Monte Carlo approach to PDF's with errors

Outline of Talk:

1. Theory of PDF fitting
2. Examples
3. Conclusion and Future

Walter Giele, CERN 2003

Theory of PDF fitting

√ Goal:

↪ 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.

↪ Develop a rigorous statistical treatment. 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 problematic 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.

$\surd$ Care has to be taken in selecting experiments for a consistent PDF determination. 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.

√ All these considerations lead to Fermi2001 determined using H1+BCDMS+E665 (all F_2^P measurements in the perturbative kinematic region)

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

Examples

√ Two fundamental measurements at a hadron collider:

↪ Counting jets: One jet inclusive E_T distribution.

↪ Counting leptons: $W, Z/\gamma^\star$ production.

√ 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.

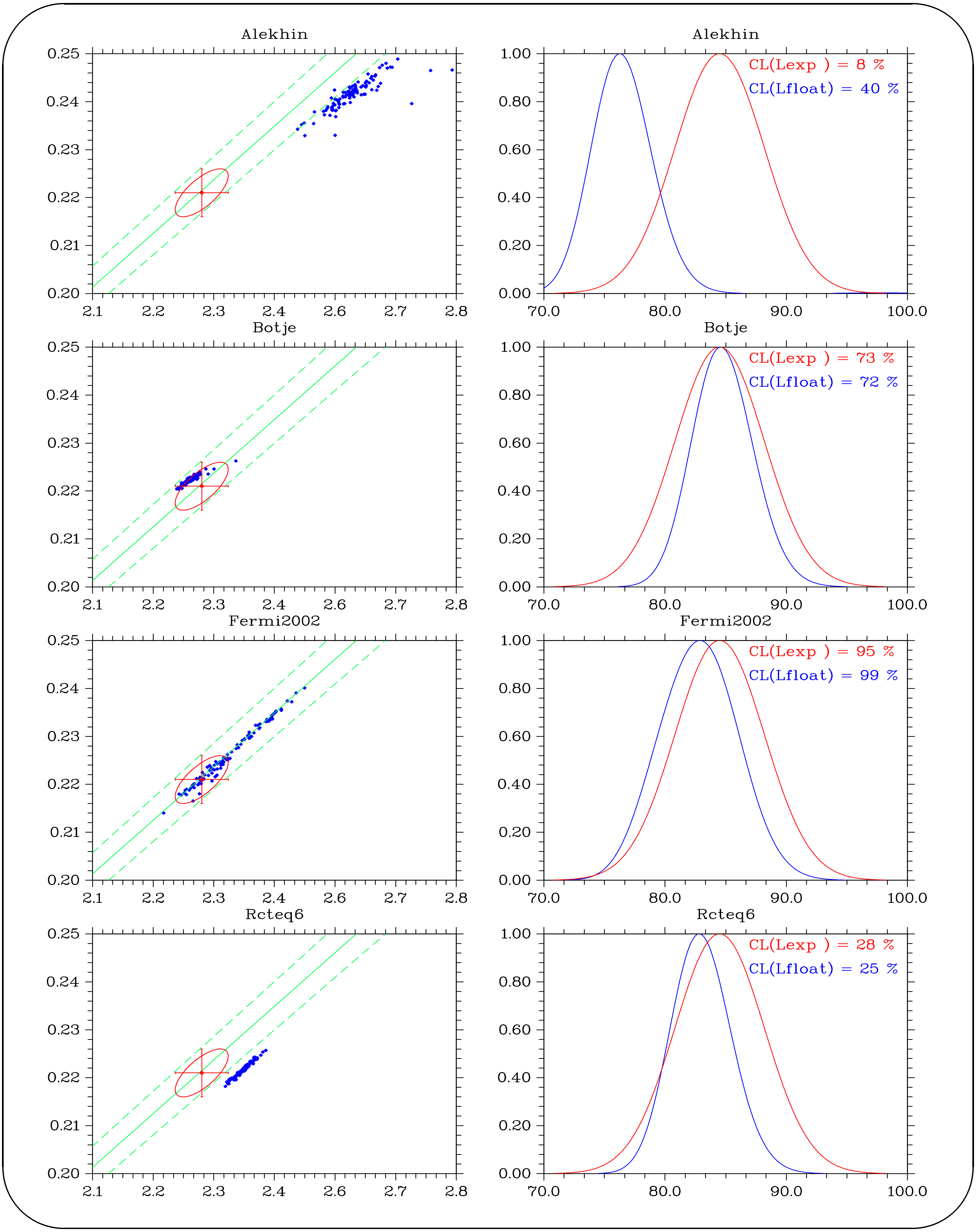
↪ The Fermi2001 PDF set has an excellent agreement with the CDF run 1 measurement one jet inclusive measurement:

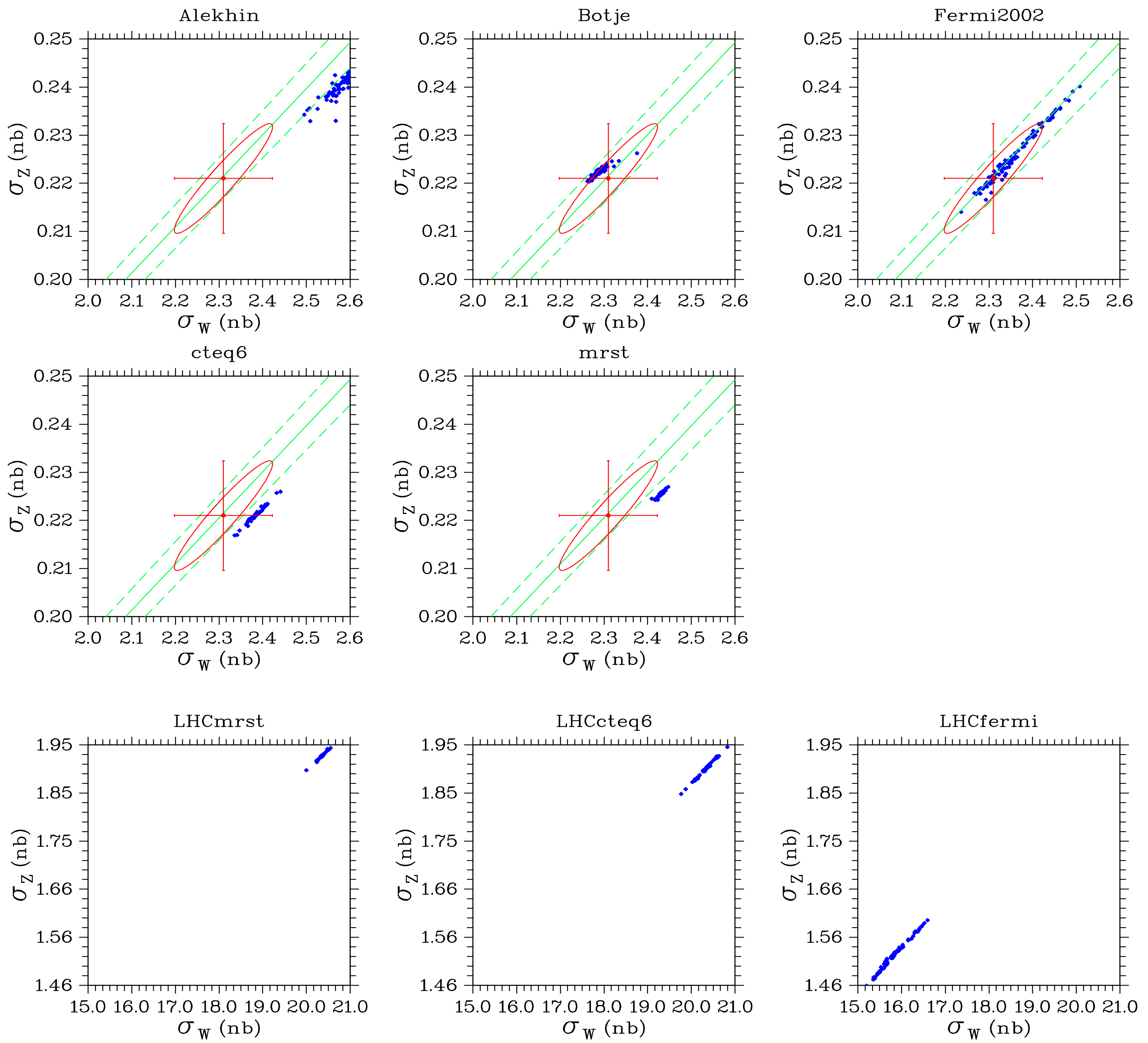
	Alekhin	Botje	Rcteq6*	Fermi2001
CL	21%	19%	94%	74%

* *Observable included in Cteq6 determination.*

↪ The Fermi2001 PDF set has an excellent agreement with the D0 run 1 measurement of W and Z cross sections:

	Alekhin	Botje	Rcteq6	Fermi2001
CL $(P(\mathcal{L}_{exp}))$	8%	73%	28%	95%
CL $(P(\mathcal{L}_{float}))$	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


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 %



Alekhin
Botje
Fermi2002
cteq6
mrst
LHCmrst
LHCcteq6
LHCfermi
σ Z (nb)
σ W (nb)
0.20
0.21
0.22
0.23
0.24
0.25
2.0
2.1
2.2
2.3
2.4
2.5
2.6
1.46
1.56
1.66
1.75
1.85
1.95
15.0
16.0
17.0
18.0
19.0
20.0
21.0

Conclusions: Results

$\surd$ The results show that the Fermi2001 set is more than an "trial" set to demonstrate a method. It can be used for a variety of phenomelogical studies allowing unambiguous statistical predictions involving PDF uncertainties.

$\surd$ 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 run 1 inclusive jet cross section and the combined D0 W/Z cross sections with a confidence level of 74% and 95% respectively.

$\surd$ 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.

$\surd$ However, one flaw remains which allows wiggle room):
The prior P_{prior} (ie the parametrization choice).
The new set has to address this problem by:

- $\hookrightarrow$ A "smoothness" measure which does not refer to a specific parametrization: $P_{prior}(\mathcal{F}) \propto e^{-S(\mathcal{F})}$ (This removes parametrization dependences.)
- $\hookrightarrow$ To describe $\mathcal{F}$ we use a complete set of functions. The smoothness measure constrains all parameters (even in the absence of data).
- $\hookrightarrow$ Release of pdffitter(.f,.cc,.py) for general use.