Markov Chain Monte Carlo technics applied to Parton Distribution Functions determination: proof of concept

Mariane Mangin-Brinet, Gabin Gbedo

Laboratoire de Physique Subatomique et de Cosmologie (LPSC/CNRS), Grenoble, France

Deep Inelastic Scattering 2017 - Birmingham - April 4th, 2017

Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

Uncertainties in global PDFs analysis

Precise assessment of PDFs uncertainties has become crucial

PDFs subject both to theoretical (input parametrization, neglected higher order corrections, heavy flavors treatment,...) and experimental uncertainties.

To deal with experimental errors:

- Hessian method: based on linear error propagation
- Lagrange multiplier method
- ...

Assumption on the permissible range of "acceptable" $\Delta\chi^2$ for the fit and choice of a tolerance parameter T [1].

[1] This excepts Neural Networks technics

Traditional propagation of experimental uncertainties

- $\Delta\chi^2 = 1$ for 68% C.L. IF fitting consistent data sets with ideal gaussian errors
- In practice: inconsistencies between fitted data sets and unknown exp. and theoretical uncertainties, so not appropriate for global PDFs analysis $\Longrightarrow$ choice of tolerance criteria $\Delta\chi^2 \neq 1$.

Ex.: MRST [hep-ph/0211080]

"We estimate $\Delta\chi^2 = 50$ to be a conservative uncertainty (perhaps of the order of a 90% confidence level or a little less than 2σ) due to the observation that an increase of 50 in the global χ^2 [...]usually signifies that the fit to one or more data sets is becoming unacceptably poor. We find that an increase $\Delta\chi^2$ of 100 normally means that some data sets are very badly described by the theory."

What can MCMC methods tell us about PDFs uncertainties?

Traditional propagation of experimental uncertainties

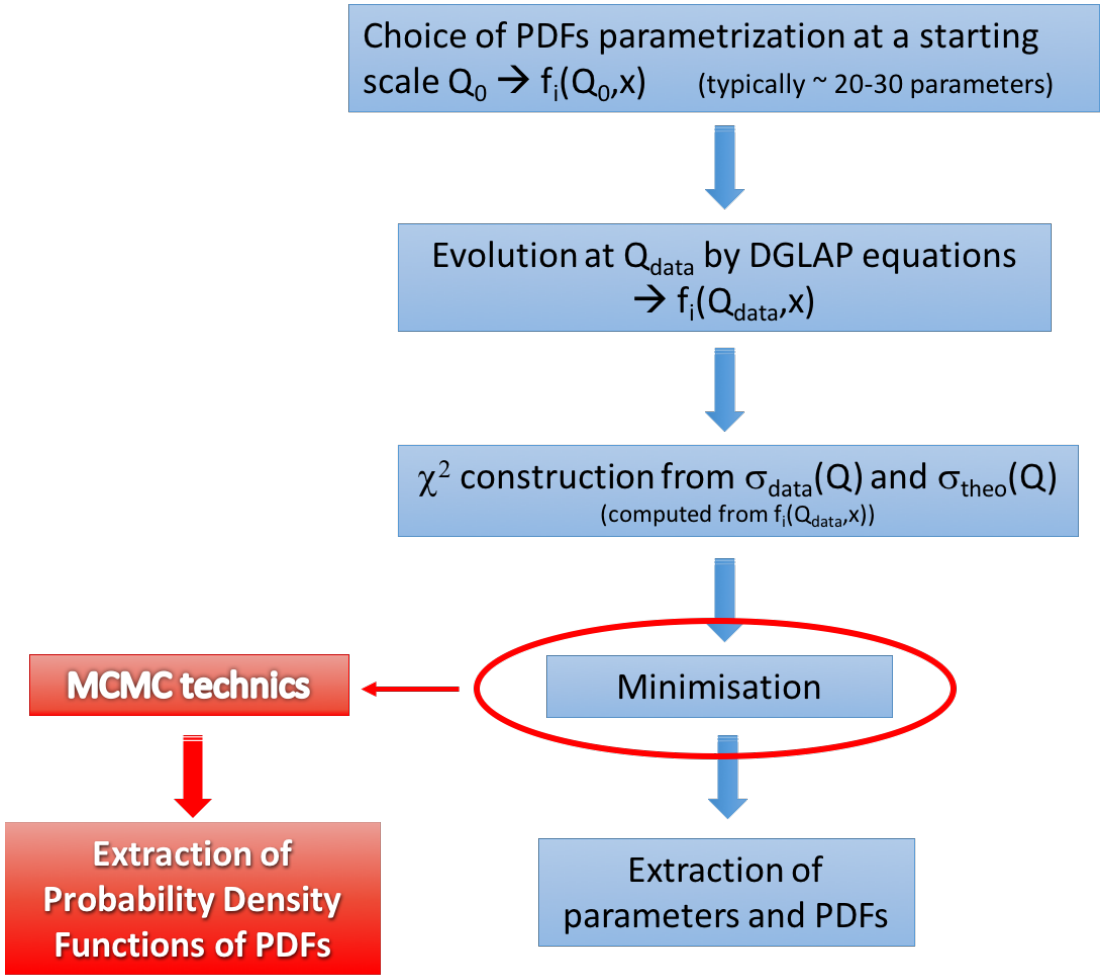
- $\Delta\chi^2 = 1$ for 68% C.L. IF fitting consistent data sets with ideal gaussian errors
- In practice: inconsistencies between fitted data sets and unknown exp. and theoretical uncertainties, so not appropriate for global PDFs analysis $\Longrightarrow$ choice of tolerance criteria $\Delta\chi^2 \neq 1$.

Ex.: MRST [hep-ph/0211080]

"We estimate $\Delta\chi^2 = 50$ to be a conservative uncertainty (perhaps of the order of a 90% confidence level or a little less than 2σ) due to the observation that an increase of 50 in the global χ^2 [...]usually signifies that the fit to one or more data sets is becoming unacceptably poor. We find that an increase $\Delta\chi^2$ of 100 normally means that some data sets are very badly described by the theory."

What can MCMC methods tell us about PDFs uncertainties?

Very general flow diagram of PDFs extraction



Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

PDFs determination in terms of Bayesian inference

- $\vec{q}$: vector of PDFs parameters to be determined:
  $\vec{q} = (q^{(1)}, q^{(2)}, \ldots, q^{(m)})^T$

  xFitter functional form at $Q_0^2 \sim 2$ GeV^2:

  $x f_a(x) = A_a\, x^{B_a}\, (1-x)^{C_a}\, (1 + D_a x + E_a x^2)$. where a labels a parton (g, u_{val}, d_{val}, ...).

- D: data

Bayesian inference:

Both model parameters and observables considered random quantities.
Aims at determining a joint probability distribution $P(D, \vec{q})$ over all random quantities:

$$P(D, \vec{q}) = P(D|\vec{q})P(\vec{q})$$

- $P(\vec{q})$ *prior* distribution
- $P(D|\vec{q})$: likelihood of the data $\mathcal{L}(\vec{q})$

Bayes theorem

Express $P(\vec{q}|D)$ in terms of the likelihood $P(D|\vec{q})$:

$$P(\vec{q}|D) = \frac{P(D|\vec{q})P(\vec{q})}{\int d\vec{q} P(D|\vec{q})P(\vec{q})} \tag{1}$$

- $P(\vec{q}|D)$: *posterior* probability density
- Can be sampled using a Monte Carlo algorithm

PDFs determination in terms of Bayesian inference

Likelihood

Let's note

- D_i and T_i: respectively the i^{th} experimental point and the corresponding theoretical calculation
- σ_i^2 the uncertainty associated with the measured data i.

If $\frac{(D_i - T_i)}{\sigma_i}$ independent and normally distributed:

$$\log \mathcal{L}(\hat{q}) = -\frac{1}{2} \sum_{i=1}^{n} \frac{(D_i - T_i)^2}{\sigma_i^2} = -\frac{1}{2}\chi^2 \tag{2}$$

Possibility to construct more involved likelihood with correlated data

Prior distribution

We choose $P(\vec{q})$ uniform

Remark: If D_i distributed according to a Gaussian law, MLE and LSM equivalent

In summary...

Least square method

- χ^2 construction
- Minimization

χ^2_{min} and neighborhood

MCMC method

- Choice of prior
- Likelihood construction
- Monte Carlo sampling of posterior probability

Probability distributions, confidence intervals...

Metropolis [Metropolis et al., J. Chem Phys. **21** (1953)]

- one of the simplest Monte Carlo algorithm
- standard computational workhorse of MCMC methods
- in principle applicable to any system
- extremely straightforward to implement and to sample a target density $P(\vec{q}|D)$
- at each Monte Carlo time t, the next state $\vec{q}_{t+1}$ is chosen by sampling a candidate point $\vec{q}\,'$ from a proposal distribution$^{(*)}$. The candidate point is then accepted with the probability

$$\alpha(\vec{q}_t, \vec{q}\,') = \min\left(1, \frac{P(\vec{q}\,'|D)}{P(\vec{q}_t|D)}\right)$$

i.e. in our case

$$\alpha(\vec{q}_t, \vec{q}\,') = \min\left(1, e^{-\frac{1}{2}\Delta\chi^2}\right)$$

$^{(*)}$ Proposal distribution assumed symmetric here

and why we do not use it.....

Initialize $\vec{q}_0$; set $t = 0$
Repeat {
Sample a point $\vec{y}$ from a proposal distribution
Sample a Uniform[0,1] random variable u
If $u \leq \alpha(\vec{q}_t, \vec{y})$ set $\vec{q}_{t+1} = \vec{y}$
otherwise set $\vec{q}_{t+1} = \vec{q}_t$
Increment t
}

Acceptance test
- typically for 1 parameter: 30-50%
- decreases as ~ 0.5^m
→ For m=10 parameters, unacceptable acceptance…

trial point far from the initial one $\Longrightarrow$ large change in the distribution to sample point close to the initial one $\Longrightarrow$ inefficient exploration of the parameter space

Metropolis algorithms (even flavored with multivariate Gaussian distributions or binary space partitioning) are NOT suited for realistic PDFs determination ($m \sim 25$).

HMC

- Developed originally for lattice field theory [Duane et al, 1987]
- Combines molecular dynamics evolution with a Metropolis accept/reject step

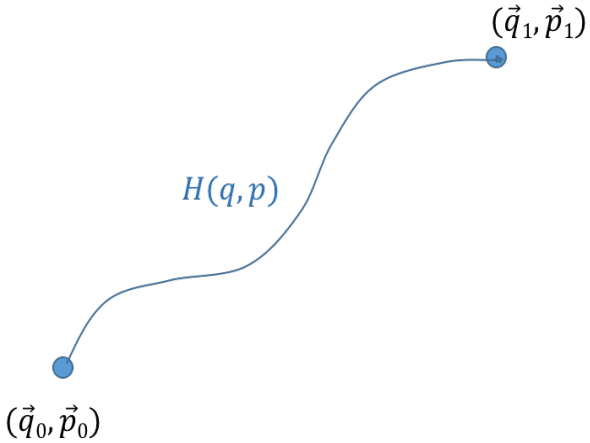
Introduce for each set of parameters $\vec{q}$ a set of conjugate momenta $\vec{p}$
Associates an Hamiltonian $H(\vec{q},\vec{p}) = \vec{p}^T M^{-1}\vec{p}/2 + \mathcal{U}(q)$, where M is a mass matrix, and $\mathcal{U}(q)$ an arbitrary potential energy.

This allows to define a joint distribution as

$$P(q,p) = \frac{1}{Z}e^{-H(q,p)} = \frac{1}{Z}e^{-\mathcal{K}(p)}e^{-\mathcal{U}(q)}$$

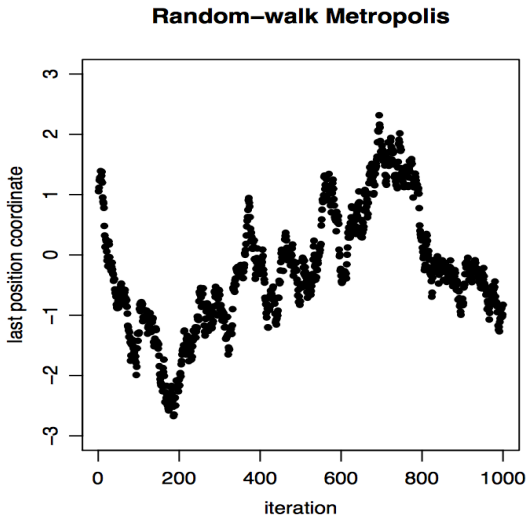
Z normalizing constant

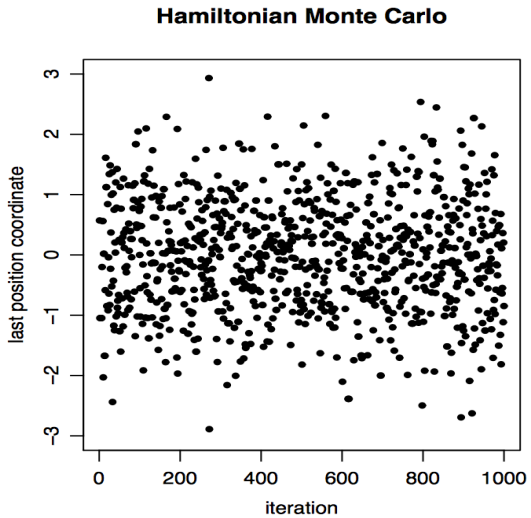
We use for the potential energy $\mathcal{U}(q) = -\log[P(D|\vec{q})P(\vec{q})]$.



- choose $\vec{p}_0$ normally distributed
- let the system evolve deterministically
- candidate point $\vec{q}_1$ accepted with probability $min(1, e^{-\Delta H})$ i.e. 100% !!!

In practice: acceptance degraded because of numerical resolution of Hamilton equations, but still very high (typically $\sim 80 - 90\%$, independently of the dimension of the chain).





Sampling of a 100D gaussian. Values of the variable with largest standard deviation.

[R. M. Neal, arXiv:1206.1901]

Implementation of HMC in xFitter package

- parametrized HERA PDFs: xu_v, xd_v, xg, $x\overline{U} = x\overline{u}$, $x\overline{D} = x\overline{d} + x\overline{s}$
- functional form at $Q_0^2 \sim 2$ GeV^2:

$$xf_a(x) = A_a\, x^{B_a}\, (1-x)^{C_a}\, (1 + D_a x + E_a x^2) \tag{3}$$

  where a labels a parton (g, u_{val}, d_{val}, ...).

- we consider 10 free parameters: B_g, C_g, $B_{u_{val}}$, $C_{u_{val}}$, $E_{u_{val}}$, $C_{d_{val}}$, $C_{\overline{U}}$, $A_{\overline{D}}$, $B_{\overline{D}}$ and $C_{\overline{D}}$
- validation study, with combined inclusive $e^{\pm}p$ scattering cross-sections from H1 and ZEUS ($\sim$ 600 data points), and ZMVFN scheme.
- comparison with HERAPDF1.0 with ZMVFNS.

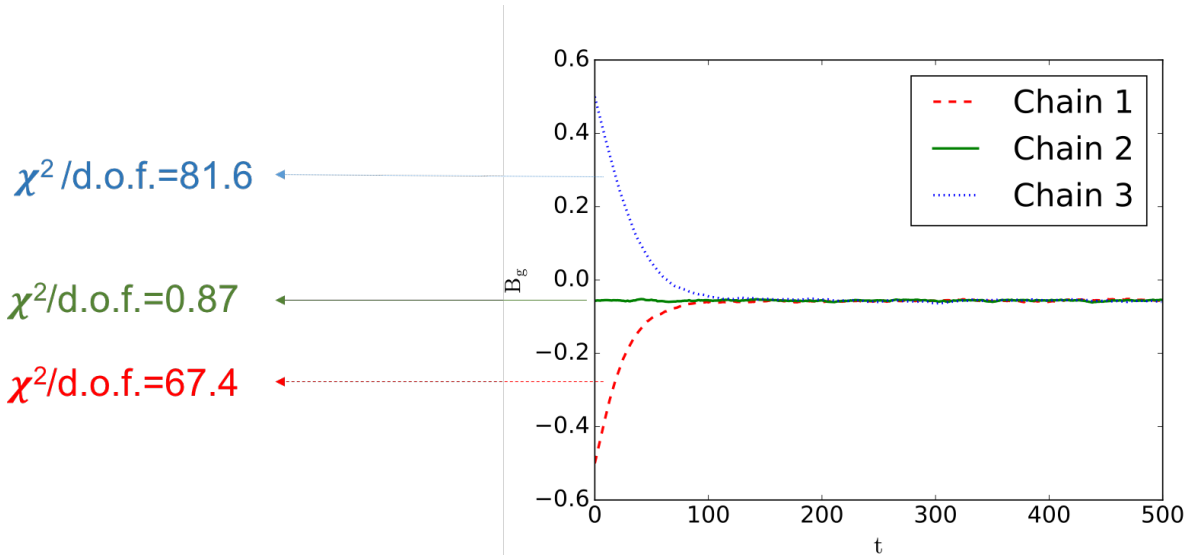
Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

Analysis (1/3)

Extracting observables and their statistical errors from Monte Carlo simulations: subtle task, requiring careful treatment of the Markov chain.

- The thermalization time (or burn-in length b) : number of states $\{\hat{q}_t\}_{t=1,\ldots,b}$ to be discarded from the beginning so that the chain forgets its starting point.



We have taken $P(\hat{q}_b|D) > P_{1/2}$

Analysis (2/3)

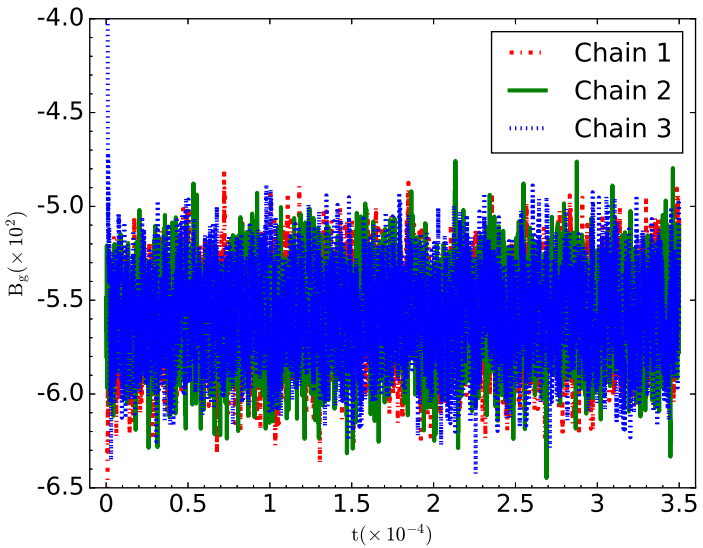
- Autocorrelation: inherent correlations from one point to the next.
  Usual estimate of root-mean-square deviation of an observable O:

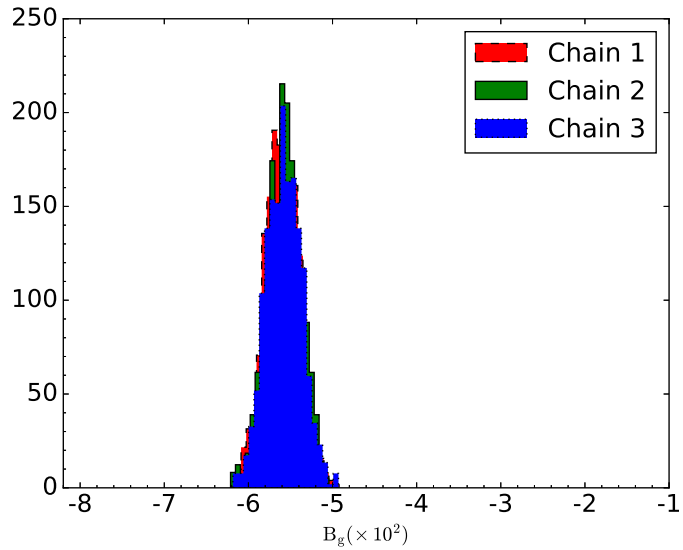
$$\sigma^2_{naive} = \frac{N}{N-1}\left(\langle O^2\rangle - \langle O\rangle^2\right)$$

  - Relies on the assumption that measurements performed on the Markov chain are NOT correlated
  - Several methods to account for the correlations: jackkniffe binning (pre-averaging over blocks of data), $\Gamma-$method (explicit determination of autocorrelation functions and times),...
  - in practice for this work: $\tau_{corr} \sim 2$.

Analysis (3/3)

- Convergence: start from different points, and check we reach the same distribution





Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

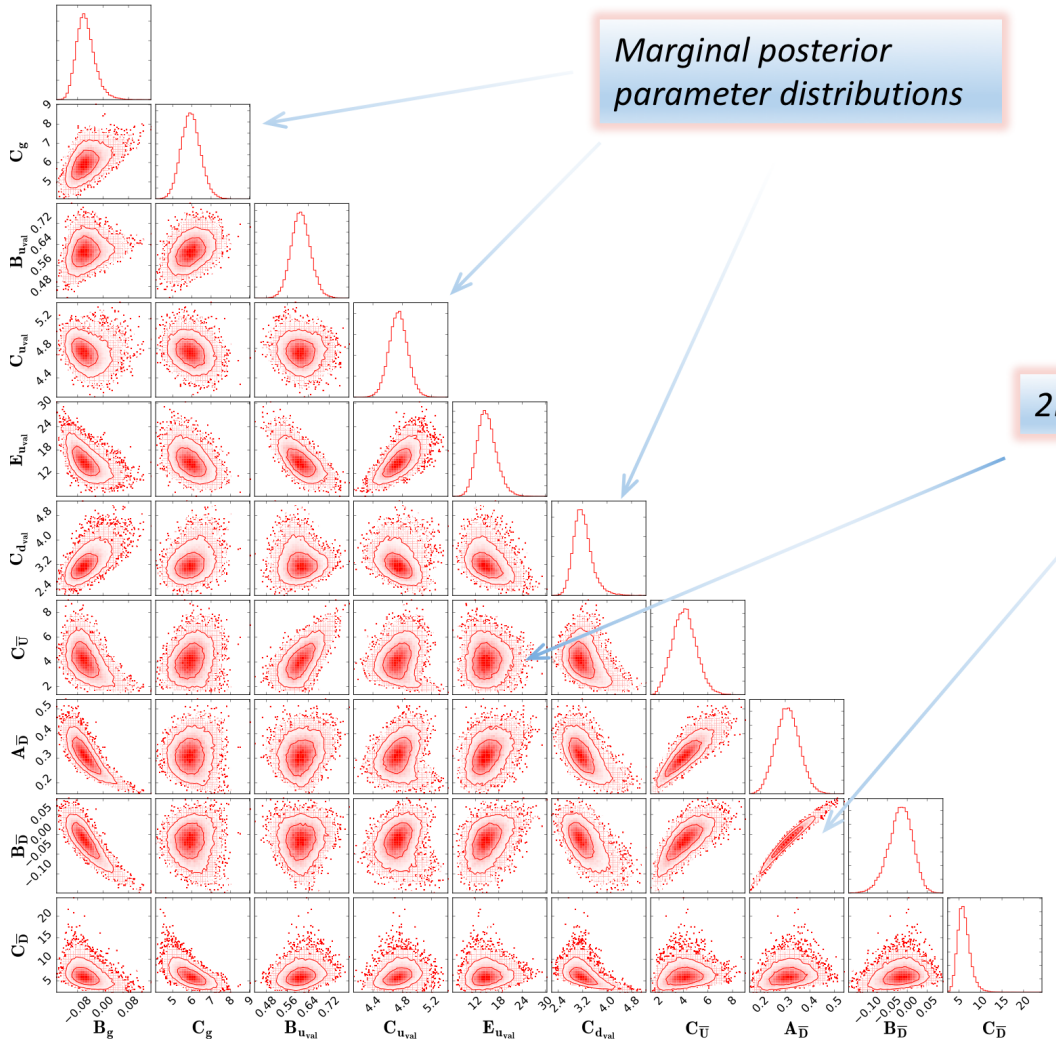
Ready to compare standard minimization and Monte Carlo method....

Comparison of results obtained for 2 parameters of the gluon PDF using Markov Chain procedure or χ^2 (MINUIT) minimisation:

parameter	values	MCMC	MINUIT minimization
$\mathbf{B_g}$	mean	-0.0537 ± 0.0002	
	most probable	-0.0632 ± 0.0168	-0.0559
	standard deviation	0.0299 ± 0.0001	0.0288
$\mathbf{C_g}$	mean	5.9483 ± 0.0025	
	most probable	5.8952 ± 0.0615	5.9274
	standard deviation	0.5037 ± 0.0019	0.5078

Much more information with MCMC

PDF Parameters: Marginal probability distribution and correlations

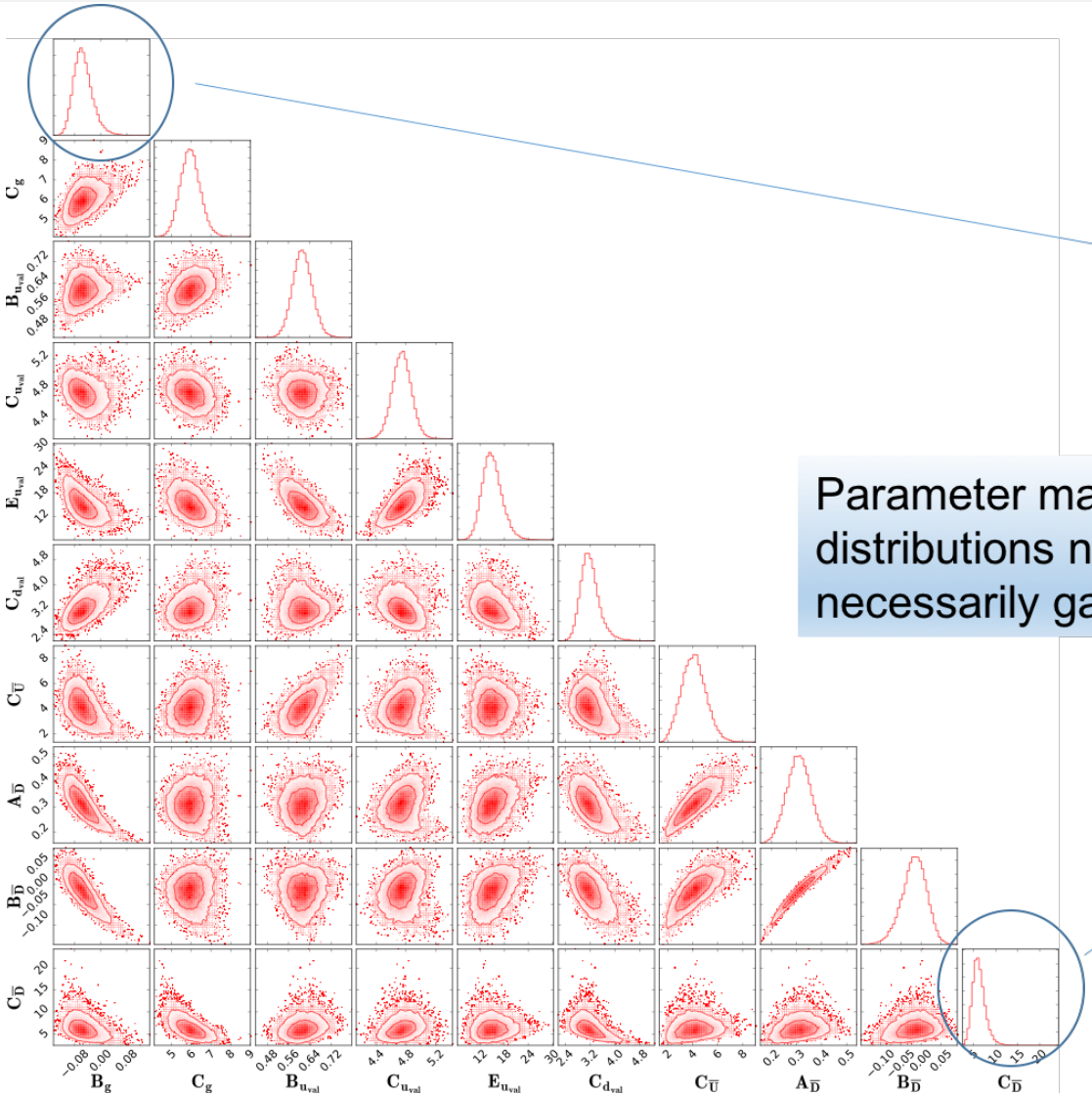


Marginal posterior parameter distributions

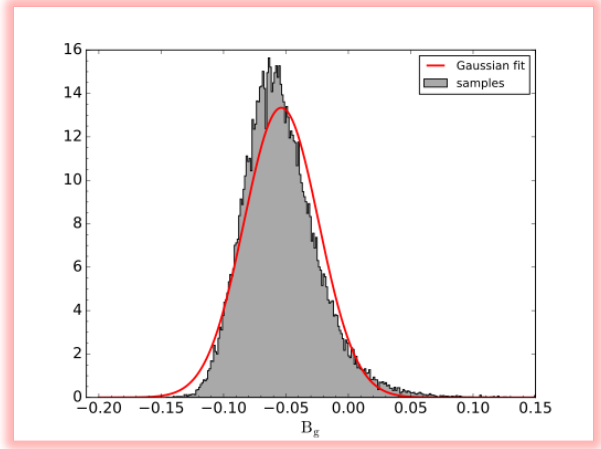
2D correlations

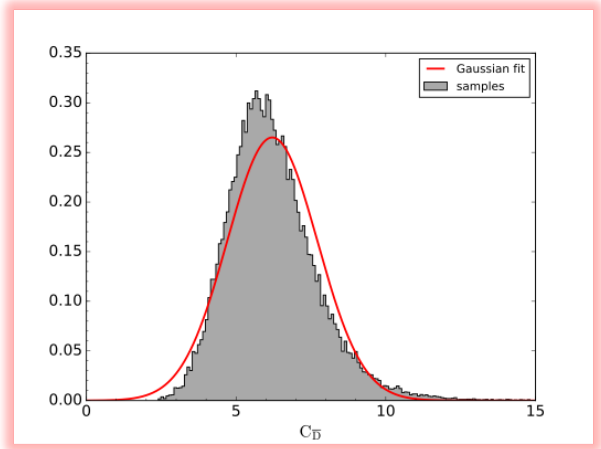
Inner and outer contours: 68% and 95% of the probability density resp.

PDF Parameters: Marginal probability distribution and correlations



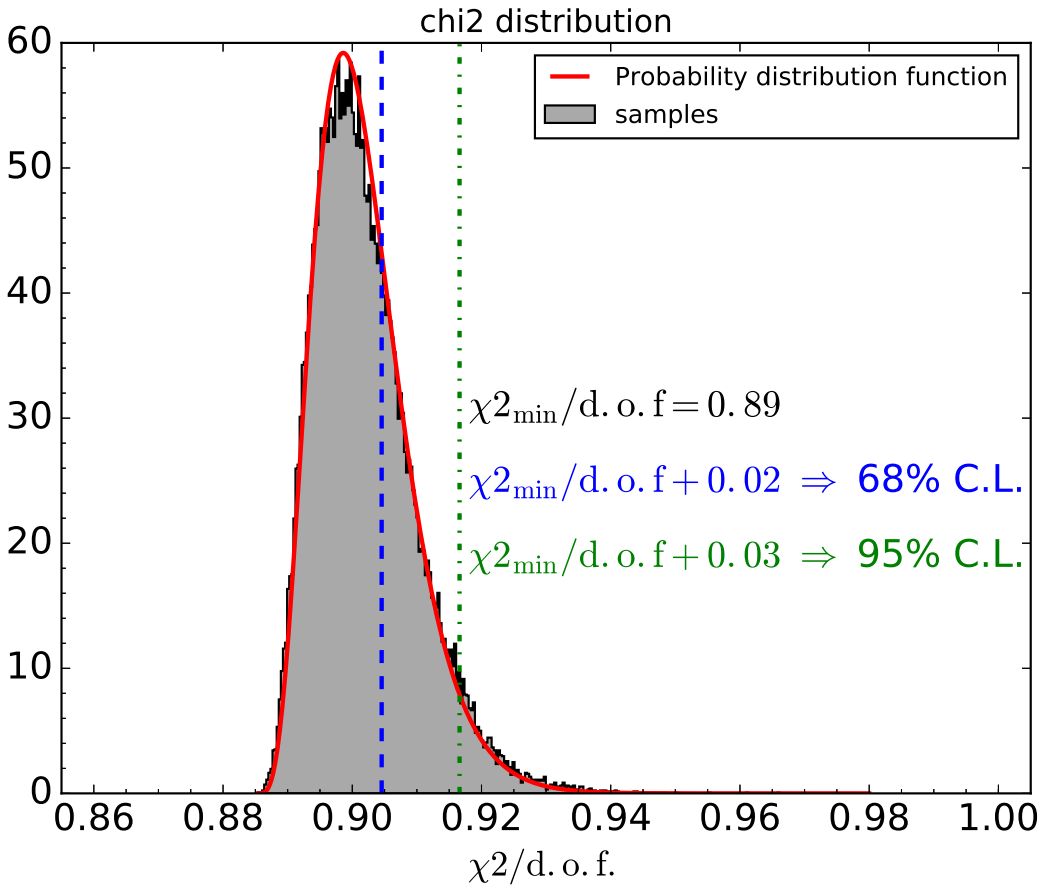
Parameter marginal distributions not necessarily gaussian





PDF Parameters: χ^2 distribution

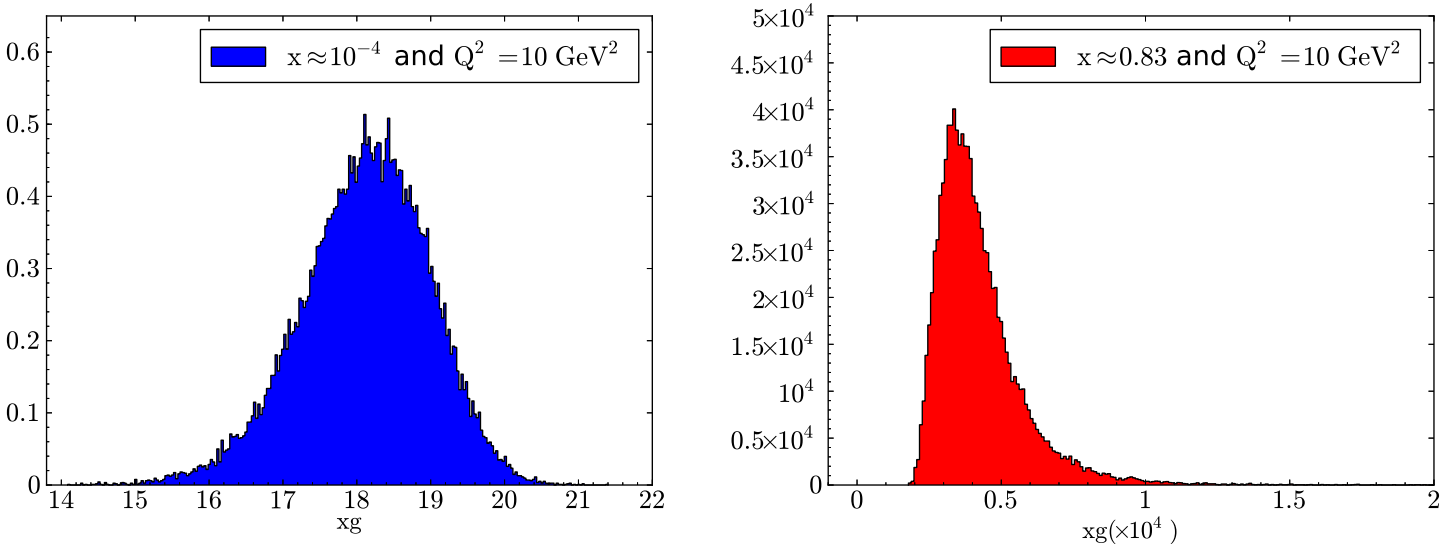
χ^2 distribution for a 10D MCMC. The solid line is an adjustment with a χ^2 distribution law.



Probability distribution functions of PDFs (PDFs of PDFs...)

For each MC parameter set $\vec{q}_t$, compute the corresponding PDFs

probability distribution of PDFs at fixed (x, Q^2)



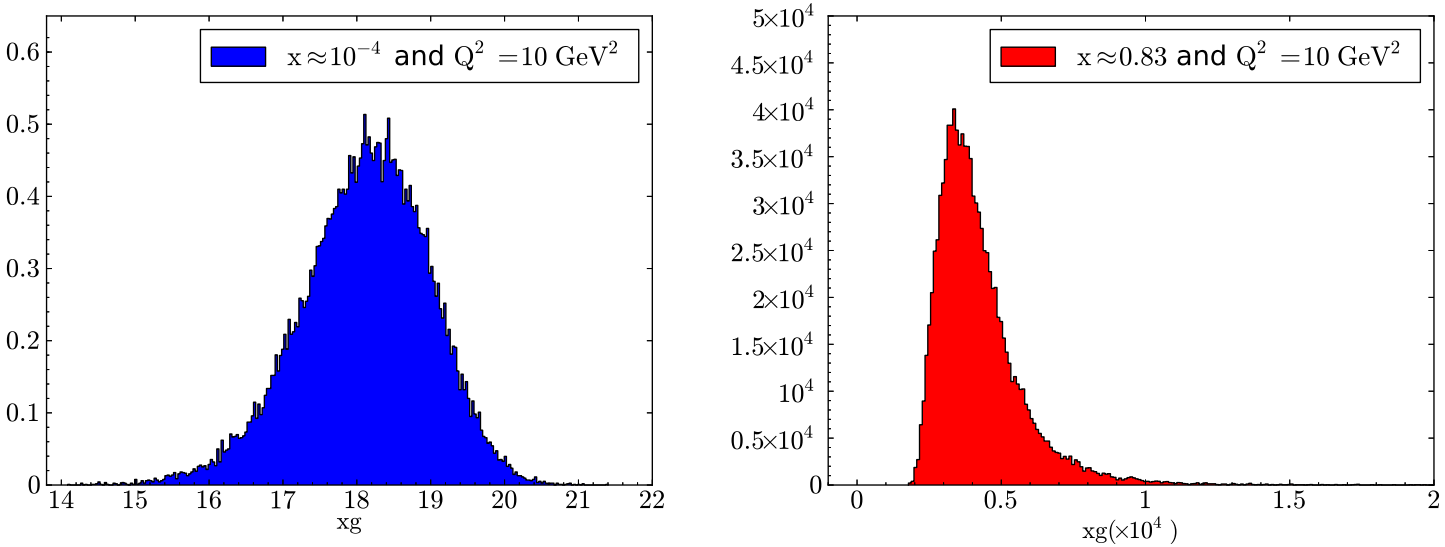
Gluon PDF probability distribution function for $x \approx 10^{-4}$ (l.h.s.) and $x \approx 0.83$ at fixed $Q^2 = 10$ GeV^2.

The 68% confidence interval is obtained considering the region of the distribution containing 68% of the data remaining on each side of the most probable value.

Probability distribution functions of PDFs (PDFs of PDFs...)

For each MC parameter set $\vec{q}_t$, compute the corresponding PDFs

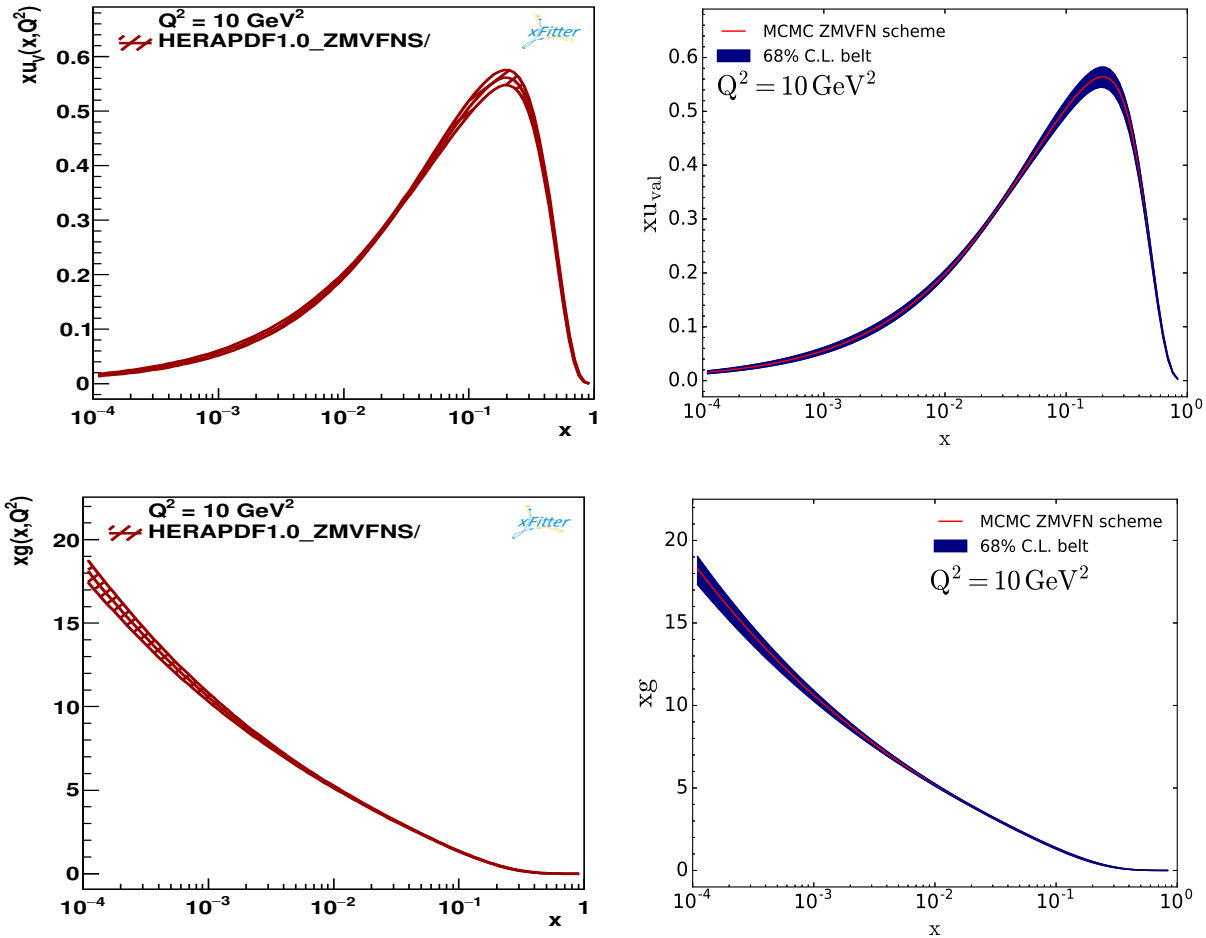
probability distribution of PDFs at fixed (x, Q^2)



Gluon PDF probability distribution function for $x \approx 10^{-4}$ (l.h.s.) and $x \approx 0.83$ at fixed $Q^2 = 10$ GeV^2.

The 68% confidence interval is obtained considering the region of the distribution containing 68% of the data remaining on each side of the most probable value.

MCMC PDFs



The parton distribution functions obtained using MCMC (right) compared to HERAPDF1.0 (ZMVFN scheme) from xFitter output (left) for $\mathrm{xu_{val}}$ *and* xg*, at* $\mathrm{Q}^2 = 10\ \mathrm{GeV}^2$*. The bands show the* 68% *confidence interval around the most probable value for the MCMC PDFs, and the standard* $\Delta\chi^2 = 1$ *deviation for HERAPDF.*

Outline

1. Introduction
2. From minimization to Bayesian inference
   - Formulation of PDFs determination in terms of Bayesian inference
   - Markov chains
   - Metropolis algorithm
   - Hybrid (or Hamiltonian) Monte Carlo
3. Markov chain analysis
   - Thermalisation, autocorrelation and all that....
4. First results
   - PDFs Parameters
   - PDFs probability distribution functions
5. Conclusions

Conclusions

What we have done....

- MCMC, well-suited to multi-parameter determination, applicable to PDFs extraction
- Overcome the technical difficulties by a nice recycling of a lattice algorithm (HMC)
- Obtained probability densities of PDFs and as by-product, confidence intervals

What we would like to do....

- Consider more complex χ^2 functions including correlations and study prior influence
- Extend this work to a competitive PDF ensemble with more parameters and data
- Extract α_s and possibly m_c, m_b from data using MCMC

What we have NOT done....

- Tackled the problem of potentially incompatible data, but....

Bayesian approach applied to global analyses can lead to a deeper insight into PDFs uncertainties determination

Conclusions

What we have done....

- MCMC, well-suited to multi-parameter determination, applicable to PDFs extraction
- Overcome the technical difficulties by a nice recycling of a lattice algorithm (HMC)
- Obtained probability densities of PDFs and as by-product, confidence intervals

What we would like to do....

- Consider more complex χ^2 functions including correlations and study prior influence
- Extend this work to a competitive PDF ensemble with more parameters and data
- Extract α_s and possibly m_c, m_b from data using MCMC

What we have NOT done....

- Tackled the problem of potentially incompatible data, but....

Bayesian approach applied to global analyses can lead to a deeper insight into PDFs uncertainties determination

Conclusions

What we have done....

- MCMC, well-suited to multi-parameter determination, applicable to PDFs extraction
- Overcome the technical difficulties by a nice recycling of a lattice algorithm (HMC)
- Obtained probability densities of PDFs and as by-product, confidence intervals

What we would like to do....

- Consider more complex χ^2 functions including correlations and study prior influence
- Extend this work to a competitive PDF ensemble with more parameters and data
- Extract α_s and possibly m_c, m_b from data using MCMC

What we have NOT done....

- Tackled the problem of potentially incompatible data, but....

Bayesian approach applied to global analyses can lead to a deeper insight into PDFs uncertainties determination


THANK YOU