Self Organizing Maps Parametrization of Parton Distribution Functions

Simonetta Liuti
&
Katherine Holcomb

University of Virginia

ACAT 2011, Uxbridge, London
September 5-9, 2010

Outline

- *Introduction*
- *Algorithm*
- *SOMPDFs*
- *Comparison with NNPDFs*
- *Future Work: Extension to GPDs*
- *Conclusions/Outlook*

History/Organization of work

2005 An interdisciplinary group - Physics/Computer Science - was formed in order to investigate new computational methods in theoretical particle physics (NSF)

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)***

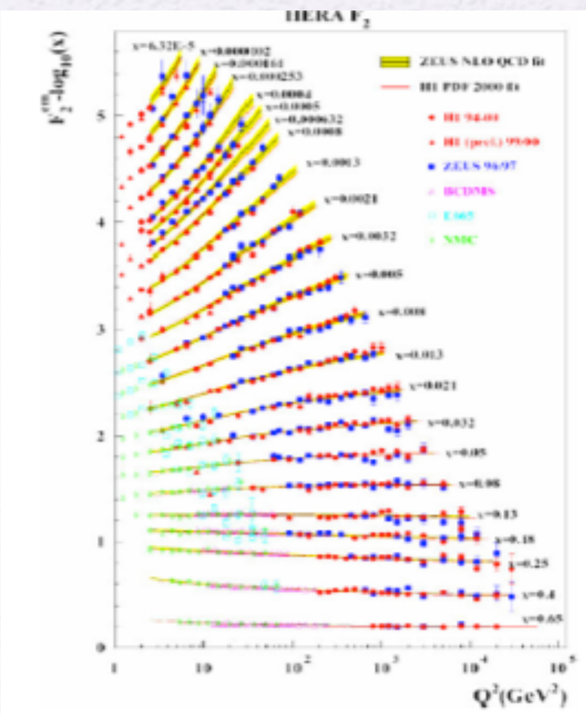
2009 New group formed (K. Holcomb, D. Perry, S. Taneja + Jlab) Rewriting, reorganization and translation of First Code into a uniform language, fortran 95.
2010 Implementation of Error analysis. Extension to new data analyses.
2011 PDF Parametrization Code ready to be released- SOMPDF.1

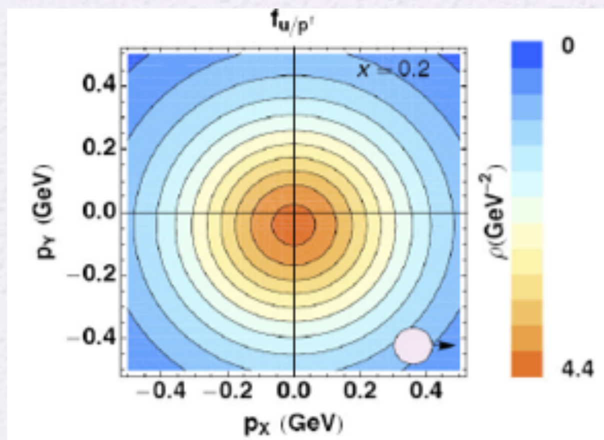
Group Website: http://faculty.virginia.edu/sompdf/

Introduction

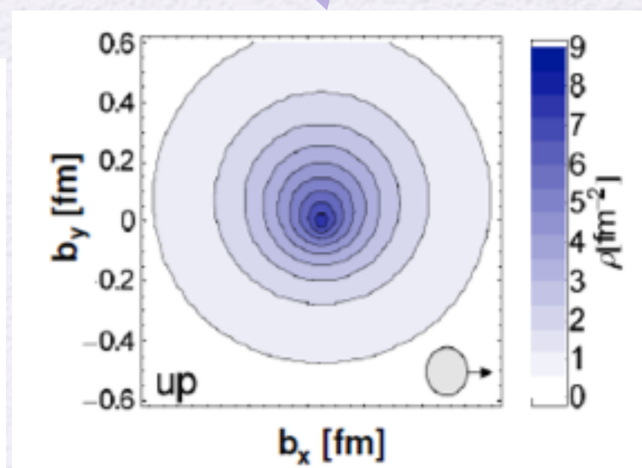
✓ The study of hadron structure in the LHC era an beyond (!) involves a large set of increasingly complicated and diverse observables

Parton Longitudinal Momentum Distribution Functions (PDFs),
Parton Transverse Momentum Distributions (TMDs),
Generalized Parton Distributions (GPDs),
Fragmentation Functions (FFs)
Fracture Functions (FFs)...



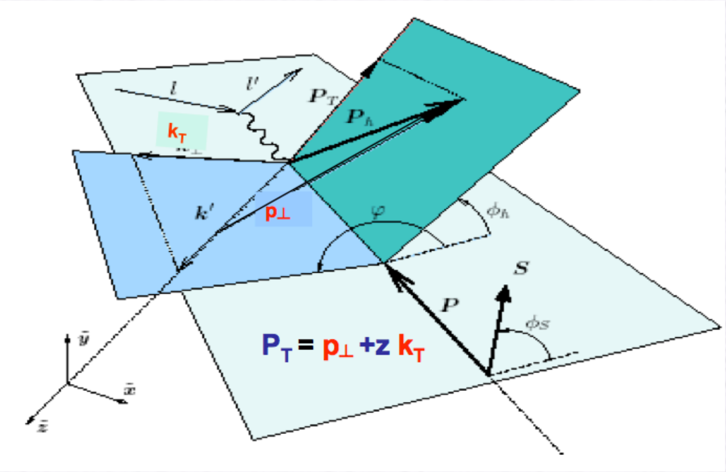


A.B., F. Conti, M. Radici, PRD78 (08)



QCDSF/UKQCD, PRL 98 (07)

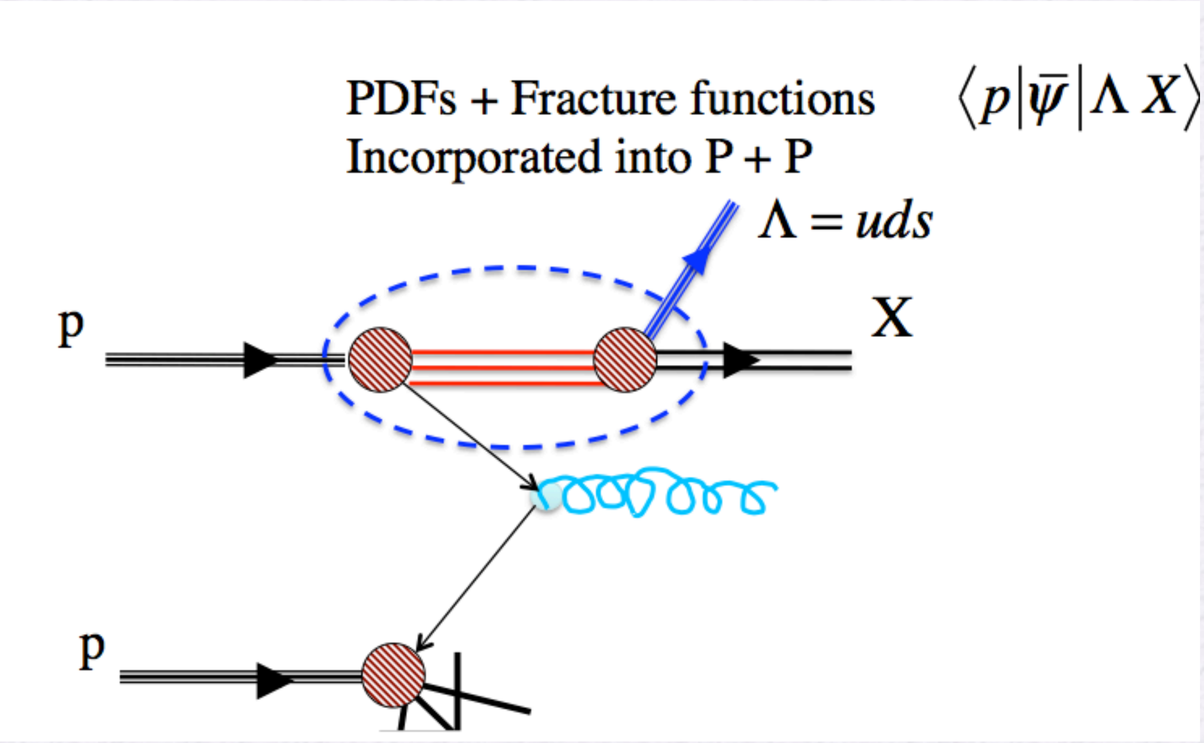
Experimental observations allow us to study the hadrons momentum, spin, spatial distributions, and their correlations



Example: Semi-Inclusive DIS

Conventional models give interpretations in terms of the microscopic properties of the theory (based on two-body interactions).

Example: pp ➔ΛX



- We now attack the problem from a different perspective:

Study the behavior of multiparticle systems as they evolve from a
large and varied number of initial conditions.

- This goal is at reach with HPC

The Use of Neural Networks in Data Analysis

- ✓ Neural Networks (NN) have been widely applied for the analysis of HEP data and PDF parametrizations (Cerutti'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 via a cost/object function.

Cost function measures the importance to detect or miss a particular occurrence

Example: If all patterns have equal probability, then the cost of predicting pattern S_i instead of S_k is simply

$$C(S_i, S_k) = 1 - \delta_{ik}$$

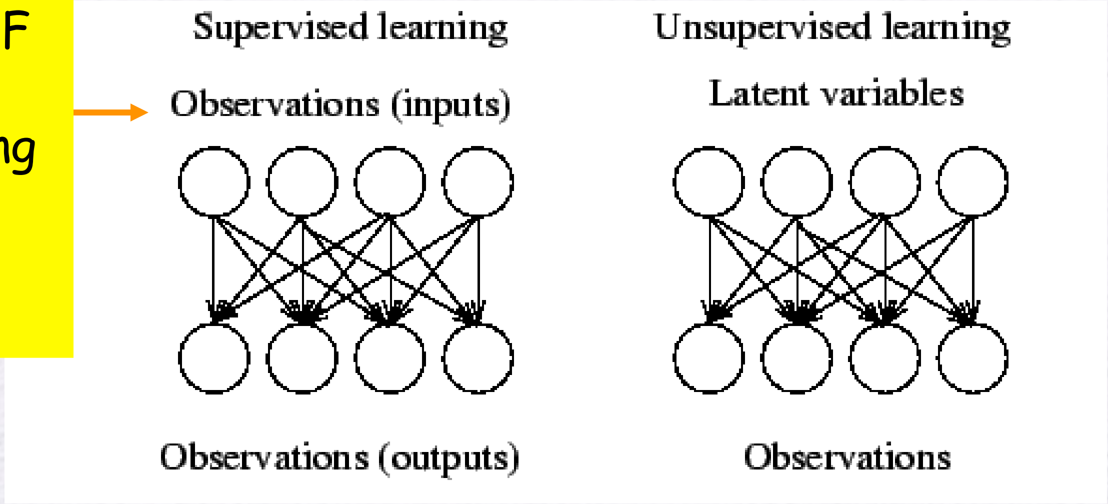
In general the aim is to minimize the cost

Most NNs (including NNPDFs) learn with supervised learning

Supervised Learning → A set of examples is given.
The goal is to force the data
To match the examples as closely as possible.
The cost function includes information about the domain

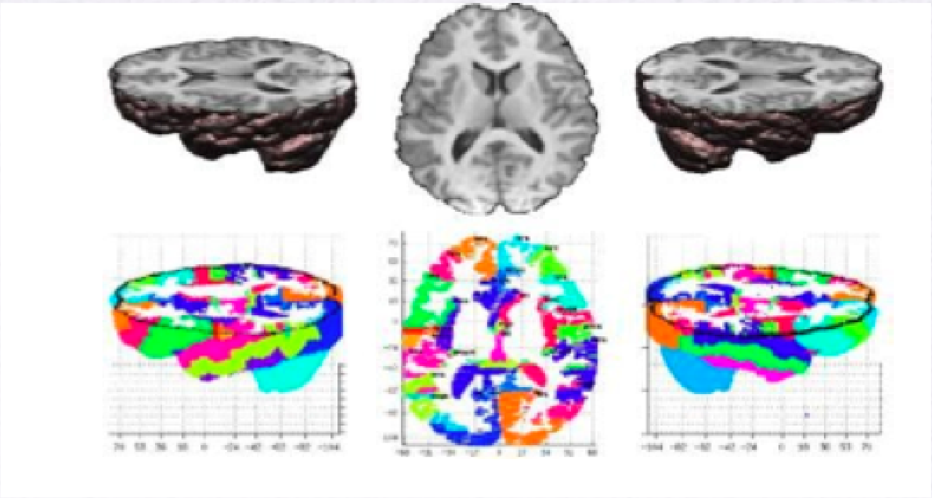
Unsupervised Learning → No a priori examples are given.
The goal is to minimize the cost function by similarity relations, or by finding how the data cluster or self-organize
➔ global optimization problem

Important for PDF analysis!
If data are missing it is not possible to determine the output!



SOMs in a nutshell

SOMs were developed by T. Kohonen in ‘80s (T. Kohonen, Self-Organzing Maps, Springer, 1995, 1997, 2006)



Inspired by the patterns in cerebral Cortex ➔ associative memory is based on the topographical order of neural connections forming localized maps

SOMs are a type of neural network whose nodes/neurons -- map cells --are tuned to a set of input signals/data/samples according to a form of adaptation (similar to regression).

Principles:

1) The neurons behave according to a form of
unsupervised self-organization

2) The representation of knowledge assumes the form of a map
geometrically organized over the brain so that
similar learning functions are associated to adjacent areas

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.

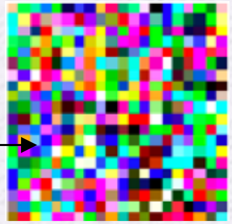
SOMs Algorithm

Each cell (neuron) is sensitized to a different 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:

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

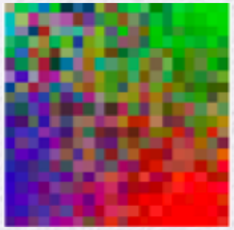
isomorphic

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

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

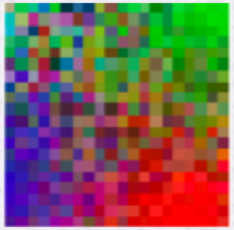
(Aggawal et al., 2000)

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

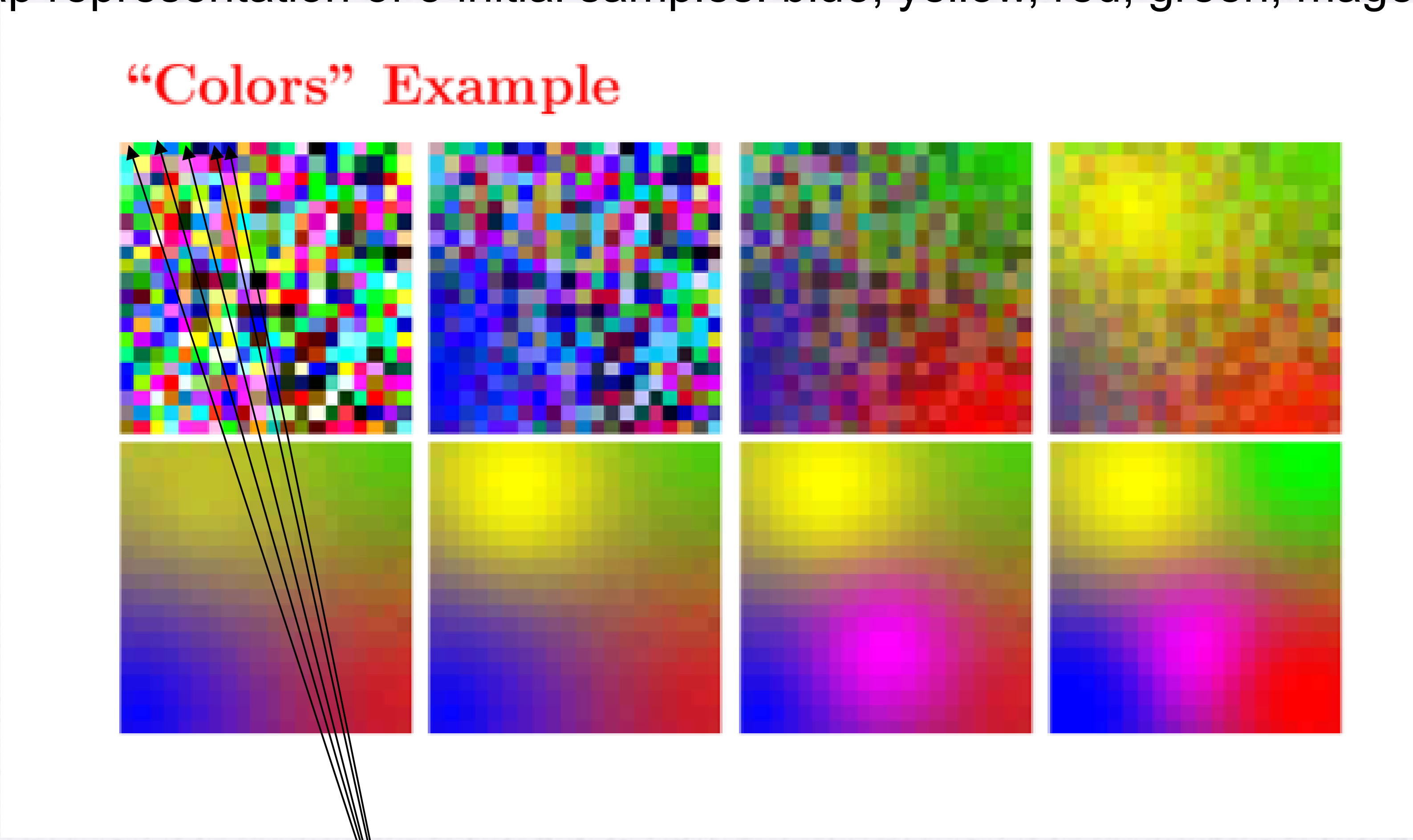
$$V_i(n+1) = V_i(n) + h_{ci}(n)\left[x(n) - V_i(n)\right]$$

iteration number

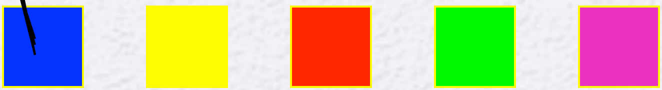
$$h_{ci}(n) = f(\|r_c - r_i\|) \equiv \alpha(n)\exp\left(\frac{-\|r_c - r_i\|^2}{2\sigma^2(n)}\right)$$

neighborhood function decreases with "n" and "distance"

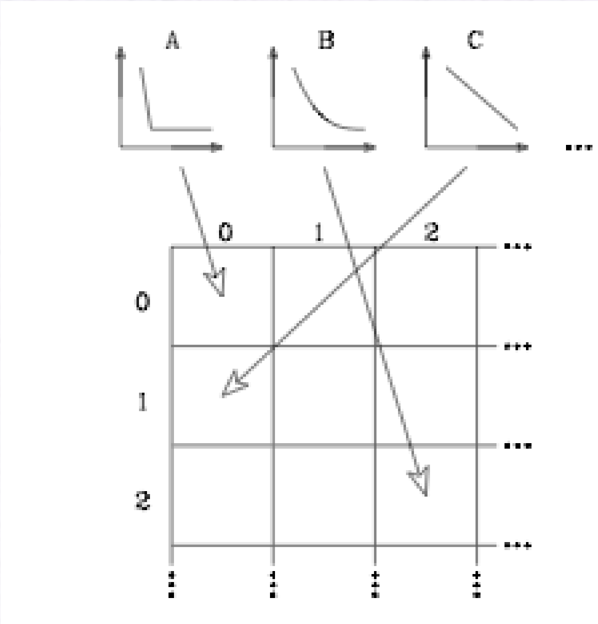
Map representation of 5 initial samples: blue, yellow, red, green, magenta

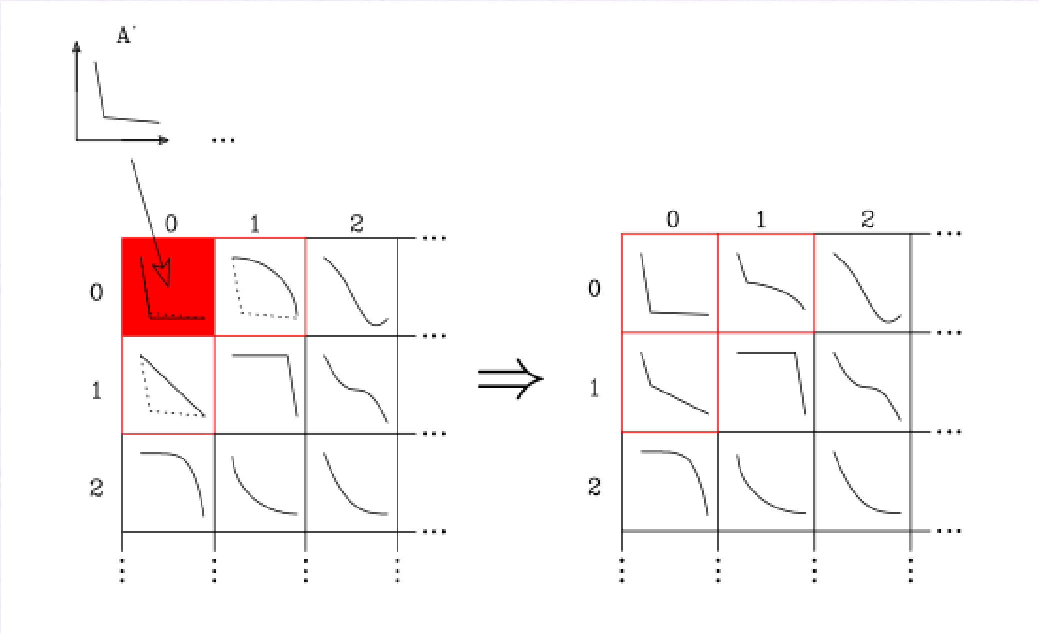


V_i



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

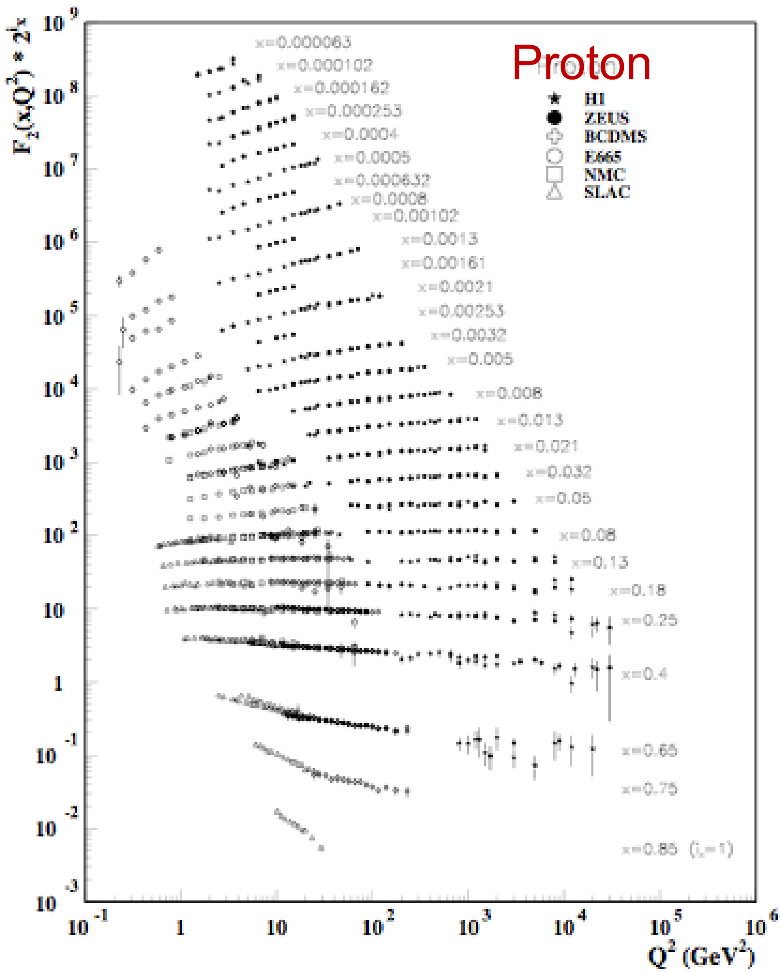
SOMPDFs

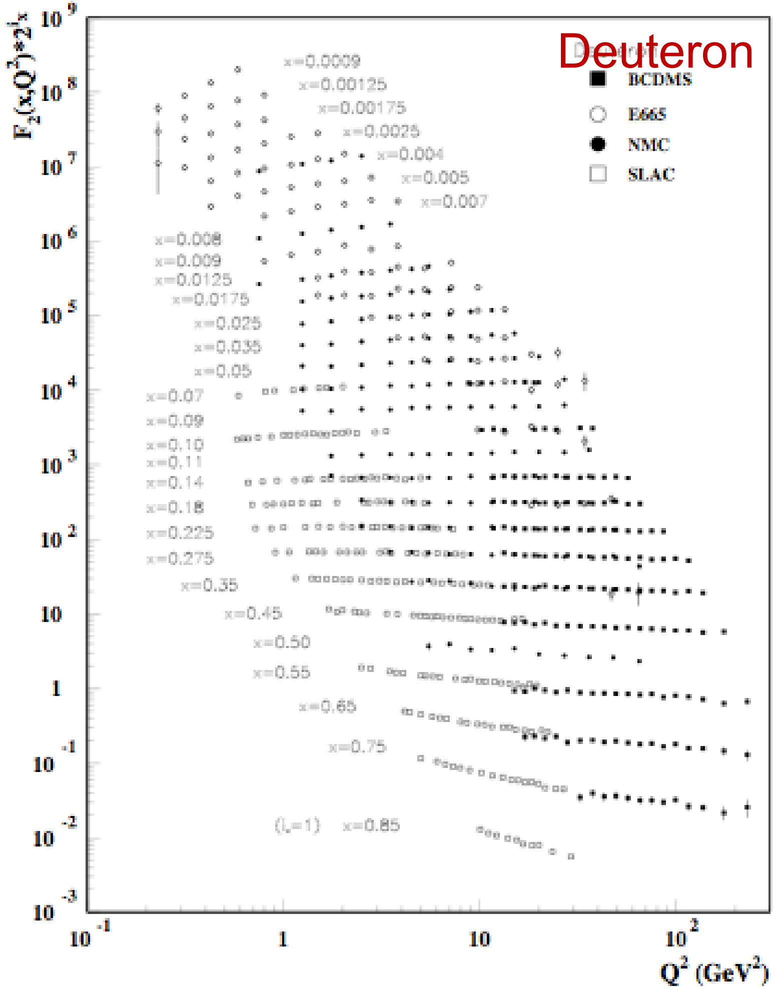
SOMPDF.0

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

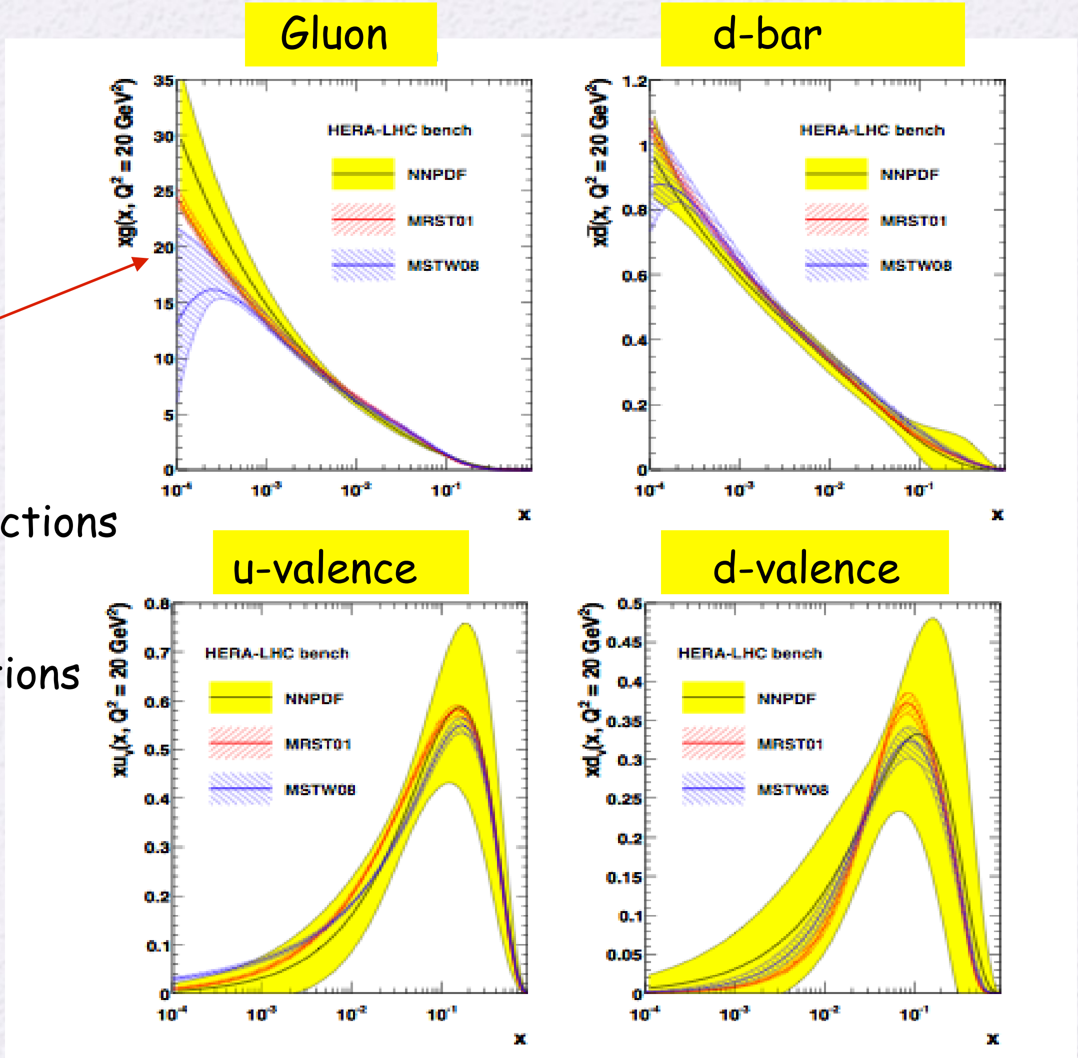
SOMPDF.1,

K. Holcomb, S.L., D.Z.Perry, hep-ph (2010)





Main issue



Uncertainties from
different PDF evaluations/extractions
(Δ_{PDF})
are smaller than the
differences between the evaluations
(Δ_G)

$\Delta_{PDF} < \Delta_G$

Studies such as M. Dittmar et al., hep-ph 0901.2504
define 3 benchmarks 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, Thorne)
3) Comparison of H12000 and NNPDF 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!

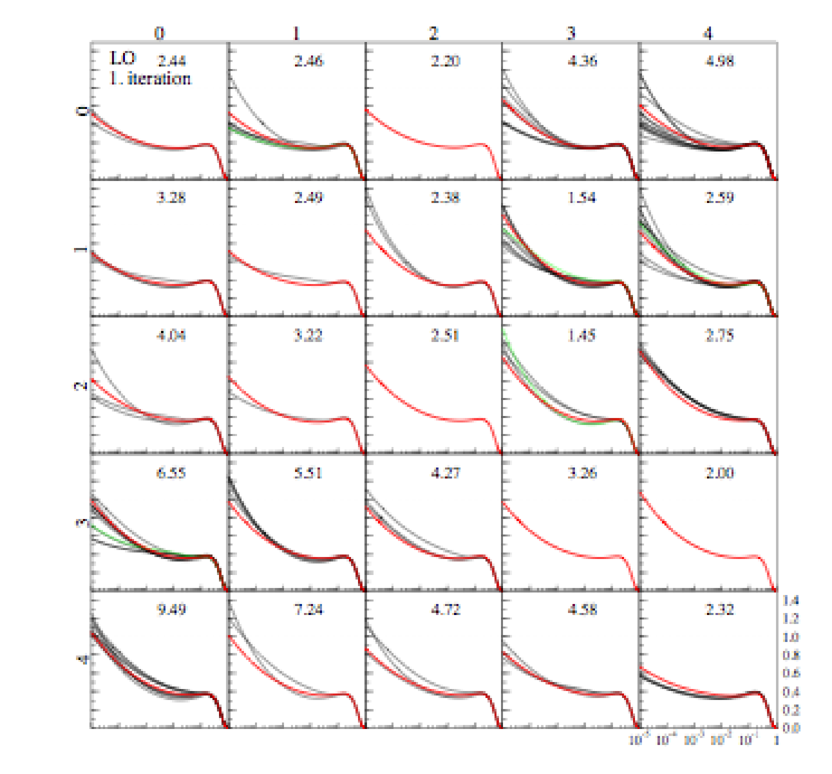
SOMPDF Method

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

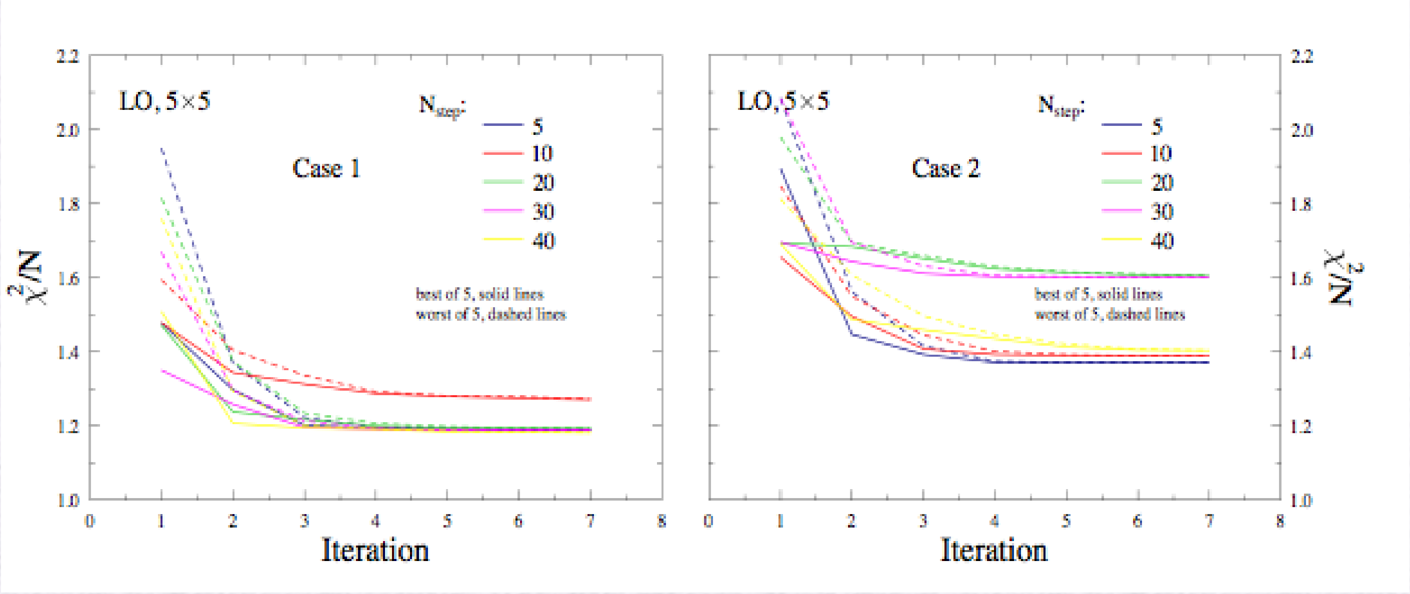
- In generating the PDFs (for the map and for the training) we need to avoid introducing a functional bias
- Thus we mix together variations of different structure functions
  - Random perturbations are used to generate a variant of a standard set of structure functions—currently based on GRV, MRST, 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.



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.



Advantages with respect to "conventional way":

- Initial scale ansatz

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

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

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

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.

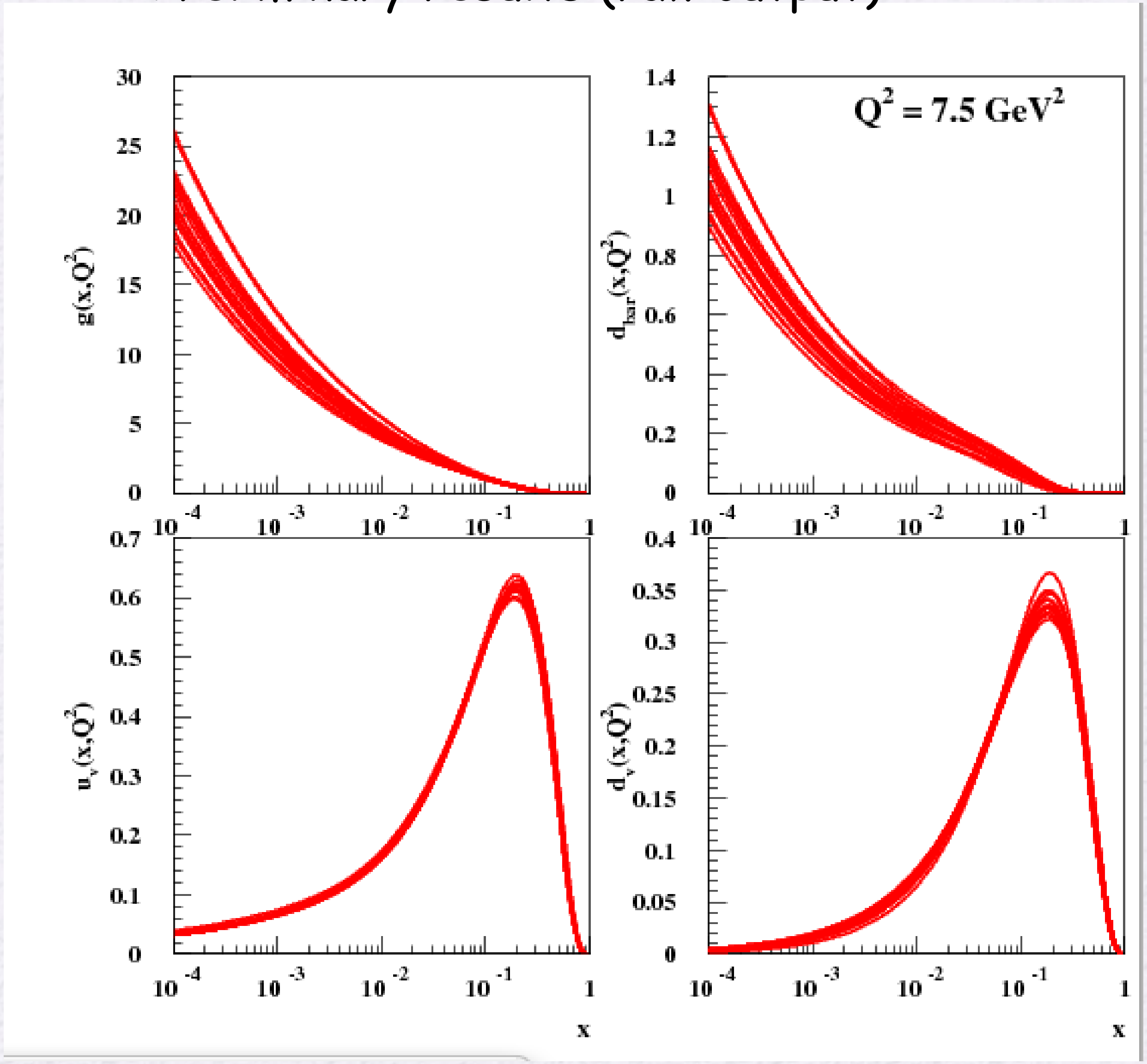
SOM work differently from ANN that do not keep track of the inter-connections among clustering of data at different stages of the network training.

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 defined a statistical error on an ensemble of SOMPDF runs
- Additional evaluation using Lagrange multiplier method is in progress

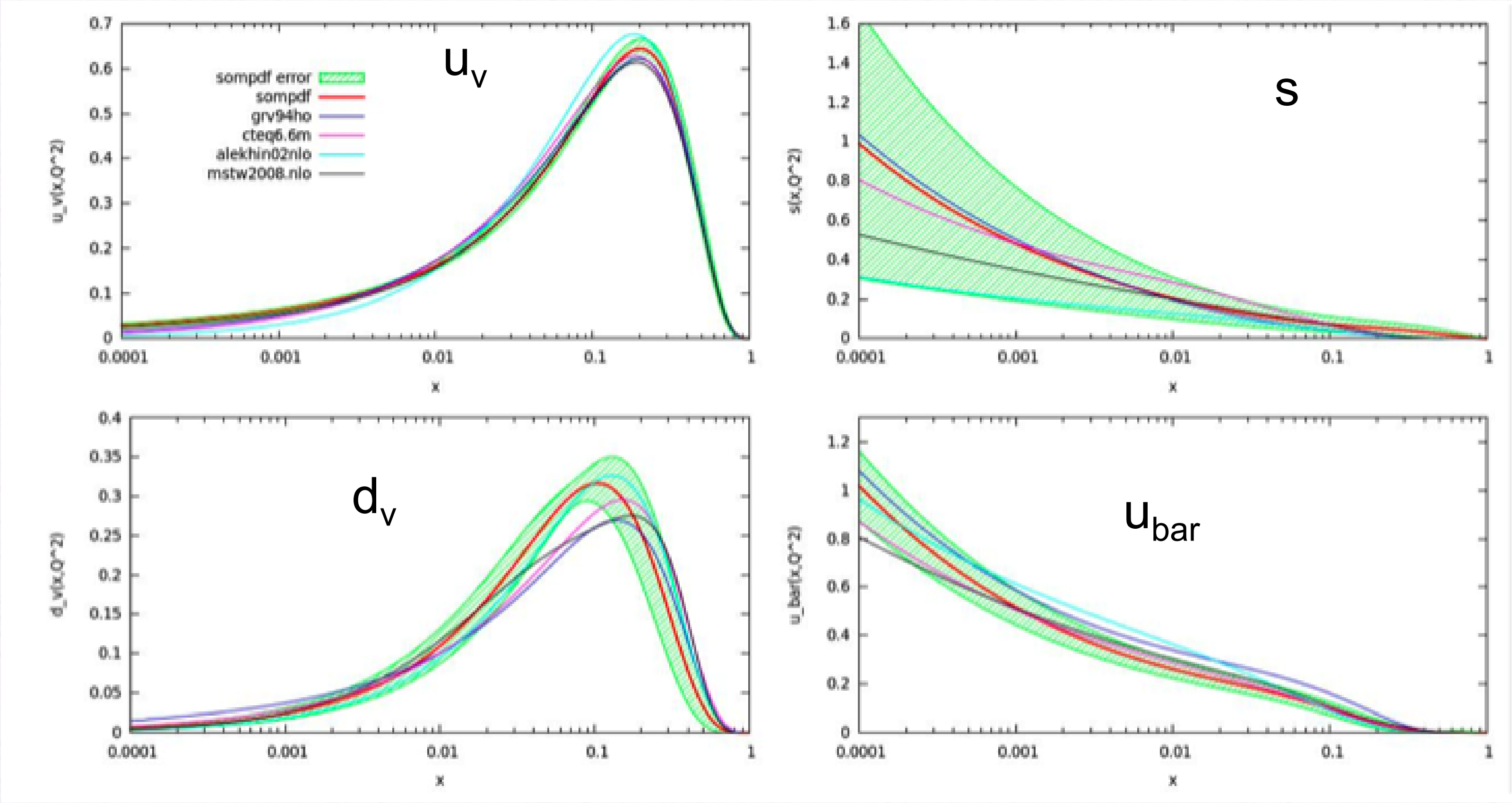
Preliminary results (raw output)



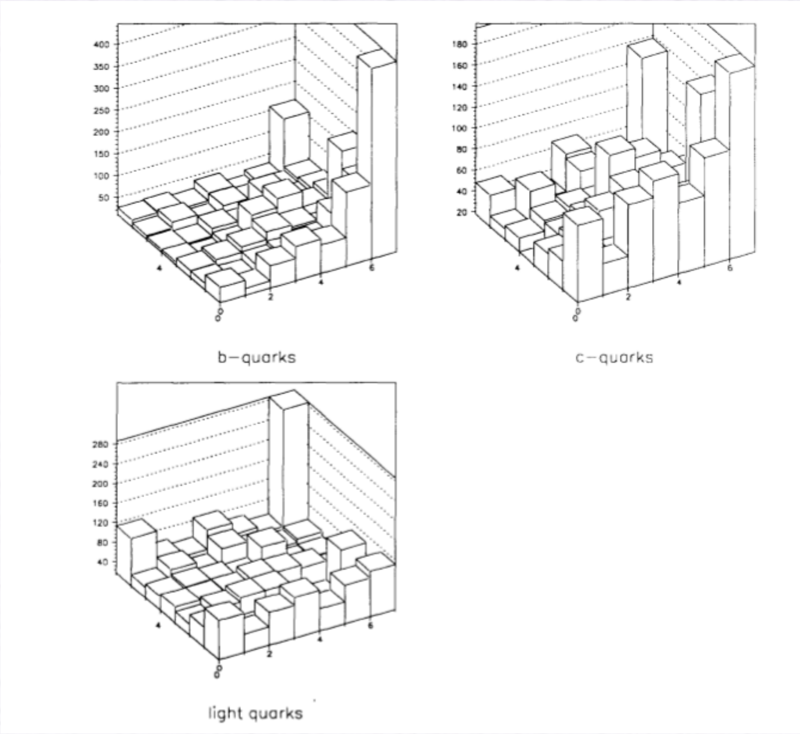
Preliminary Results

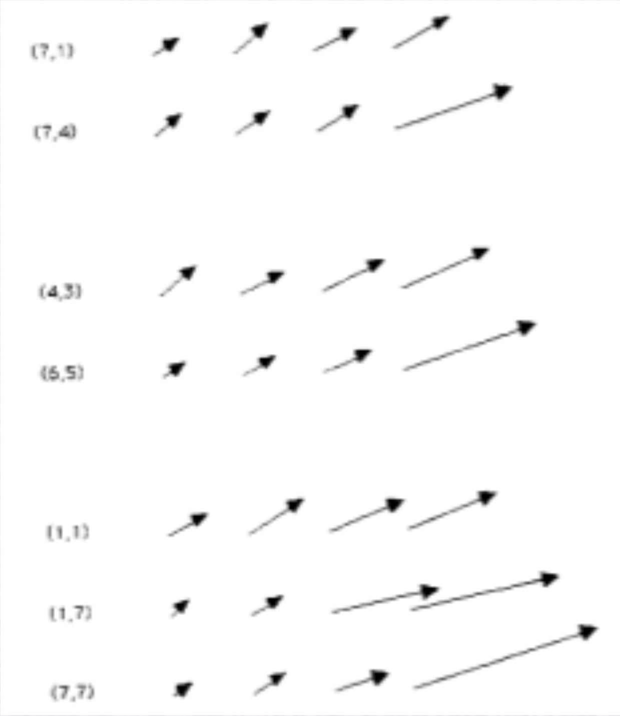
(D. Perry, DIS 2010 and MS Thesis 2010, K. Holcomb, Exclusive Processes Workshop, Jlab 2010)

$Q^2 = 7.5\ GeV^2$



Extension to multidimensional parton distributions/multiparton correlations: jet physics example (Lonnblad, Peterson et al., 1991)





b

c

u,d,s

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)

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

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

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

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

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

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

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

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

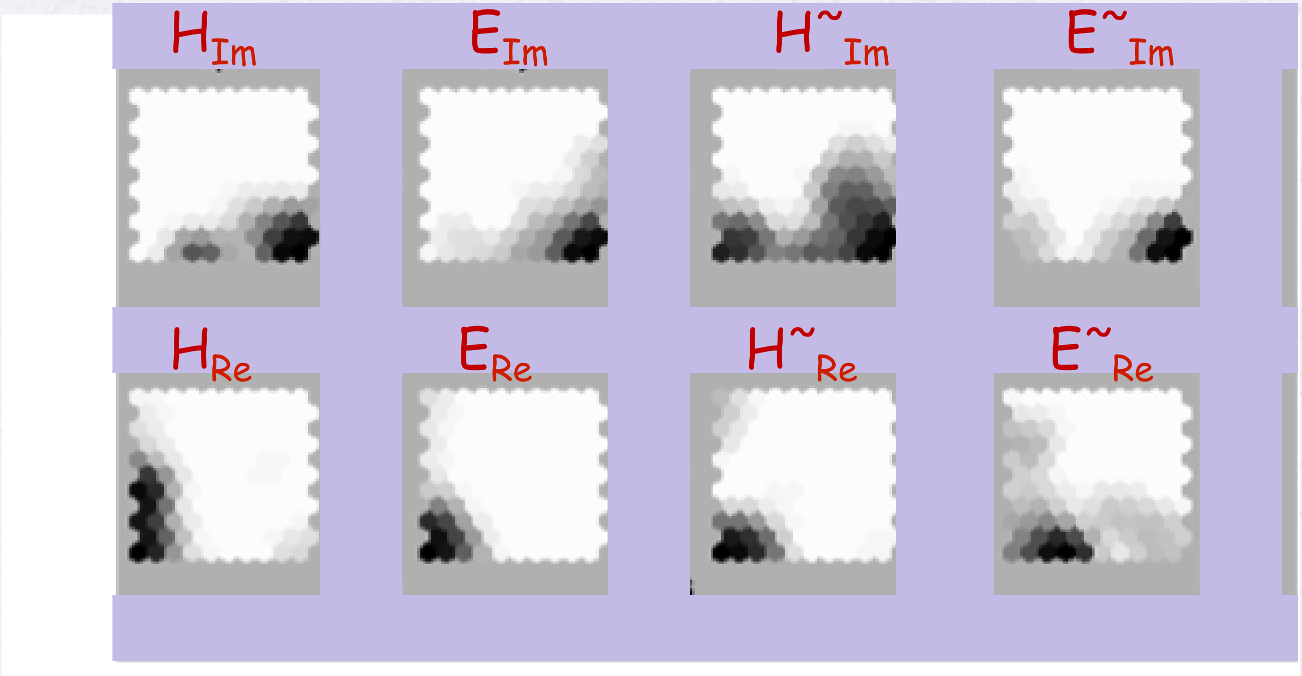
8 GPD-related functions

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

17 obsvervables (6 LO) from HERMES + Jlab data

"a challenge for phenomenology…" (Moutarde) + "theoretical bias"

The 8 GPDs are the dimensions in our analysis



Work in progress…

Conclusions/Outlook

✓ We presented a new computational method,

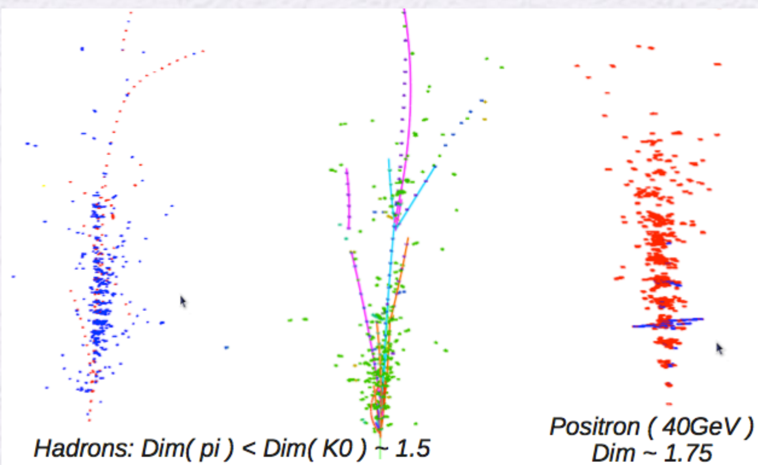
Self-Organizing Maps

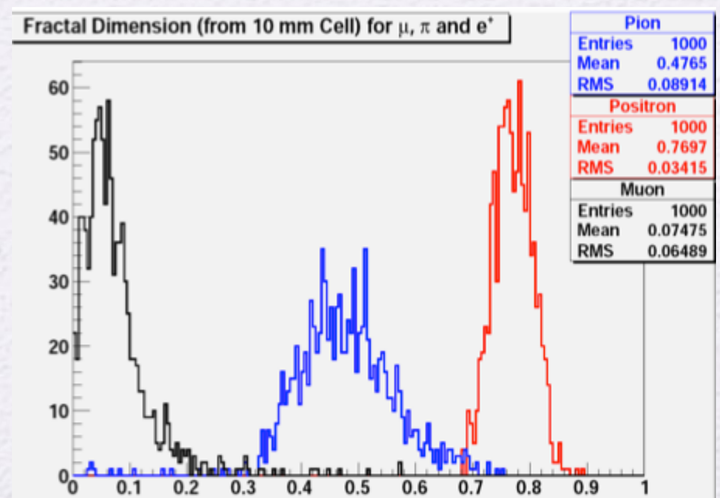
for parametrizing nucleon PDFs

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

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

✓ More distant Future: apply to GPDs, theoretical developments, connection with "similar approaches", complexity theory…





M. Ruan, this workshop