PDF's: Present and Future

Walter Giele

Publications: hep-ph/9803393;hep-ph/0104052;hep-ph/0104053
LHAPDF website: pdf.fnal.gov

Outline of Talk:

1. Theory of PDF fitting
2. "Core" Phenomenology
3. Conclusion and Future

ICHEP 2002
Amsterdam, July 25 2002

Theory of PDF fitting

√ Uncertainty fits (e.g. Alekhin, Botje, Fermi2001, ...):

↪ My goal is to predict the PDF uncertainties induced in TEVATRON/LHC observables. The PDF's should be fitted to a minimal set of experiments in order to maximize the predictive power of the PDF set.

↪ The fact that we have to worry about PDF uncertainties is a testimony to the dramatic increase in accuracy of hadron collider experiments.

↪ A rigorous statistical treatment is required for any sort of progress in this area. Purging of all wiggle room is required with a clear definition of the core assumptions.

√ Methods for General PDF fitting:

$$P^{\mathcal{O}}_{pdf}(x_e) = \int_{V(\mathcal{F})} \mathcal{D}\,\mathcal{F}\; P_{prior}(\mathcal{F}) \times P^{input}_{exp}(\mathcal{F}) \times P^{\mathcal{O}}_{MC}\left(x_e|x_t(\mathcal{F})\right)$$

↪ $P_{prior}(\mathcal{F})$: Defines initial assumptions (it can be seen as the metric definition/regulator function/smoothness definition).

↪ $P^{\mathcal{O}}_{MC}\left(x_e|x_t(\mathcal{F})\right)$: The conditional probability density for measuring a value x_e for observable $\mathcal{O}$ given the theory prediction $x_t(\mathcal{F})$.

↪ $P^{input}_{exp}(\mathcal{F})$: Input experiments used in the PDF determination.

$\surd$ Numerical solution through sequential MC implementation:

$$\mathcal{D}\,\mathcal{F}' \;= \mathcal{D}\,\mathcal{F}\; P_{prior}(\mathcal{F}) \times P_{exp}^{input}(\mathcal{F})$$

such that

$$\begin{aligned} P_{pdf}^{\mathcal{O}}(x_e) &= \int_{V(\mathcal{F}')} \mathcal{D}\,\mathcal{F}'\; P_{MC}^{\mathcal{O}}\left(x_e | x_t(\mathcal{F}')\right) \\ &\simeq \frac{1}{N_{PDF}} \sum_{i=1}^{N_{PDF}} P_{MC}^{\mathcal{O}}\left(x_e | x_t(\mathcal{F}_i)\right) \end{aligned}$$

$\hookrightarrow$ The PDF list $\{\mathcal{F}_i\}$ is generated using a sequential MC approach. This easily deals with not only non-gaussian behavior but also with multiple maxima and zero modes.

$\hookrightarrow$ A gaussian approximation is not acceptable as it forces one to reduce the number of PDF parameters such that no zero modes exist (ie the second derivatives at the χ^2_{min} exist). This leads to an unavoidable underestimate of the PDF uncertainties.

$\surd$ In the "real world" experiments should stand behind and defend their published results. Discrepancies between an experiment and theory can have many sources (in order of likelyhood):

$\hookrightarrow$ Incorrect modeling of non-perturbative physics (eg nuclear modeling, higher twist).

$\hookrightarrow$ PDF prior (eg restricted parametrization).

$\hookrightarrow$ Mistake in experimental analysis (this is rare).

√ This leads (at the moment) to a restrictive set of experiments which one should consider in PDF uncertainty analysis:

↪ Proton DIS data (with appropriate (Q^2, W^2) cuts): BCDMS, E665, H1 and ZEUS.

↪ Liquid hydrogen fixed target data: E772 DY results.

↪ A collection of $P\bar{P}$ experimental results: UA1, UA2, CDF, D0.

√ Using PDF's based on these restrictive sets of experiments we can start to study the modeling of nuclear effects. Once we understand how to model these effects we can include many more experiments (without error inflation).

√ If we want to make predictions for an experiment using a PDF set we have to be careful whether or not data from the same experiment was used in the PDF determination. If so, the PDF uncertainties become correlated with the experimental uncertainties and need to be calculated for a correct result. (e.g. if the PDF determination is in part based on the D0 one-jet inclusive jet cross section, the top quark measurement uncertainties will be correlated with the PDF uncertainties). *In general we should avoid such situations.*

√ All these considerations lead to Fermi2001: determined using H1+BCDMS+E665

	H1	BCDMS	E665
H1-fit	-	67%	21%
BCDMS-fit	85%	-	23%
E665-fit	30%	82%	-

"Core" Phenomenology

- √ Three fundamental measurements at a $P\bar{P}$ collider:
  - ↪ Counting jets: One jet inclusive E_T distribution.
  - ↪ Counting leptons: W,$Z/\gamma^\star$ production.
  - ↪ ~~Counting photons: prompt photon E_T distribution~~.
- √ These cross sections are/will be up to NNLO. For now comparisons at NLO, once the forthcoming NNLO evolution is available we can redo the analysis one order higher.
- √ The analysis will by shown using the following 4 PDF sets (all in random sampling representation):
  - ↪ Alekhin-fit (hep-ph/0011002): E-Gaussian, P/D fit DIS, variable α_S, fixed $\mu^2_{R/F}$, 5 param. higher twist model, no nuclear modeling.
  - ↪ Botje-fit (hep-ph/9912439): Diagonalized E-Gaussian, P/D DIS fit, fixed α_S, fixed $\mu^2_{R/F}$, 5 param. higher twist model, no nuclear modeling.
  - ↪ RCTEQ6-fit (hep-ph/0201195): Inflated E-Gaussian, "global" fit, fixed α_S, fixed $\mu^2_{R/F}$, appropriate (Q^2, W^2) cuts for DIS, nuclear model ?
  - ↪ Fermi2001-fit (hep-ph/0104052): T-sequential MC, P DIS fit, variable α_S, variable $\mu^2_{R/F}$ appropriate (Q^2, W^2) cuts.
- √ All these sets are downloadable from `pdf.fnal.gov` integrated in a single LHAPDF interface. (Note that the CTEQ6 PDF set on the site is not the random sampling version.)

√ One-jet inclusive transverse energy distribution:

↪ Observable sensitive to color charges of parton: $p \to (g, \sum q)$.

↪ To calculate the agreement between the PDF set using NLO and the CDF 1a data we calculate the confidence level:

$$CL(x_m) = \int_{V(meas.)} \mathcal{D}x \; P_{pdf}(x) \times \Theta\left(P_{pdf}(x_m) - P_{pdf}(x)\right)$$

$$\simeq \frac{1}{N_{CL}} \sum_{i=1}^{N_{CL}} \Theta\left(P_{pdf}(x_m) - P_{pdf}(x_i)\right)$$

$$P_{pdf}(x_e) = \int_{V(\mathcal{F})} \mathcal{D}\mathcal{F} \; P_{prior}(\mathcal{F}) \times P_{exp}^{input}(\mathcal{F}) \times P_{MC}\left(x_e | x_t(\mathcal{F})\right)$$

$$\simeq \frac{1}{N_{pdf}} \sum_{i=1}^{N_{pdf}} P_{MC}\left(x_e | x_t(\mathcal{F}_i)\right)$$

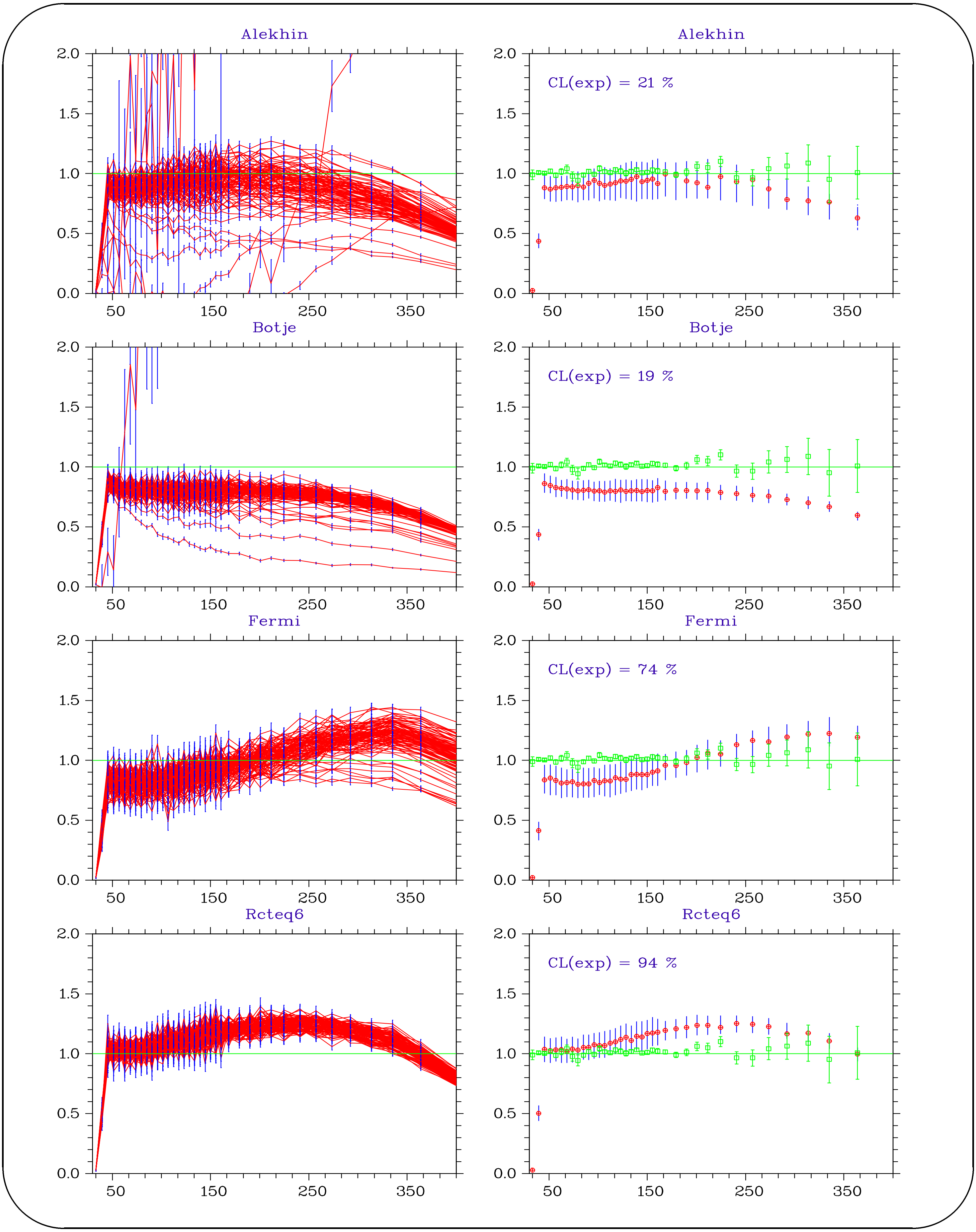
$$P_{MC}(x|x_t(\mathcal{F})) = \int_{V(\mathcal{S})} d\mathcal{S} \; P_{sys}(\mathcal{S}) \times P_{MC}(x|\mathcal{S}(x_t^{nlo}(\mathcal{F})))$$

$$\simeq \frac{1}{N_{pseudo}} \sum_{i=1}^{N_{pseudo}} P_{MC}(x|\mathcal{S}_i(x_t^{nlo}(\mathcal{F})))$$

↪ The Fermi2001 PDF set has an excellent agreement with the CDF 1a measurement.

	Alekhin	Botje	Rcteq6*	Fermi2001
CL(CDF 1a)	21%	19%	94%	74%

* *Observable included in Cteq6 determination.*
Not sure how to interpret Cteq6 confidence level number.



Alekhin
CL(exp) = 21 %
Botje
CL(exp) = 19 %
Fermi
CL(exp) = 74 %
Rcteq6
CL(exp) = 94 %

$\surd$ Vector boson cross sections:

$\hookrightarrow$ Observable adds sensitive to EW charges of partons:

$p \to (g, \sum q) \to (g, \sum U, \sum D)$.

$\hookrightarrow$ To calculate the agreement between the PDF set using NLO and the D0 1b data we calculate the confidence level:

$$CL(N_W^m, N_Z^m) = \int_{V(meas.)} dN_W \; dN_Z \; P_{pdf}(N_W, N_Z) \times \Theta\left(P_{pdf}(N_W^m, N_Z^m) - P_{pdf}(N_W, N_Z)\right)$$

$$\simeq \frac{1}{N_{CL}} \sum_{i=1}^{N_{CL}} \Theta\left(P_{pdf}(N_W^m, N_Z^m) - P_{pdf}(N_W^{(i)}, N_Z^{(i)})\right)$$

$$P_{pdf}(N_W, N_Z) = \int d\mathcal{L} P(\mathcal{L}) \times P_{pdf}(N_W, N_Z|\mathcal{L})$$

$$\simeq \frac{1}{N_{\mathcal{L}}} \sum_{i=1}^{N_{\mathcal{L}}} P_{exp}\left(N_W, N_Z|\mathcal{L}_{\rangle}\right)$$

$$P_{pdf}(N_W, N_Z|\mathcal{L}) = \int_{V(\mathcal{F})} \mathcal{D}\mathcal{F} \; P_{prior}(\mathcal{F}) \times P_{exp}^{input}(\mathcal{F}) \times P_{exp}(N_W, N_Z|\sigma_W(\mathcal{F}), \sigma_Z(\mathcal{F}), \mathcal{L})$$

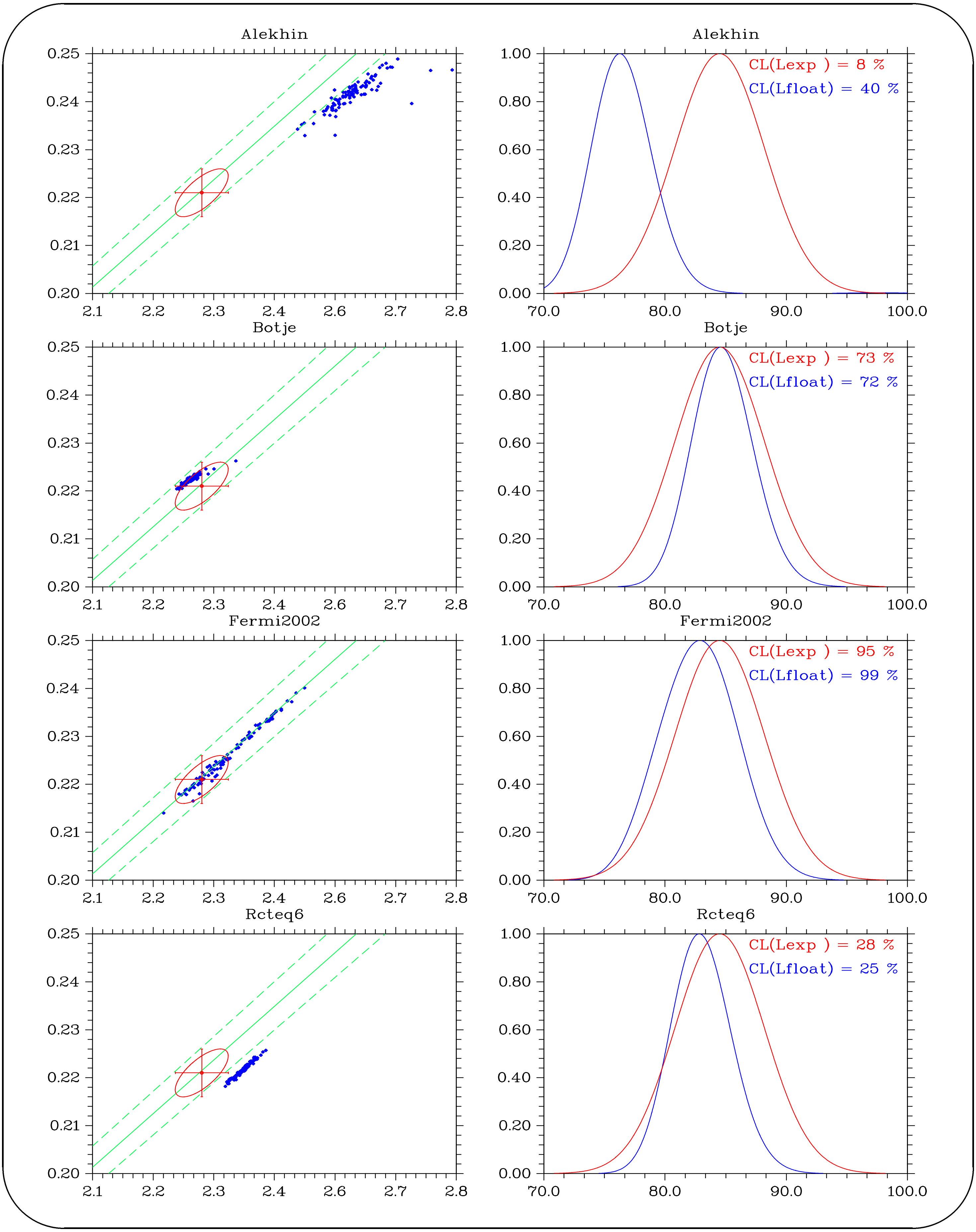
$$\simeq \frac{1}{N_{pdf}} \sum_{i=1}^{N_{pdf}} P_{exp}(N_W, N_Z|\sigma_W(\mathcal{F}_i), \sigma_Z(\mathcal{F}_i), \mathcal{L})$$

$$P_{exp}(N_W, N_Z|\sigma_W, \sigma_Z, \mathcal{L}) = \frac{1}{2\pi\sqrt{|C|}} e^{-\frac{1}{2}(\mathcal{L}\sigma - N)_i C_{ij} (\mathcal{L}\sigma - N)_j}$$

$\hookrightarrow$ The Fermi2001 PDF set has an excellent agreement with the D0 1b measurement.

	Alekhin	Botje	Rcteq6*	Fermi2001
CL(D0 1b) ($P(\mathcal{L}_{exp})$)	8%	73%	28%	95%
CL(D0 1b) ($P(\mathcal{L}_{float})$)	40%	72%	25%	99%
$\mathcal{L}_{float}$, 75% CL (pb^{-1})	76.3 ± 2.5	84.6 ± 2.6	82.8 ± 2.5	82.9 ± 3.3

* *Not sure how to interpret Cteq6 confidence level number.*



Alekhin
Alekhin
CL(Lexp) = 8 %
CL(Lfloat) = 40 %
Botje
Botje
CL(Lexp) = 73 %
CL(Lfloat) = 72 %
Fermi2002
Fermi2002
CL(Lexp) = 95 %
CL(Lfloat) = 99 %
Rcteq6
Rcteq6
CL(Lexp) = 28 %
CL(Lfloat) = 25 %

Conclusions: Results

- √ The results show that the Fermi2001 set is more than an "experimental" set to demonstrate a method. It can be used for a variety of phenomelogical studies allowing unambiguous statistical predictions involving PDF uncertainties.
- √ Using the Fermi2001 set we can make the following clear (bayesian) statistical statement:
  Given the MRST-parametrization and the F_2^P data from H1, BCDMS and E665, we predict the measured CDF 1a and D0 1b W/Z cross sections with a confidence level of 74% and 95% respectively.
- √ Given we have a high confidence that PQCD is applicable for all observables in the analysis the statistical statements does not come as a surprise. A low confidence level would be hard to explain indicating a breakdown in the PQCD framework.
- √ However, one flaw remains which allows wiggle room):
  The prior P_{prior} (ie the parametrization choice).
  The new Fermi2002 set will address this problem by:
  - ↪ No explicit parametrization, but a "smoothness" measure: $P_{prior}(\mathcal{F}) \propto e^{-S(\mathcal{F})}$ (This removes parametrization dependences.)
  - ↪ To describe $\mathcal{F}$ we use a complete set of functions. The smoothness measure constrains all parameters (even in the absence of data).
  - ↪ Release of pdffitter(.f,.cc,.py) for general use.