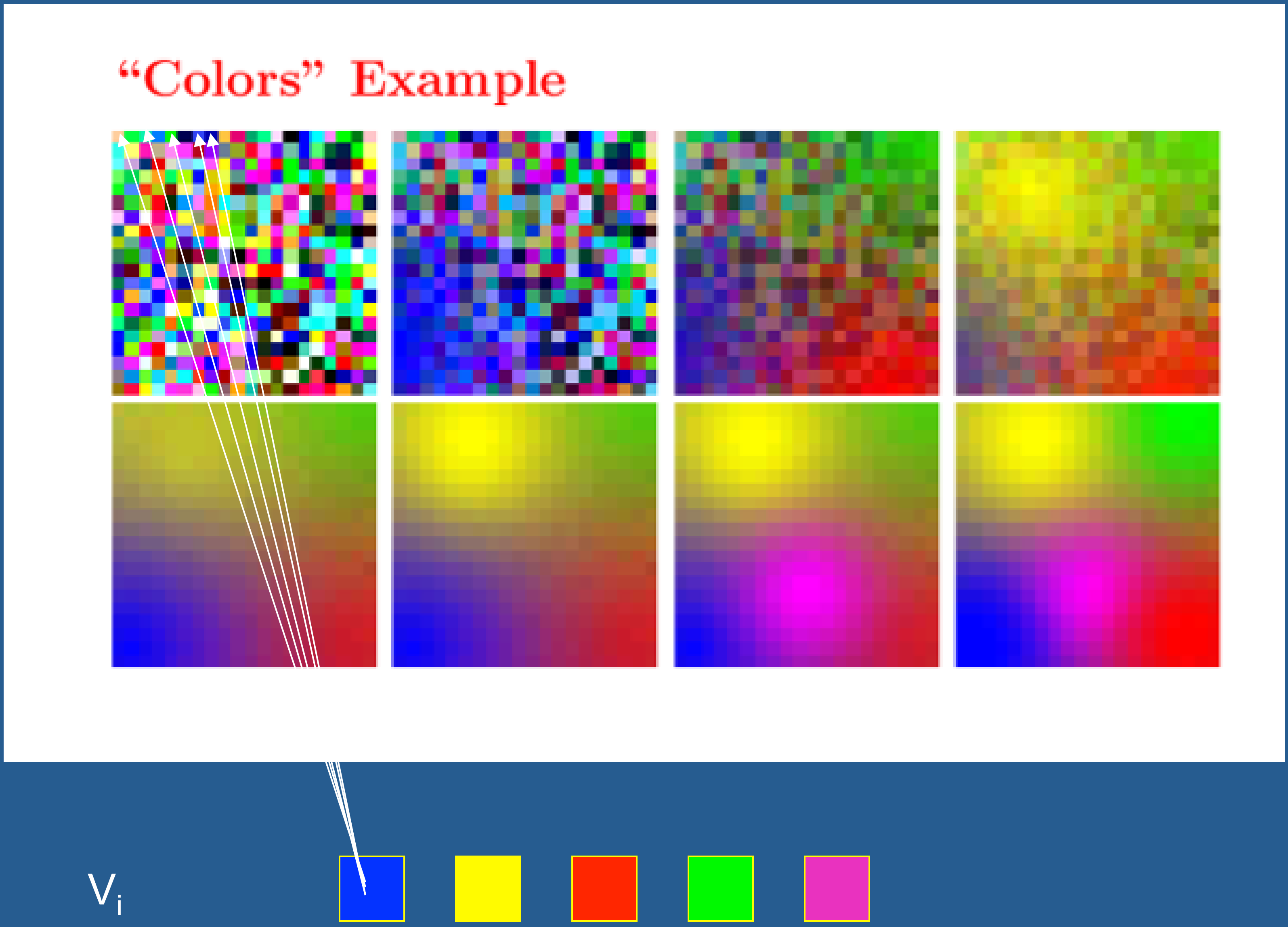
$$V_i(n+1) = V_i(n) + h_{ci}(n)\left[x(n) - V_i(n)\right]$$

iteration number

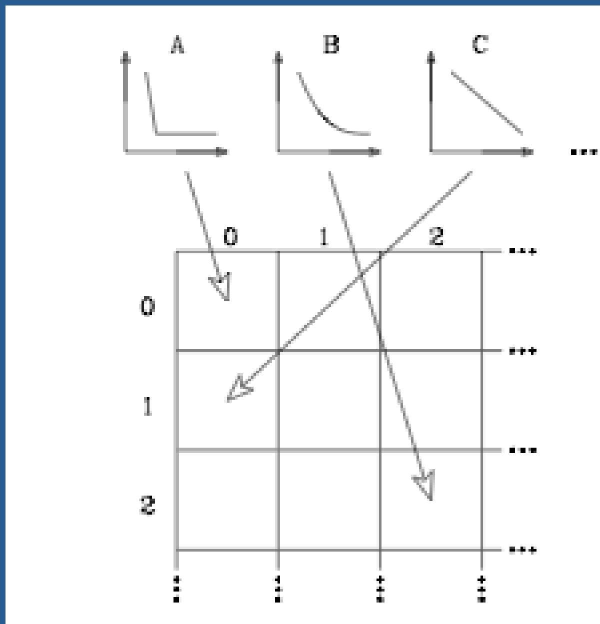
$$h_{ci}(n) = f(\|r_c - r_i\|) \equiv \alpha(n)\exp\left(\frac{-\|r_c - r_i\|^2}{2\sigma^2(n)}\right)$$

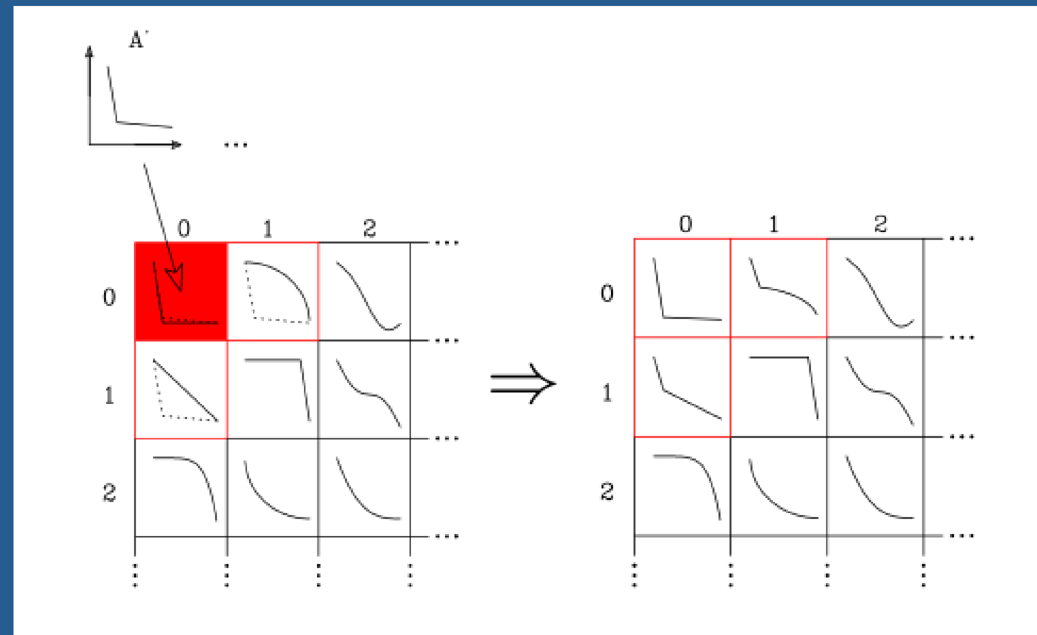
neighborhood function decreases with "n" and "distance"

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



Simple Functions Example





Initialization: functions are placed on map

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

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

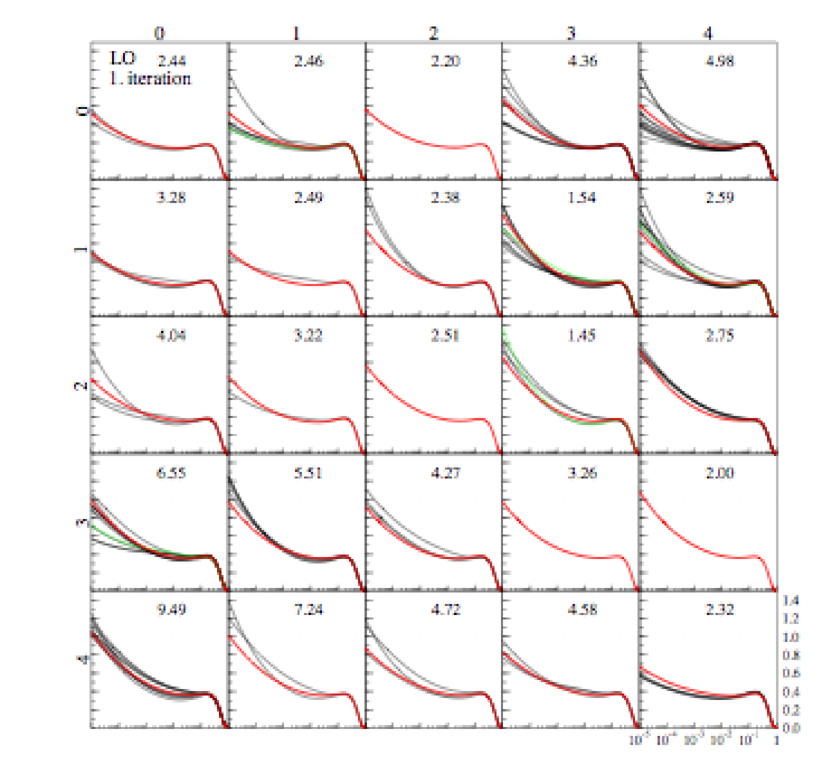
SOMPDF Method

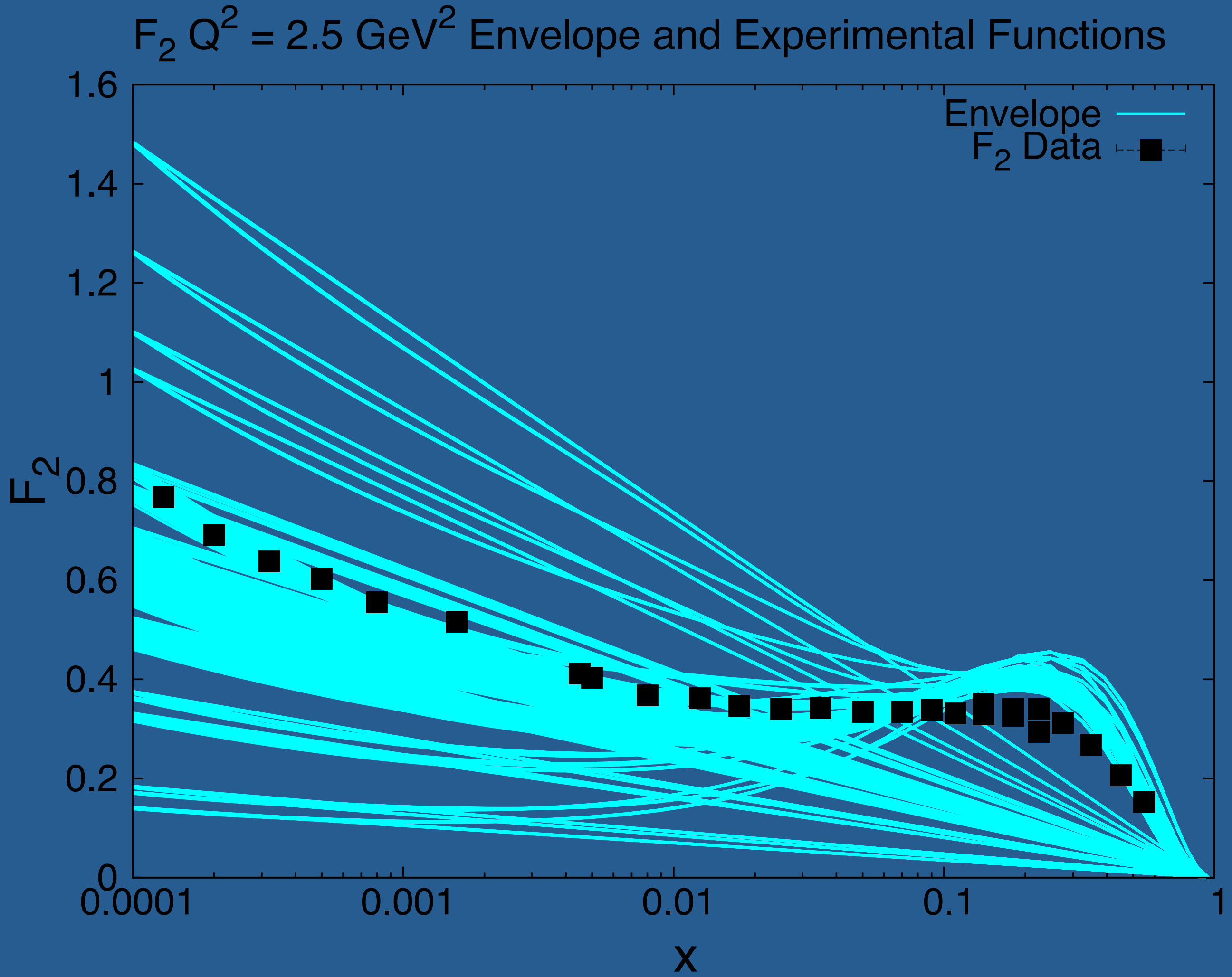
Initialization: a set of database/input PDFs is formed by selecting at random from existing PDF sets and varying their parameters.
Baryon number and momentum sum rules are imposed at every step.
These input PDFs are used to initialize the map.

Mixing

- In generating the PDFs (for the map and for the training) we need to avoid introducing a functional bias
- Thus we mix together variations of different structure functions
  - Random perturbations are used to generate a variant of a standard set of structure functions—currently based on GRV, MRST, AMP. We select some number of these varied functions, then combine them in a weighted-average linear combination to obtain a final candidate PDF.
  - Sum rules are enforced on each candidate “mixed” PDF

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



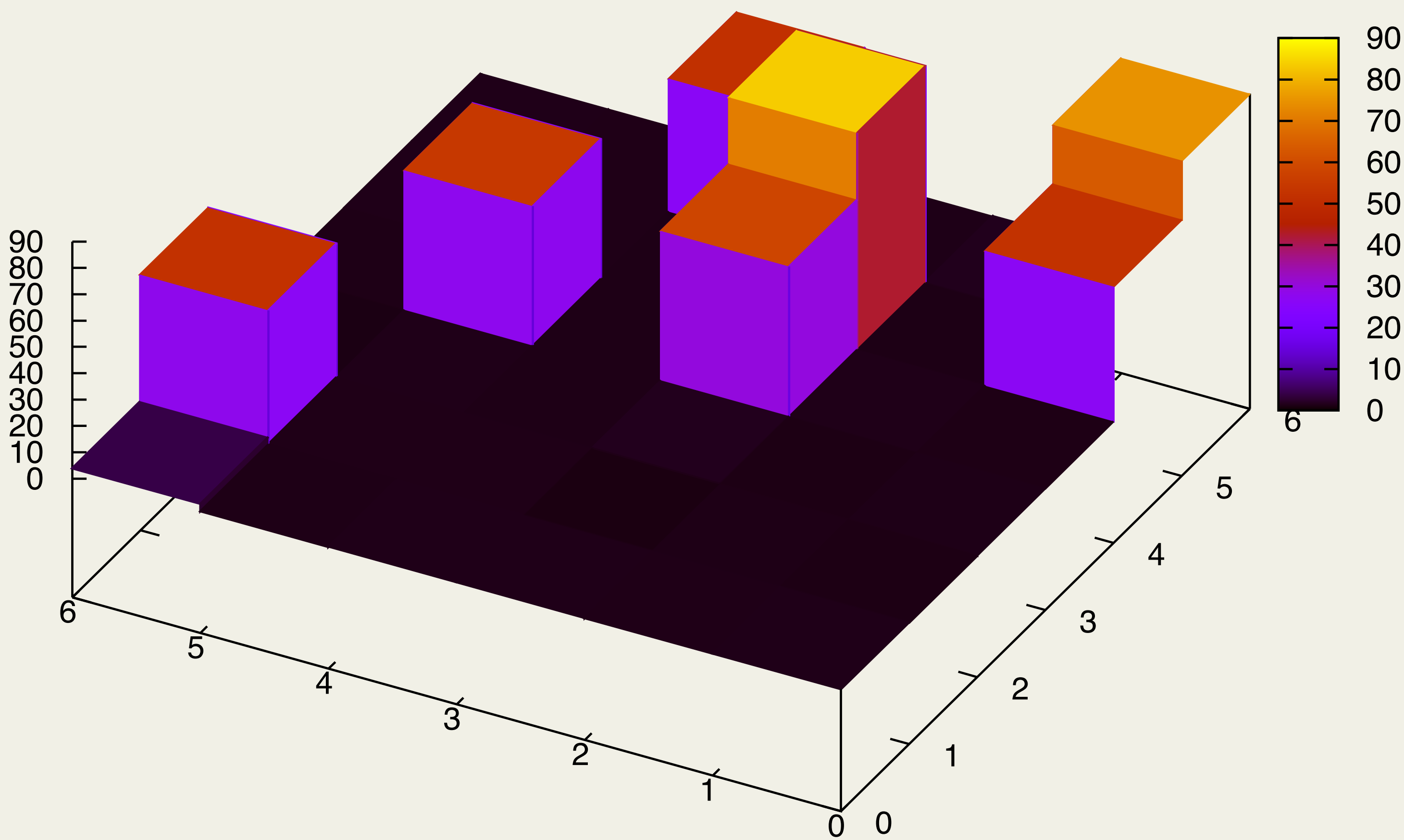


F2 Q2 = 2.5 GeV2 Envelope and Experimental Functions
Envelope
F2 Data
F2
x
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
0.0001
0.001
0.01
0.1
1

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).

chi squared



90
80
70
60
50
40
30
20
10
0
6
5
4
3
2
1
0
0
1
2
3
4
5
6
90
80
70
60
50
40
30
20
10
0

Advantages with respect to "conventional way":

- Initial scale ansatz

$$F(x,Q_0) = A_0 x^{A_1}(1-x)^{A_2} P(x; A_3, \ldots)$$

- Evolve to higher scale
- Compute observables e.g. $F_2^p(x,Q^2)$
- Compare with the data e.g.

$$\chi^2(\{a\}) = \sum_{\mathrm{expt.}} \left\{ \sum_{i=1}^{N_e} \frac{(D_i - T_i)^2}{\alpha_i^2} - \sum_{k,k'=1}^{K} B_k \left(A^{-1}\right)_{kk'} B_{k'} \right\}$$

where $B_k = \sum_{i=1}^{N_e} \frac{\beta_{ki}(D_i - T_i)}{\alpha_i^2}$, $\qquad A_{kk'} = \delta_{kk'} + \sum_{i=1}^{N_e} \frac{\beta_{ki}\beta_{k'i}}{\alpha_i^2}$

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

Advantages over NNPDFs

Mechanism responsible for the self-organization of the different representations of information: the response of the network changes in such a way that the location of the cell holding a given response corresponds to a specific input signal.

Geometrical arrangement of information is maintained during the training.

SOM work differently from ANN that do not keep track of the inter-connections among clustering of data at different stages of the network training.

Important because it allows for "user/expert's" intervention: evaluate the impact of possibe theoretical input

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

History/Organization of work

2006-2007 PDF Parametrization Code - SOMPDF.0 - using Python, C++, fortran. Preliminary results discussed at conferences: DIS 2006,…

2008 First analysis published -
J. Carnahan, H. Honkanen, S.Liuti, Y. Loitiere, P. Reynolds, Phys Rev D79, 034022 (2009)

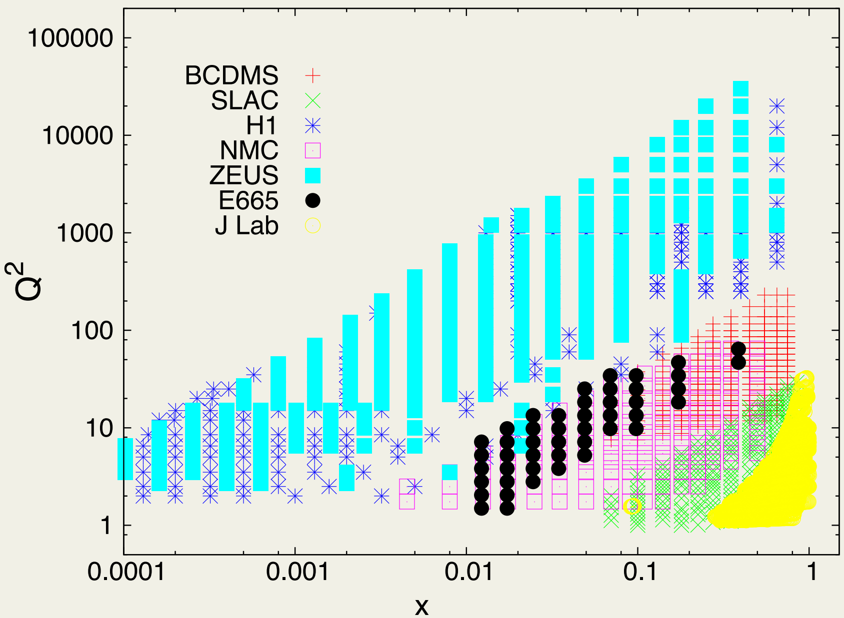
2009 New group formed (K. Holcomb, D. Perry)
Rewriting, reorganization and translation of First Code into a uniform language, fortran 95.

2010/11 Implementation of Error analysis. Extension to new data analyses.
(E. Askanazi, K. Holcomb)

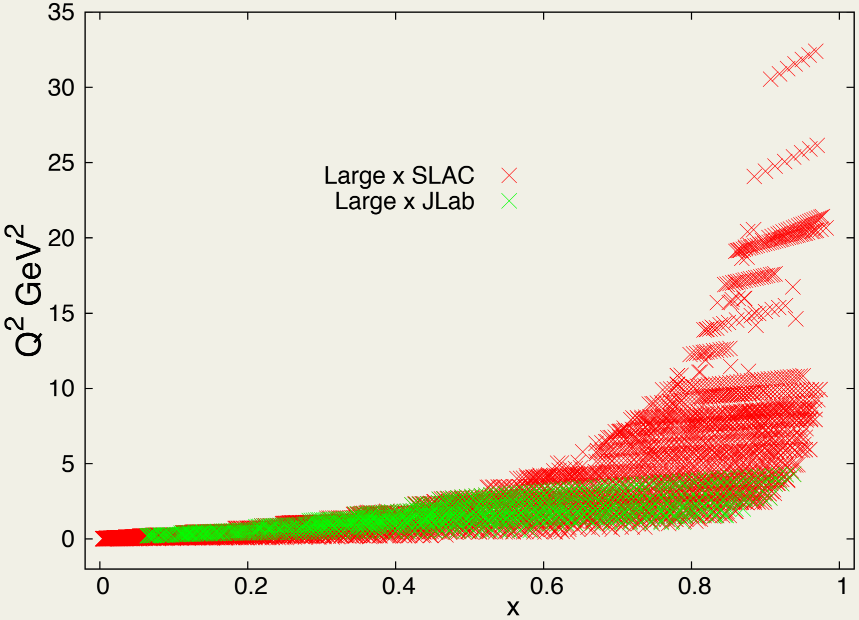
2013 PDF Parametrization Code ready to be released- SOMPDF.1

Group Website (under construction): http://faculty.virginia.edu/sompdf/

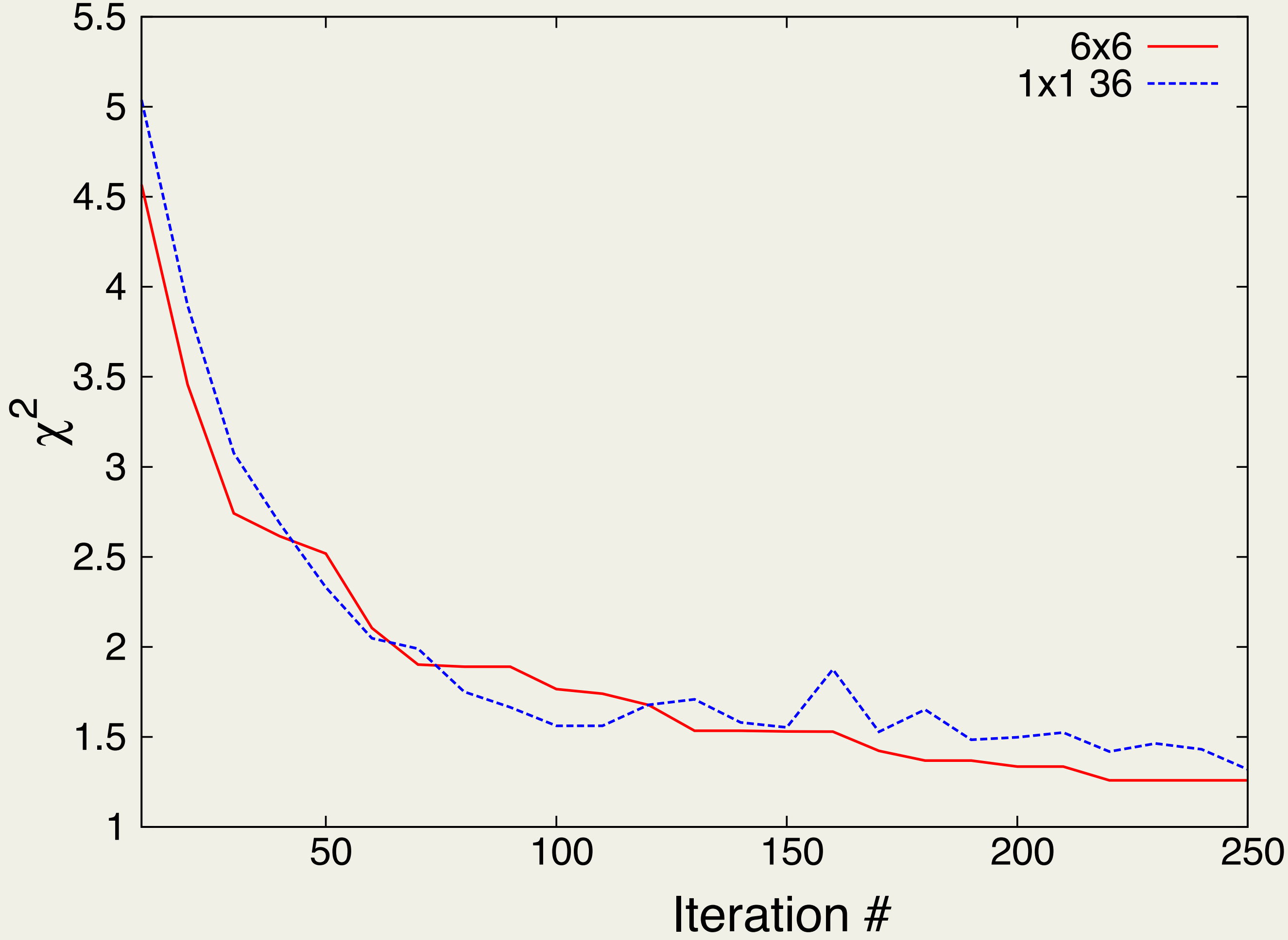
SOMPDF.0
J. Carnahan, H. Honkanen, S.L., Y. Loitiere, P. Reynolds, Phys Rev D79, 034022 (2009)
SOMPDF.1,
K. Holcomb, S.L., D.Z.Perry, hep-ph (2010)



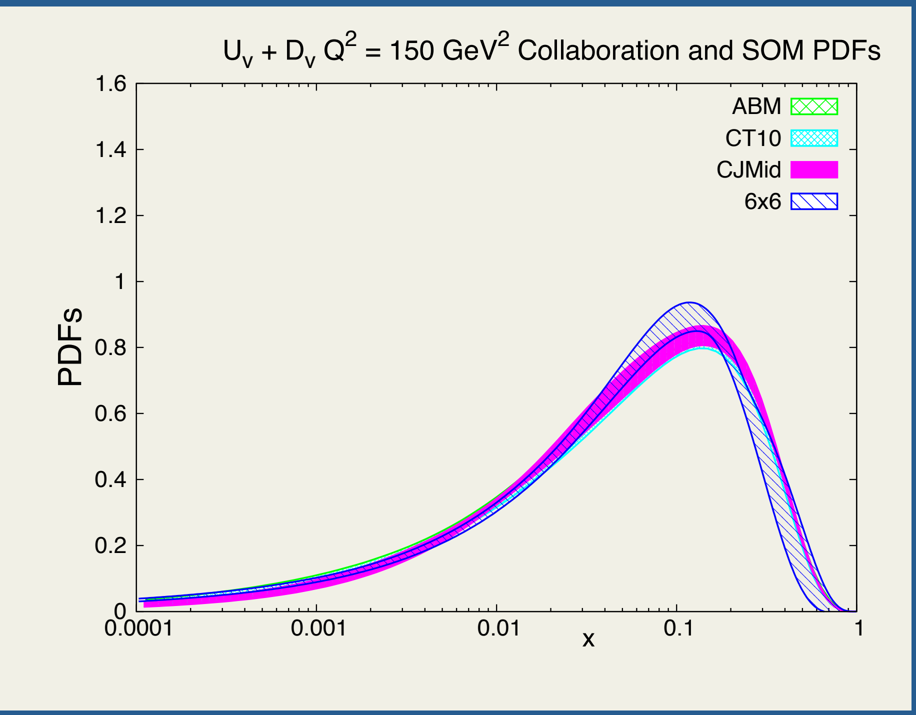
BCDMS
SLAC
H1
NMC
ZEUS
E665
J Lab
100000
10000
1000
100
10
1
0.0001
0.001
0.01
0.1
1
Q^2
x



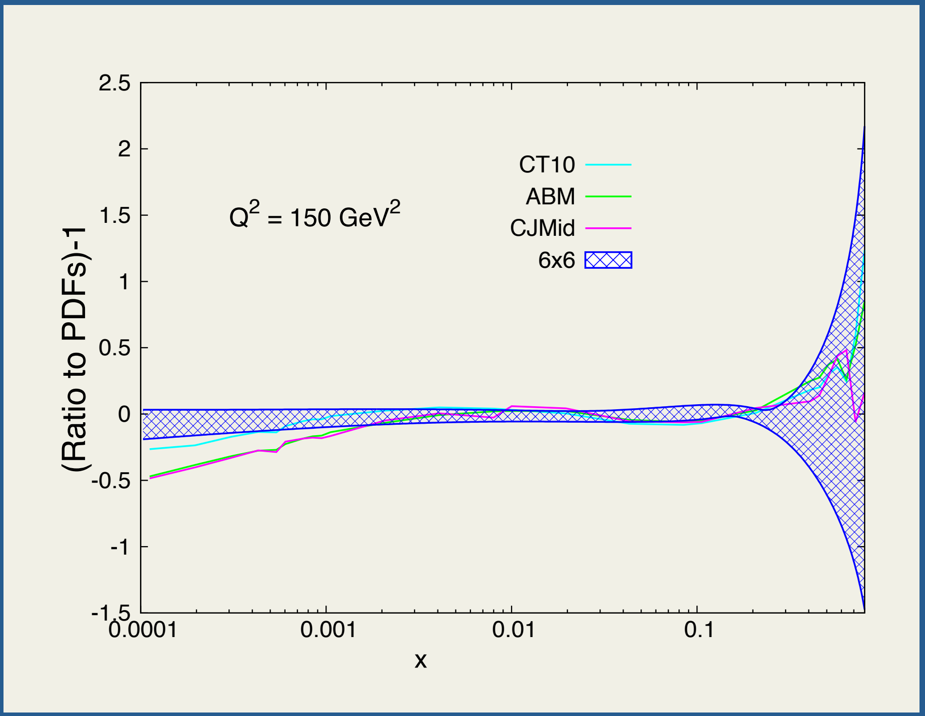
Large x SLAC
Large x JLab
35
30
25
20
15
10
5
0
0.2
0.4
0.6
0.8
1
Q^2 GeV^2
x



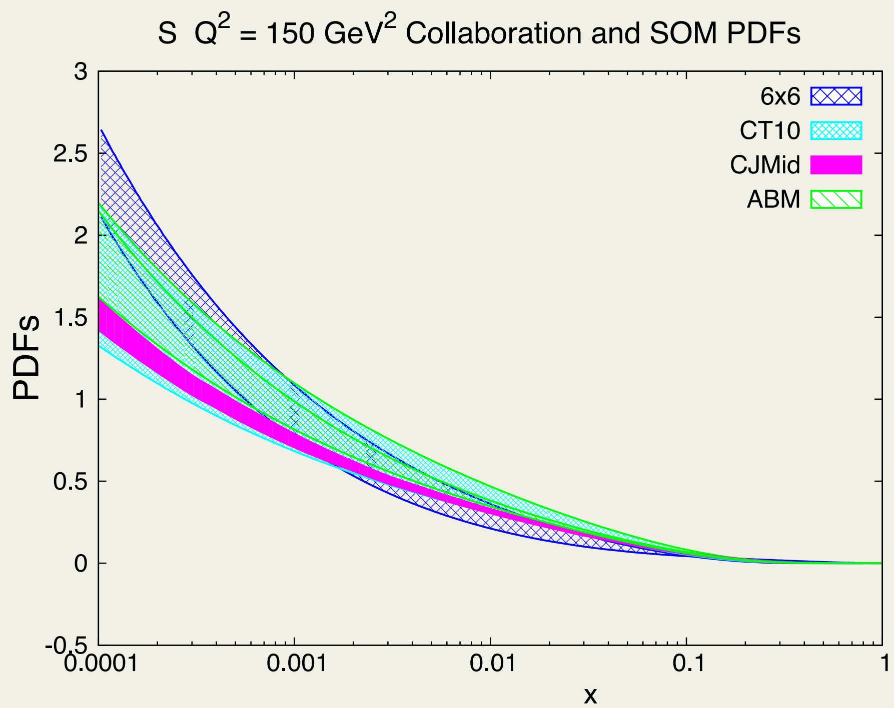
6x6
1x1 36
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
χ^2
50
100
150
200
250
Iteration #



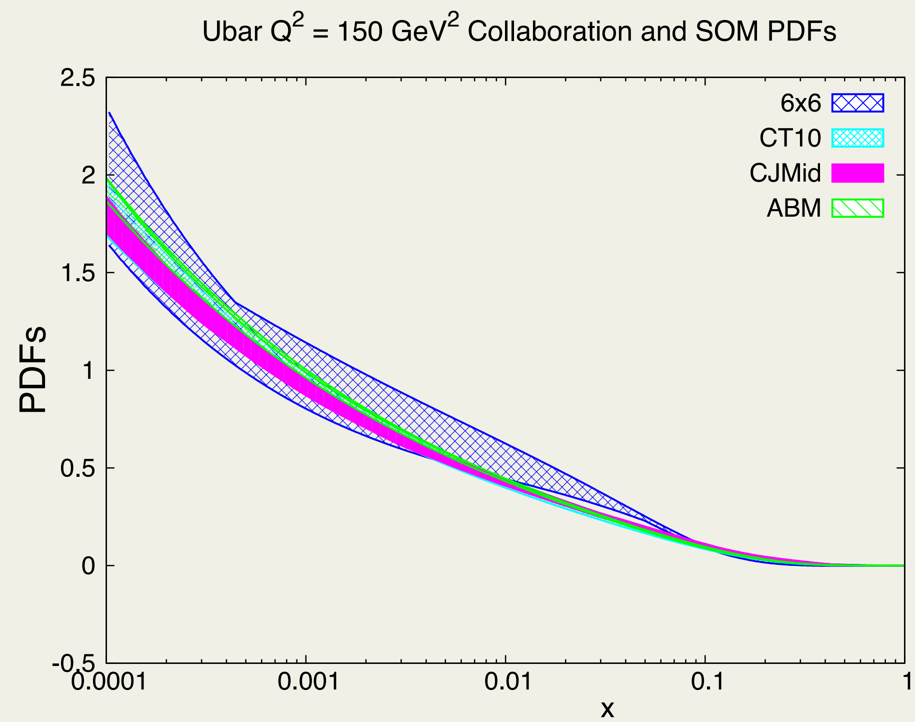
$U_v + D_v$ Q^2 = 150 GeV^2 Collaboration and SOM PDFs
ABM
CT10
CJMid
6x6
PDFs
x
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.0001
0.001
0.01
0.1
1



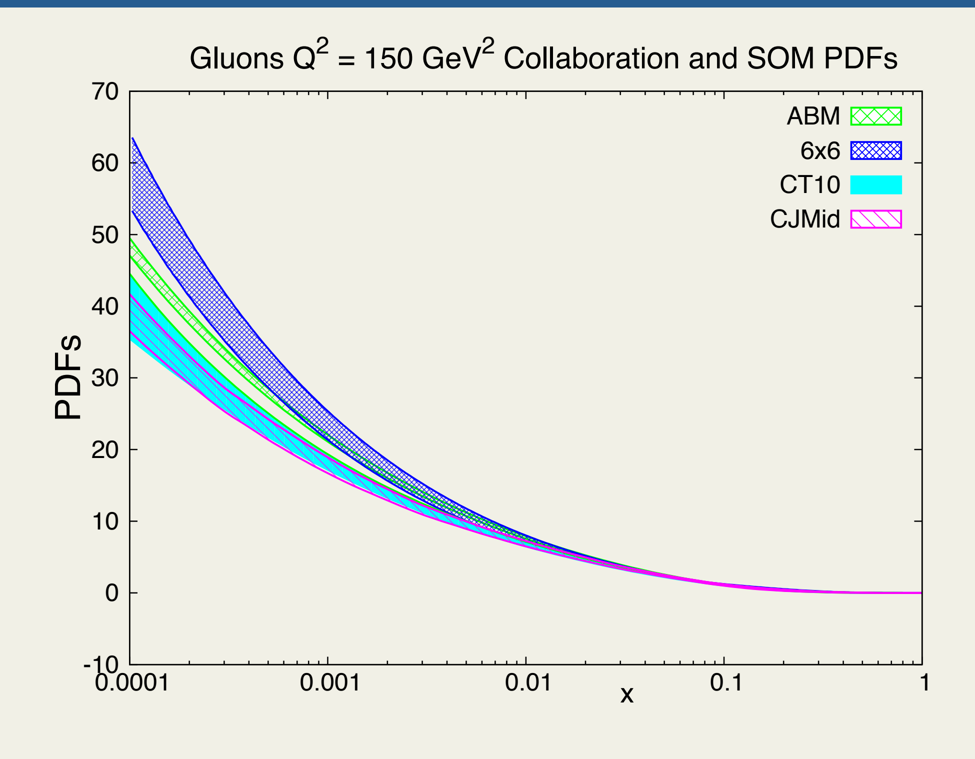
Q^2 = 150 GeV^2
CT10
ABM
CJMid
6x6
(Ratio to PDFs)-1
x
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
0.0001
0.001
0.01
0.1

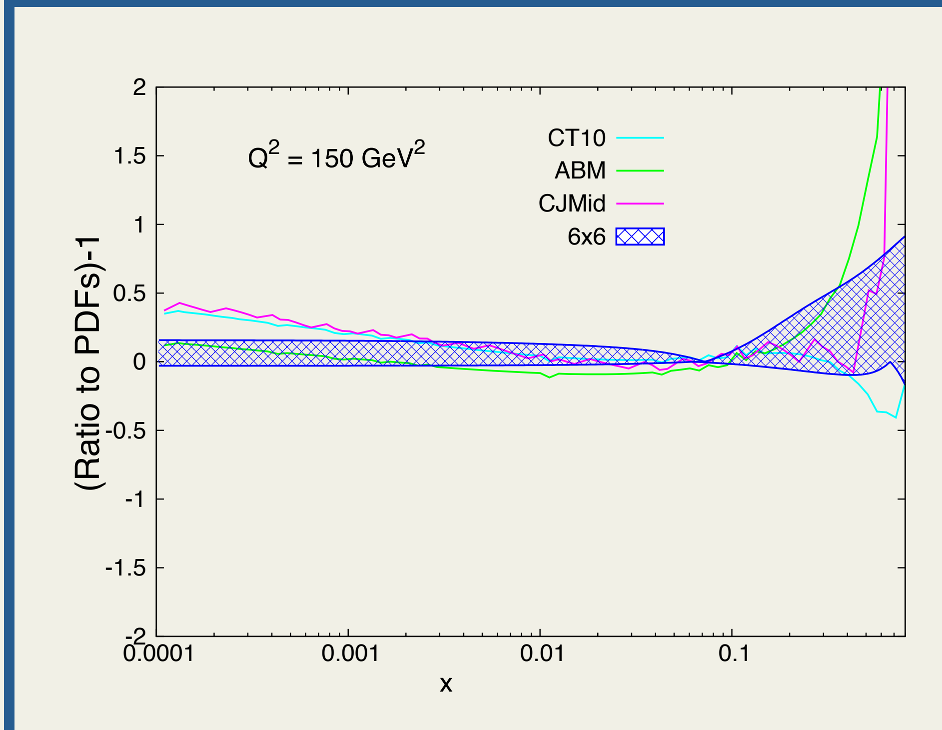


S Q^2 = 150 GeV^2 Collaboration and SOM PDFs
6x6
CT10
CJMid
ABM
PDFs
x
3
2.5
2
1.5
1
0.5
0
-0.5
0.0001
0.001
0.01
0.1
1



Ubar Q^2 = 150 GeV^2 Collaboration and SOM PDFs
6x6
CT10
CJMid
ABM
PDFs
x
2.5
2
1.5
1
0.5
0
-0.5
0.0001
0.001
0.01
0.1
1

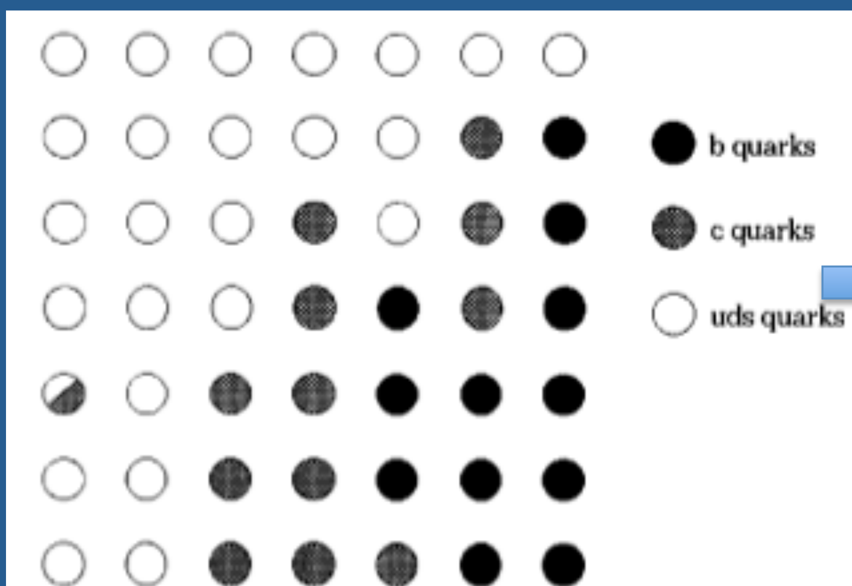

Gluons Q^2 = 150 GeV^2 Collaboration and SOM PDFs
PDFs
x
ABM
6x6
CT10
CJMid
70
60
50
40
30
20
10
0
-10
0.0001
0.001
0.01
0.1
1


Q^2 = 150 GeV^2
(Ratio to PDFs)-1
x
CT10
ABM
CJMid
6x6
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0.0001
0.001
0.01
0.1

Extension to multidimensional parton distributions/multiparton correlations: GPDs

SOMs differently from standard ANN methods are "unsupervised": they find similarities in the input data without a training target.

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
(see e.g. Der, Hermann, Phys.Rev.E (1994), Guimera et al., Phys. Rev.E (2003))

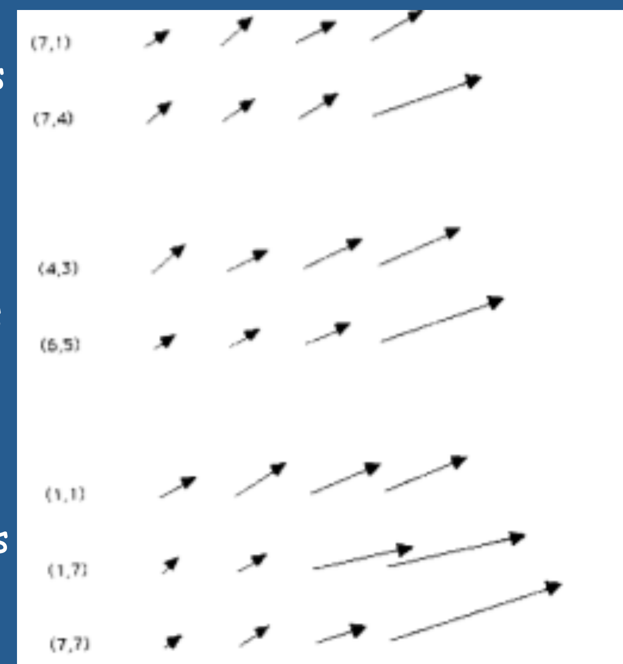


Lonnblad, Peterson, Pi, Computer Physics Comm. 1991

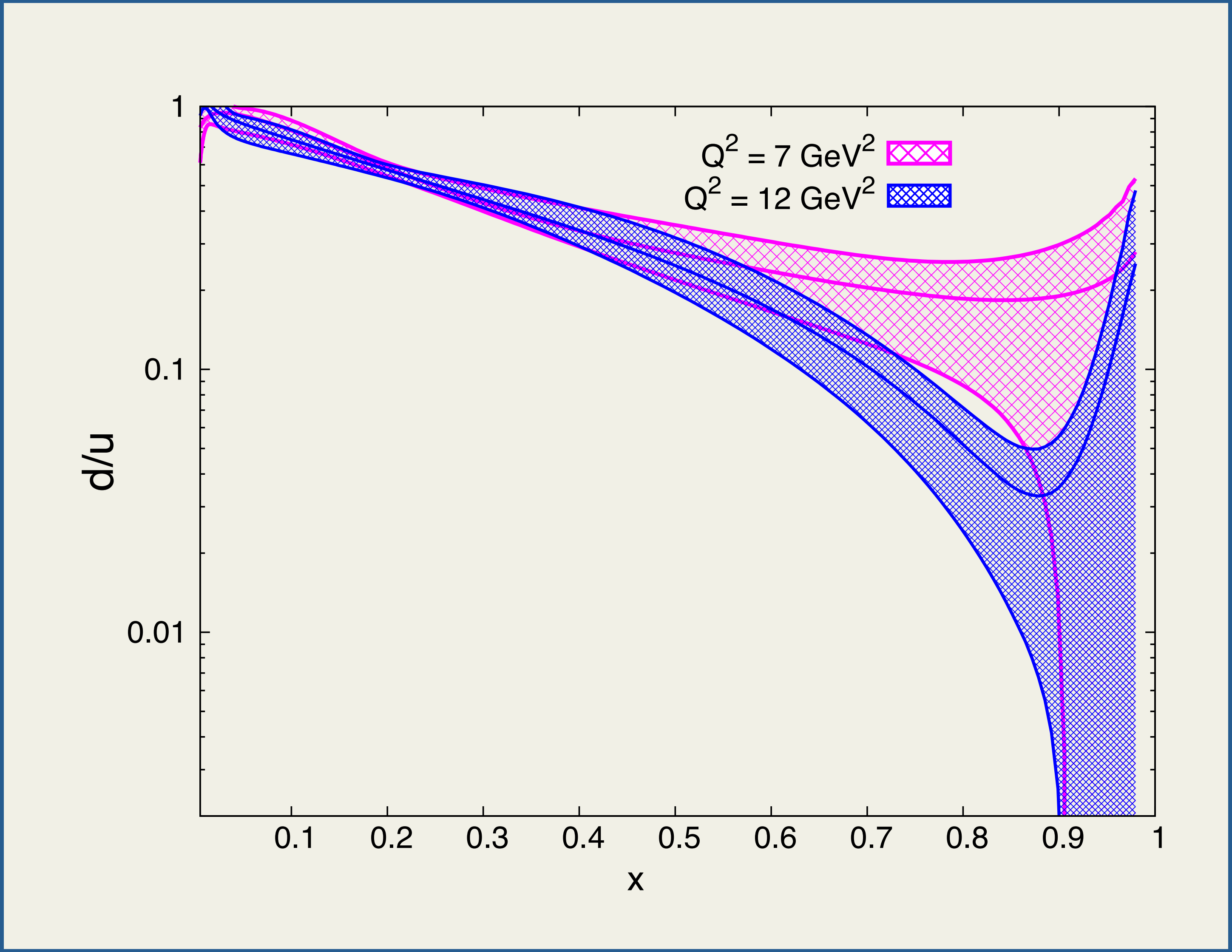
b quarks

c quarks

uds quarks



Initial application: d/u ratio



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)

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

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

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

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

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

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

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

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

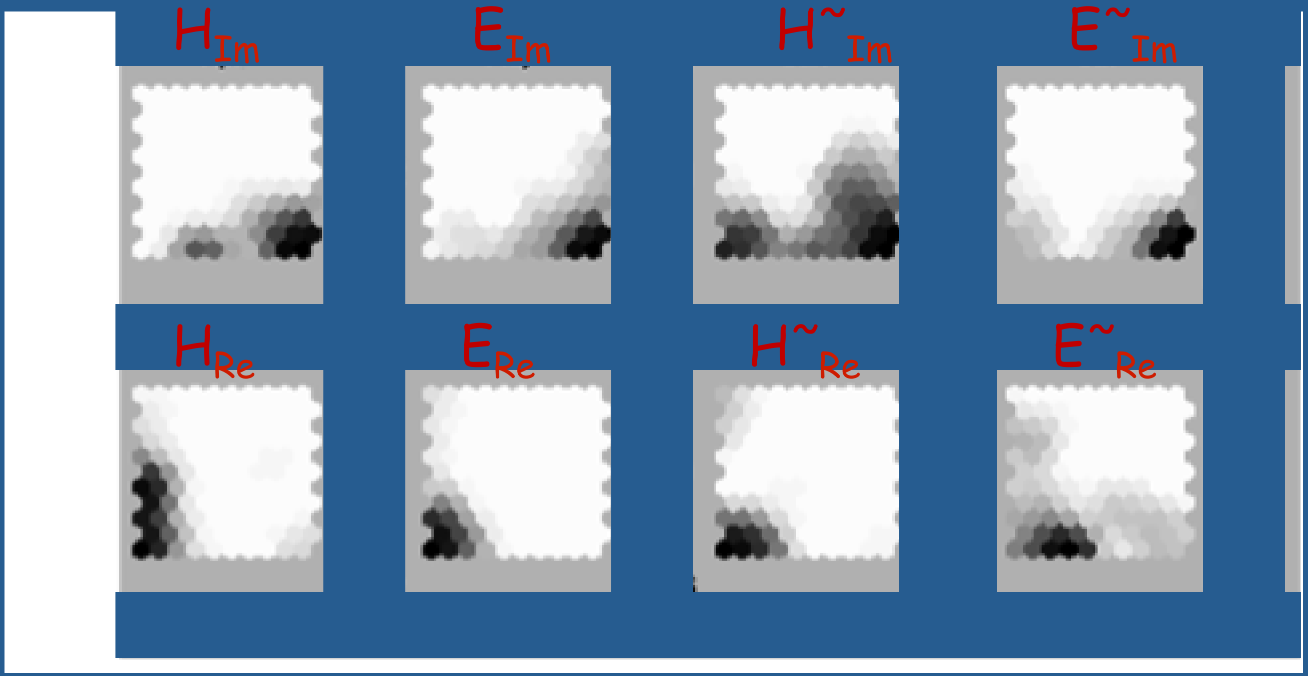
8 GPD-related functions

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

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

✓ We presented a new computational method,

Self-Organizing Maps

for parametrizing nucleon PDFs

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

✓ In progress: applications to more varied sets of data where predictivity is important (polarized scattering, $x \rightarrow 1$, …)

✓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