Phenomenological Implications of Parton Density Function Uncertainties

Walter Giele

Collaborators: Stéphane Keller (and David Kosower)

Website: `pdf.fnal.gov`

Outline of Talk:

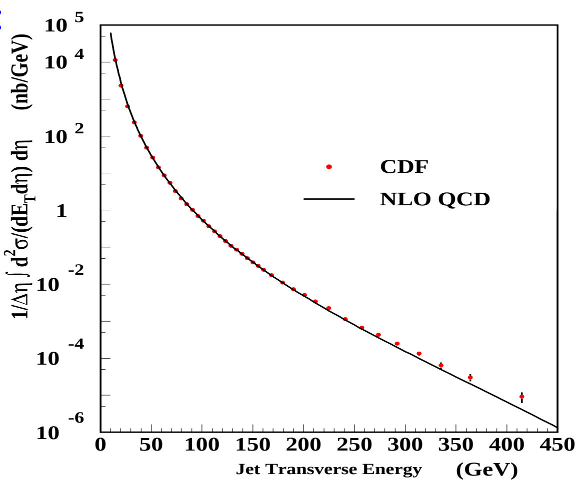
1. Introduction: Motivation *and* Goals
2. The Method: Framework *and* Implementation
3. The Fits: Priors *and* Experiments
4. Tevatron Predictions:
   - (W, Z) production *and* luminosity measurements
   - one-jet inclusive *and* α_S measurement
   - prompt photon production (*and* k_T-smearing ?)
5. Outlook: Conclusion *and* Future

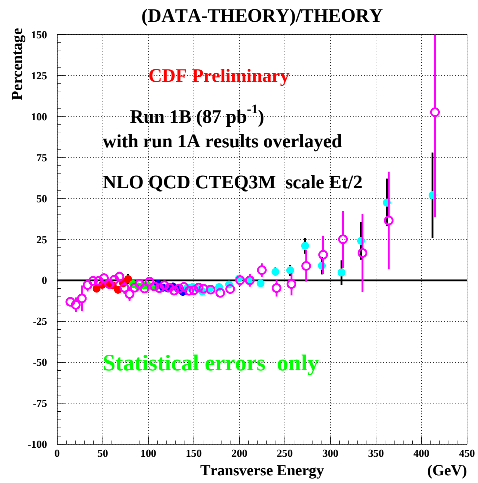
RADCOR-2000 Workshop

Carmel, September 12 2000

Introduction: Motivation *and* Goals

The watershed moment was the run 1a one-jet inclusive result from CDF:





CDF in Phys. Rev. Lett. 77,438, 1996):

> *$\cdots$ Above 160 GeV, however, each of the four [statistical test] methods yield a probability of 1% that the excess is due to a fluctuation*

Structure Function Subgroup in Snowmass 1996 Workshop: New Directions for High Energy Physics):

> *The much discussed high E_T inclusive jet cross-section has been shown to be compatible with all existing data within the framework of conventional pQCD $\cdots$*

Clearly from the conventional *qualitative* approach not much can be learned. A *quantitative* approach is needed with well defined probabilistic predictions given a set of priors to proceed.

In the *quantitative* approach one uses the procedure:

$$\left.\begin{array}{l}\text{DIS \& DY}\\ \text{Some ``choices''}\\ \text{Simplified error analysis}\end{array}\right\} \Longrightarrow \boxed{\chi^2_{\text{min}}\ \text{PDF}} \Longrightarrow \left\{\begin{array}{l}\text{Predictions}\\ \text{Conclusions}\end{array}\right.$$

However the *quantitative*, probabilistic method follows a different approach :

- Well defined probabilistic interpretations of results
- Traceable and changeable priors
- No compromise treatment of *published* experimental uncertainties (including non-gaussian uncertainties)

While the priors are subject to interpretation the procedure to get from priors to results has to be unique. That is, given the priors there can be only one correct answer with no room of intermediate choices or interpretations.

$$\left.\begin{array}{l}\text{Input Experiments}\\ \text{Physics Model}\\ \text{(Extra Priors)}\end{array}\right\} \Longrightarrow \boxed{\begin{array}{c}\forall\ \text{PDF}\\ Prob(\text{PDF})\end{array}} \Longrightarrow \left\{\begin{array}{l}\text{Predictions}\\ \text{Conclusions}\end{array}\right.$$

Many people reached the conclusion we have reached the end of the road for the *qualitative* phenomenological approach of testing QCD. Instead we are evolving to the *quantitative* probabilistic approach of using QCD as a tool to probe beyond.

- S. Alekhin, hep-ph/9611213, Eur. Phys. J. C10 (1999) 395-403.
- W. Giele and S. Keller, Phys. Rev. D58 (1998) 094023.
- M. Botje, hep-ph/9912439.
- V. Barone, C. Pascaud and F. Zomer, Eur.Phys.J. C12 (2000) 243-262.
- J. Bloch and R. Ball ?
- CTEQ ?
- MRS ?

It is clear from all the progress and results in the probabilistic approaches to PDF's in Run II CDF and D0 will (hopefully) say:

> *Given the priors our analysis yield a probability of 1% that the excess is due to a fluctuation.*

As it should, the *only* possible point of contest will now be the priors. Varying/changing/stretching the priors is needed to convince oneself such an excess is really due to new physics. Remember the prior consists of the physics model and input experiments, exactly the points one needs to discuss.

The Method: Framework *and* Implementation

This section assumes an *ideal* world. Of course in practice compromises have to be made. But the mathematical underpinning and numerical implementation have to be defined in a strict manner.
First we have to defining the physical model to make approximations to the true nature value $\{x_t\}$.

The theory prior:

- Order of PQCD (LO, NLO, NNLO,. . .).
- PDF parameterization/smoothness/constraints.
- Resummation scheme (if any).
- Higher twist model (if any).
- Heavy Nuclei model or shadow model (if any).
- Fragmentation models (if any).

Notes:

- All choices have to be implemented consistently for all predictions.
- Adding non-perturbative phenomenological models will involve additional parameters which have to be determined and which will build up correlations with the PDF parameters.

Next we must define the interface with the measurement.

The experimental prior:

The definition will be based on the axiom

> *Any experimental measurement must have a quantifiable uncertainty*

This means there must exist a experimental response function $P_{exp}(\{x_t\}|\{x_e\})$ which measures the probability density of measuring $\{x_e\}$ given the true nature value $\{x_t\}$.

Notes:

- The response function P_{exp} contains all possible information we can extract from the measurement of $\{x_e\}$.
- The response function is defined without dependence on any theoretical model.

Other priors:

These can be many based on a variety of sources and mathematically contained in $P_{prior}(pdf)$. Examples:

- We might know one (or more) moments of certain pdf's from lattice QCD
- We might want to include a previous fit
- We might want to impose smoothness constraints or small/large x constraints

Now we are ready to make predictions for observable $\mathcal{O}$ including PDF uncertainties:

$$\langle \mathcal{O} \rangle = \int \mathcal{D}f \; P_{fit}(f) \times \mathcal{O}(f)$$

where

$$P_{fit}(f) = P_{exp\,1}(f) \times P_{exp\,2}(f) \times \cdots \times P_{exp\,n}(f) \times \; P_{prior}(f)$$

i.e. a functional integration over the PDF set f.
In a "perfect" world this is all what is needed. The only things needed is the definition of the prior (i.e. defining the physics model and construction of the experimental response functions).
In the "real" world we perform the numerical integration over the PDF's using a Monte Carlo approach:

$$\langle \mathcal{O} \rangle = \frac{1}{N} \sum_{i}^{N} P_{fit}(f_i) \times \mathcal{O}(f_i)$$

To optimize the Monte Carlo integration we "unweight" the PDF's according to P_{fit}. That is, we generate a set of N PDF's distributed according to P_{fit}. This means we simply get

$$\langle \mathcal{O} \rangle = \frac{1}{N} \sum_{i}^{N} \mathcal{O}(f_i^{unweighed})$$

To summarize:

A "fit" is *merely* an unweighting of to PDF integration with respect to a set of experiments to numerically make the integration over the PDF's efficient (and in fact doable).

The Fits: Priors *and* Experiments

In our method a "fit" consists of:

- A choice of priors
  - Physics Model
  - Parameterization of PDF's
  - Experiment(s) to be included in the fit
- A list of 100,000 PDF's (specifically a list of PDF parameters) distributed according to P_{exp} of the included experiments.

The Physics Model:

- NLO PQCD to calculate all theory predictions $\{x_t\}$.
- Floating renormalization/factorization scales during fit.
- No resummation model.
- No fragmentation/hadronization model.
- No shadowing model: This restricts all experiments to proton targets.
- No higher twist model: To reduce non-perturbative effects in F_2^P we apply the "MRS" cuts $Q^2 > 2\ \text{GeV}^2$ and $W^2 = Q^2 \times (1/x - 1) > 10\ \text{GeV}^2$.
- Massless quarks in calculations of F_2^P.
- NNLO (x, Q^2) evolution using QCDNUM (see "A detailed Comparison of NLO QCD evolution Codes", hep-ph/9609400, J. Blümlein, M. Botje, C. Pascaud, S. Riemersma, W. L. van Neerven, A. Vogt and F. Zomer).

The Parameterization:

For our first public fits we choose the well established 21 parameter MRST parameterization at $Q_0 = 1$ GeV:

$$
\begin{aligned}
xu_v &= A_u x^{-\lambda_u}(1-x)^{\eta_u}(1+\epsilon_u\sqrt{x}+\gamma_u x)\\
xd_v &= A_d x^{-\lambda_d}(1-x)^{\eta_d}(1+\epsilon_d\sqrt{x}+\gamma_d x)\\
xg &= A_g x^{-\lambda_g}(1-x)^{\eta_g}(1+\epsilon_g\sqrt{x}+\gamma_g x)\\
xS \equiv 2x(\bar{d}+\bar{u}+\bar{s}) &= A_S x^{-\lambda_S}(1-x)^{\eta_S}(1+\epsilon_S\sqrt{x}+\gamma_S x)\\
x\Delta \equiv x(\bar{d}-\bar{u}) &= A_\Delta x^{-\lambda_\Delta}(1-x)^{\eta_S+2}(1+\gamma_\Delta x+\delta_\Delta x^2)
\end{aligned}
$$

Notes:

- From the parameterization it follows that $2\bar{u} = 0.4S - \Delta$, $2\bar{d} = 0.4S + \Delta$ and $2\bar{s} = 0.2S$.
- Heavy quarks are generated through perturbative evolution from mass threshold ($m_c^{treshold} = 1.5$ GeV and $m_b^{treshold} = 4.5$ GeV).
- In a way the choice of the MRST-parameterization incorporates into the prior the accumulated knowledge of years of PDF studies.
- The initial probability distribution of the PDF parameters is implicitly assumed uniform. This is part of the prior assumptions.

The Experiments:

At the moment we consider only DIS F_2 proton experiments and the α_S measurement at LEP.

Choice of input experiments which can be included:

Experiment	Measurement	number of fit-points	error analysis
BCDMS	F_2^P	344	gaussian
H1	F_2^P	188	half gaussian
ZEUS	F_2^P	187	gaussian
NMC	F_2^P	127	gaussian
E665	F_2^P	53	gaussian
LEP	α_S	1	gaussian

Using a fit one can calculate the compatibility (i.e. confidence level) with the other experiments:

	H1	ZEUS	BCDMS	NMC	E665	LEP
H1-fit	-	0.5%	67%	0.1%	21%	31%
ZEUS-fit	22%	-	0.1%	0.1%	5%	24%
BCDMS-fit	85%	1.5%	-	0.1%	23%	0.5%
NMC-fit	0.1%	0.1%	28%	-	1.5%	3.2%
E665-fit	30%	1.6%	82%	1%	-	99%

Notes:

- ZEUS does not want to be combined with other experiments.
- NMC does not want to be combined with other experiments.
- BCDMS has a problem with the LEP α_S value.
- NMC has a problem with the LEP α_S value.
- H1, BCDMS and E665 can be combined in fits.
- In the future we'll need to revisit the prior. From the theory side we might want to include some higher twist effects.

Tevatron Predictions: (W,Z) production

The copious production of W and Z bosons at the TEVATRON makes these two observables ideal as an alternative to the more traditional luminosity measurement using inelastic $P\bar{P}$ scattering. Because these two observables have strongly correlated experimental and theoretical uncertainties they are best studied together. All the uncertainties tend to have a similar correlation and cancel in the ratio R of the W-boson and Z-boson cross section.

Notes:

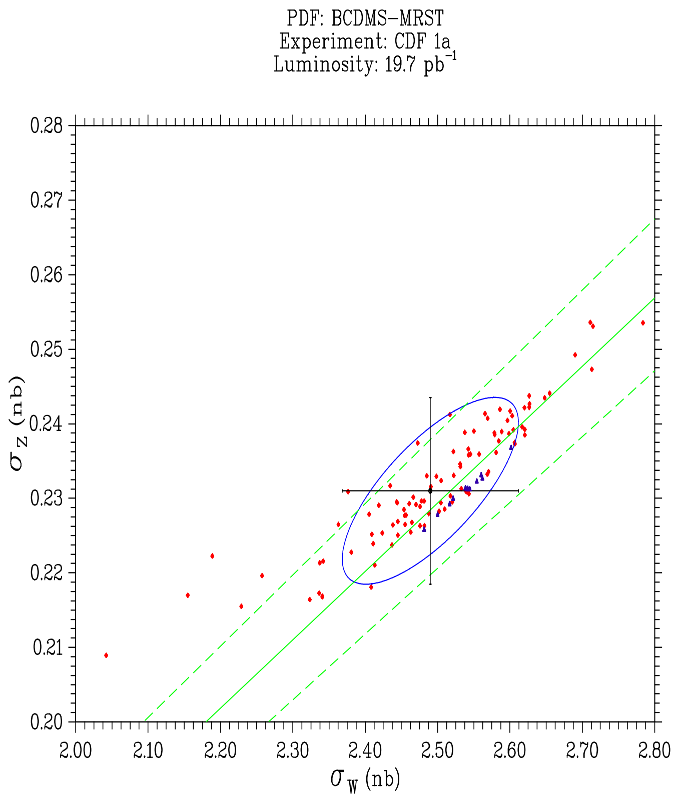
- Instead of looking at the ratio R we will look at the actual 2-dimensional measurement (σ_W,σ_Z) and specifically will attempt to extract the luminosity using the observed number of W and Z events (N_W^{exp},N_Z^{exp}) and the NLO predictions (σ_W^{Theory},σ_Z^{Theory})
- To extract the luminosity $\mathcal{L}$ we need to construct the approximate experimental response function from published CDF and D0 results. After that, we can extract the luminosity probability density function with the method described earlier

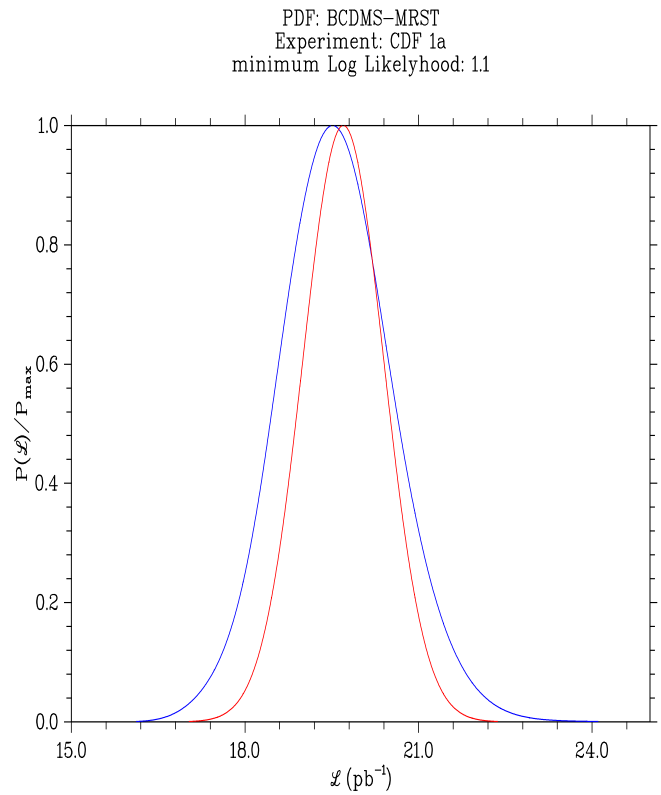
$$P_{exp}(\mathcal{L}, \sigma_W^{Theory}, \sigma_Z^{Theory} | N_W^{exp}, N_Z^{exp}) = \frac{1}{2\pi\sqrt{|C|}} e^{-\frac{1}{2} D_i C_{ij}^{-1} D_j}$$

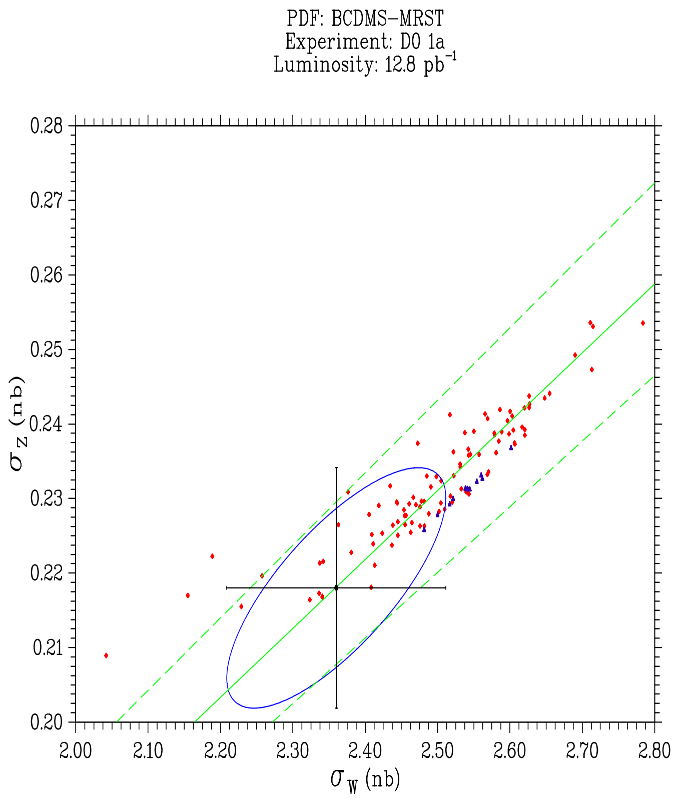
  with $D_i = (\mathcal{L}\sigma_W^{Theory} - N_W^{exp}, \mathcal{L}\sigma_Z^{Theory} - N_Z^{exp})$ and C_{ij} the correlation matrix extracted from the published results.
- We will use both the `H1+BCDMS+E665+LEP-MRST` and `BCDMS-MRST` fits.

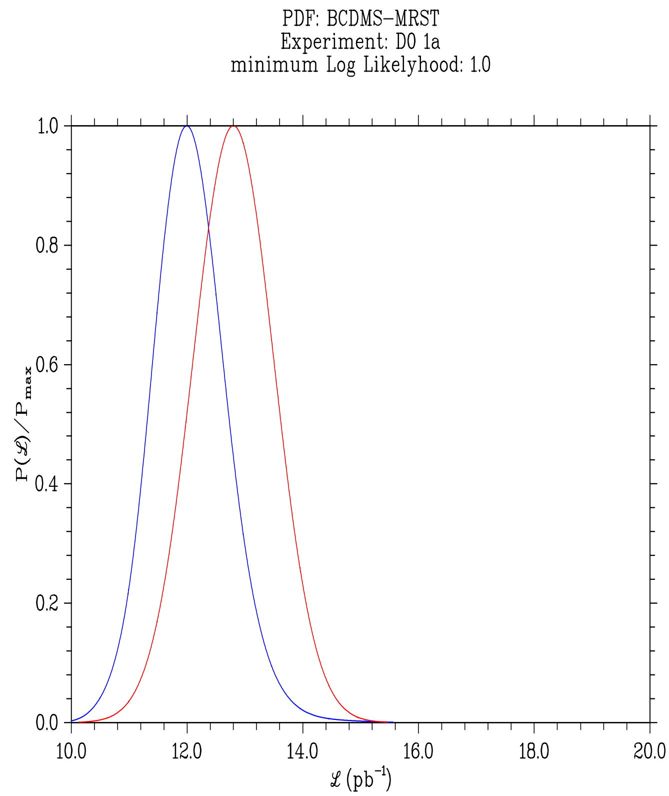
- For the scatter plots:
  - Experimental luminosity uncertainty absorbed in systematic uncertainties.
  - Blue ellipse is the one-sigma confidence level contour of measurement.
  - Black errorbars are the total W and Z-boson experimental uncertainties.
  - The green line(s) represent the ratio R measurement (with one-sigma uncertainties).
  - The red points are the 100 PDF predictions.
  - The blue points are the 12 MRS99 predictions.
- For the luminosity plots:
  - Experimental luminosity uncertainty (of course) excluded from experimental response function.
  - The red curve is the quoted luminosity from the experiments using the inelastic $\bar{P}P$ cross section.
  - The blue curve is the measured luminosity using the W and Z-boson measurements.

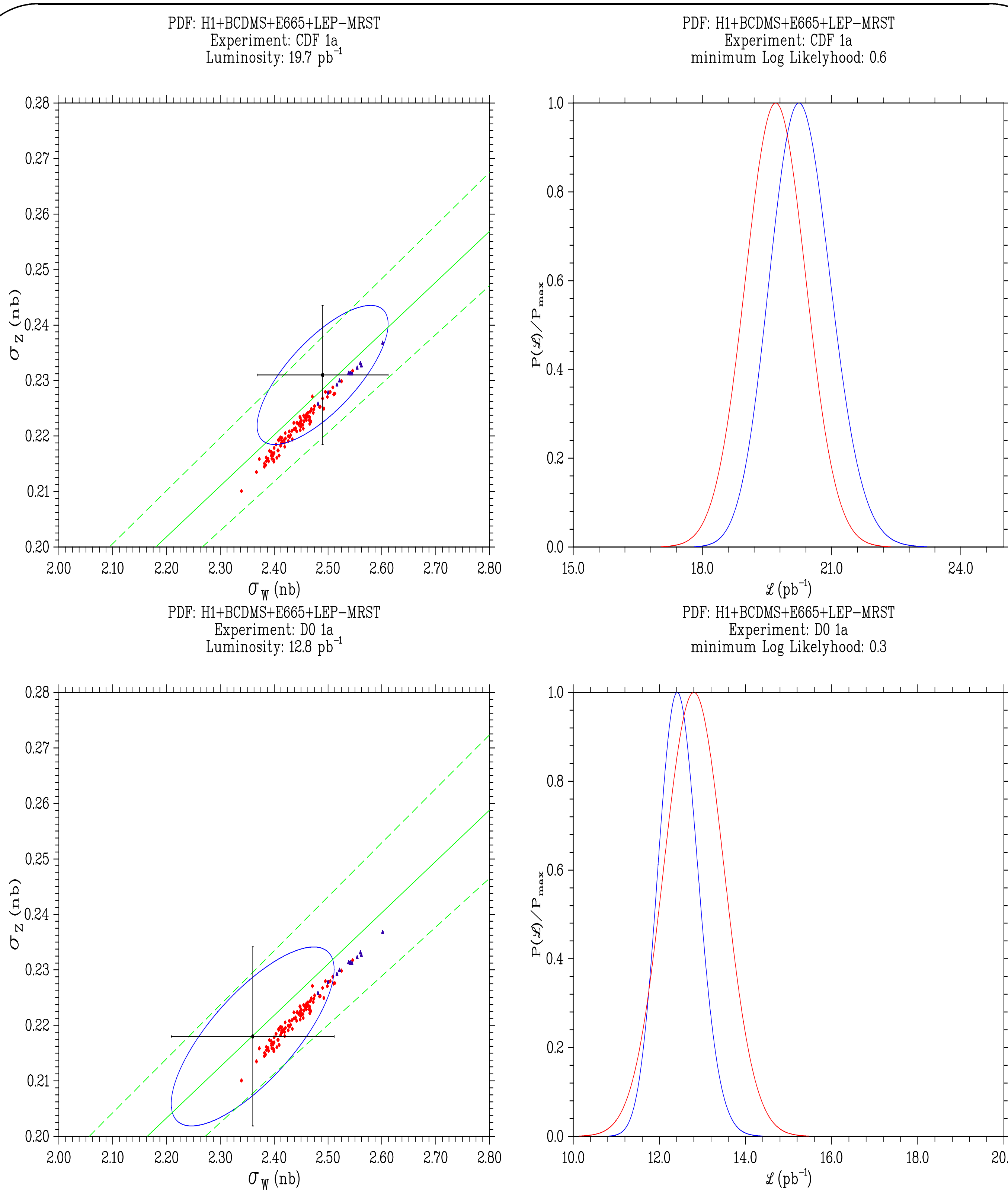
Starting off with run 1a ....





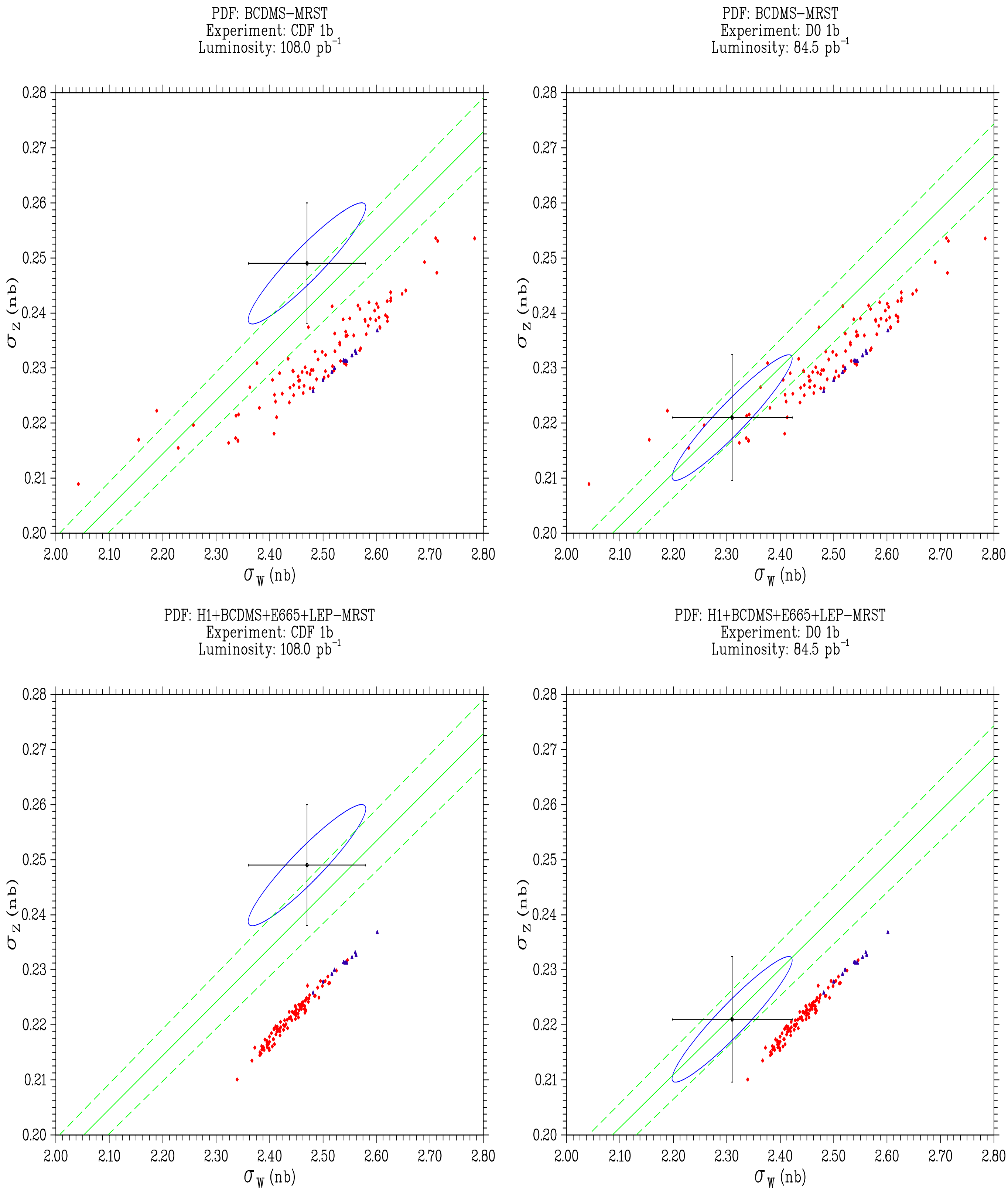






Perfect results, with luminosity uncertainty using the W and Z measurements comparable (or even better) than the inelastic $\bar{P}P$ method.

Continuing with run 1b .... Problems !



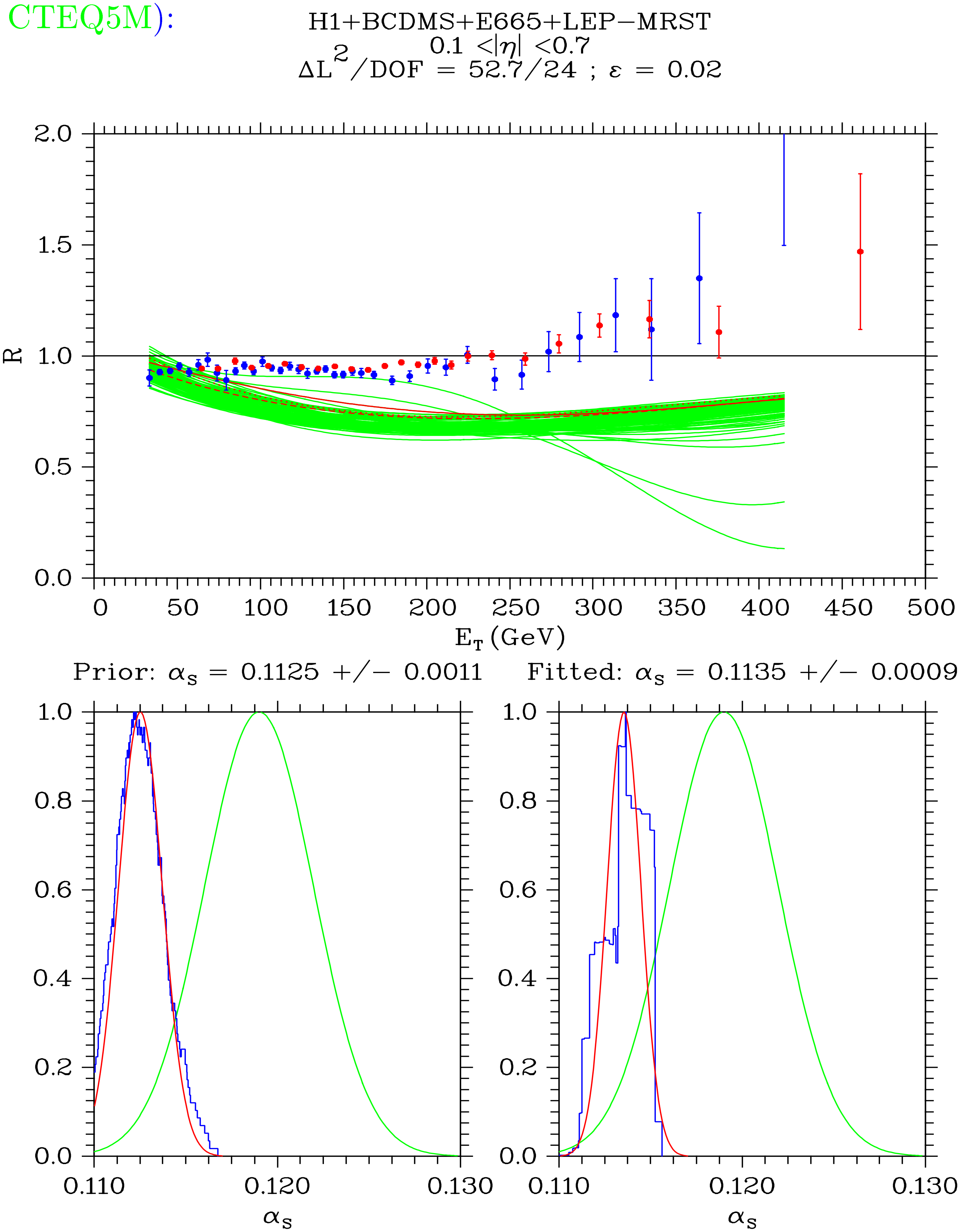
Tevatron Predictions: The one-jet inclusive

The one-jet ($\frac{d\sigma}{dE_T}$) at the Tevatron could turn out to be the measurement which will pin down the gluon PDF. Given the fact that the QCD interaction is blind to the electro-weak charges of the quarks, the measurement will in practice separate the gluon PDF from the sum of the quark PDF's as both entities carry different color charge. This combined with some carefully chosen measurements in the W and Z system which will separate the individual quark flavors could give rise to "in-house" PDF-fits. In addition the measurement also offers an opportunity to measure α_S over a wide range of momentum transfers.

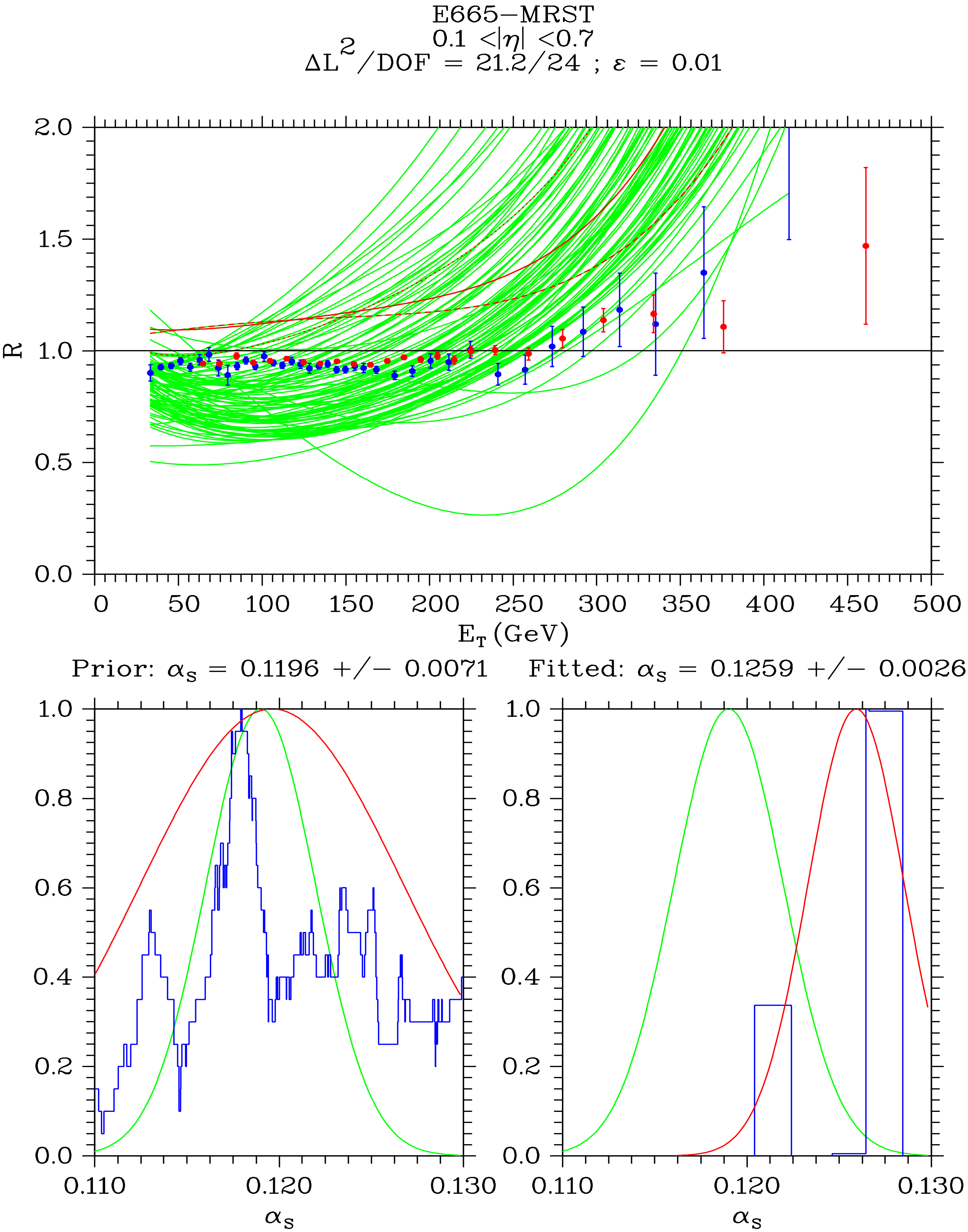
Notes:

- For now we will only show the "scatter plot". A detailed confidence level analysis is in progress.
- We will look at both the CDF and D0 measurements and compare them to two fits:
  - The *tight* `H1+BCDMS+E665+LEP-MRST` fit. This should give the smallest PDF uncertainties.
  - The *loose* `E665-MRST` fit. This should give the largest PDF uncertainties. Specifically, the value of α_S should vary over a large range.
- We will include the one-jet inclusive measurement of D0 in both fits and see the effect on the value of α_S.

The H1+BCDMS+E665+LEP-MRST comparison (the denominator is CTEQ5M):



The E665-MRST comparison (the denominator is CTEQ5M):



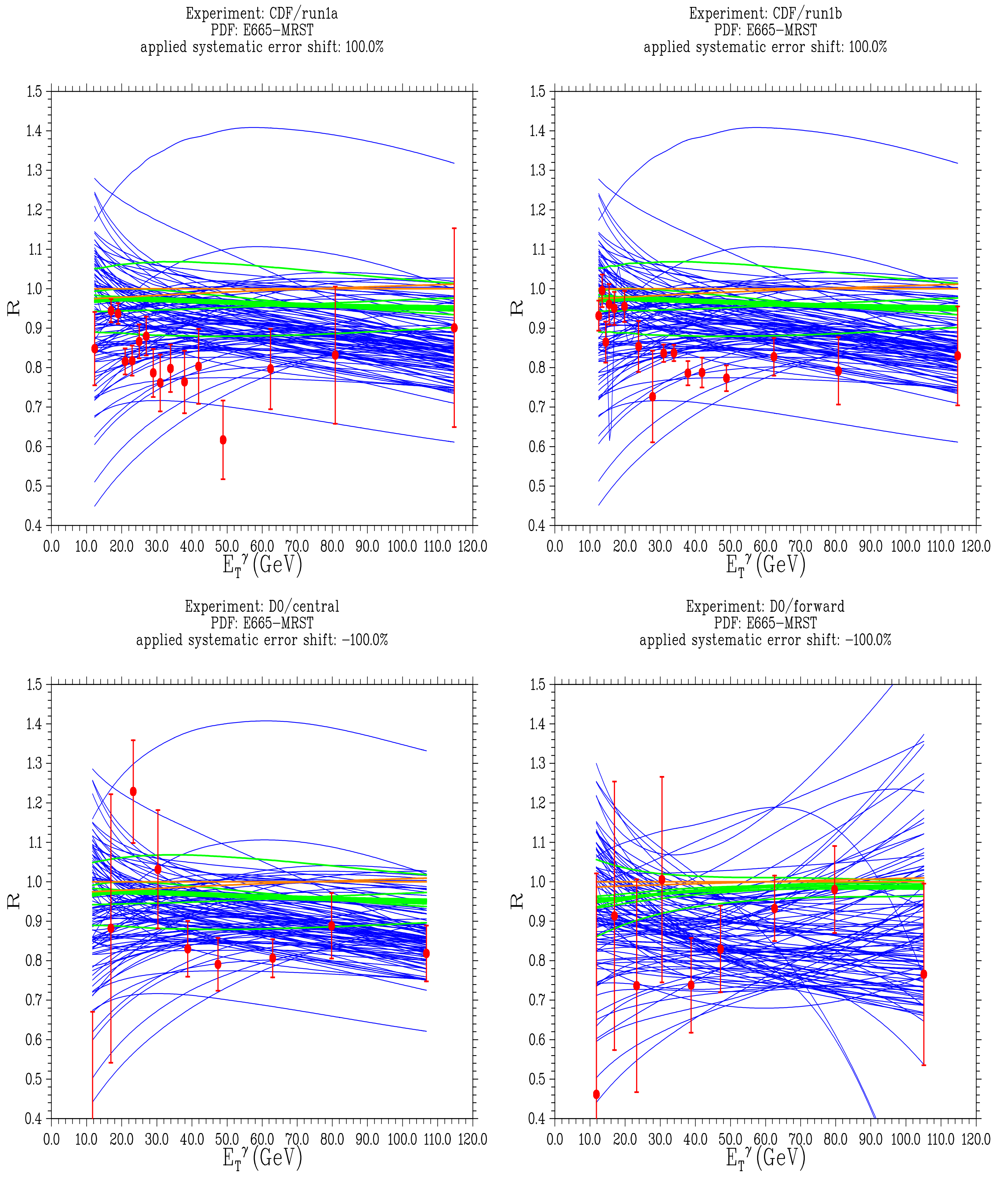
Tevatron Predictions: prompt photon production

The prompt photon production once hold the promise of pinning down the gluon distribution in parton fraction ranges relevant to hadron collider phenomenology ($qg \to \gamma q$). However, in recent years discrepancies between data and theory for a wide range of experiments have casted a dark spell on this once promising cross section which is now drowning in a swamp of non-perturbative fixes (resummation, k_T-effects,...). Can the "uncertain" PDF's come to the rescue ?

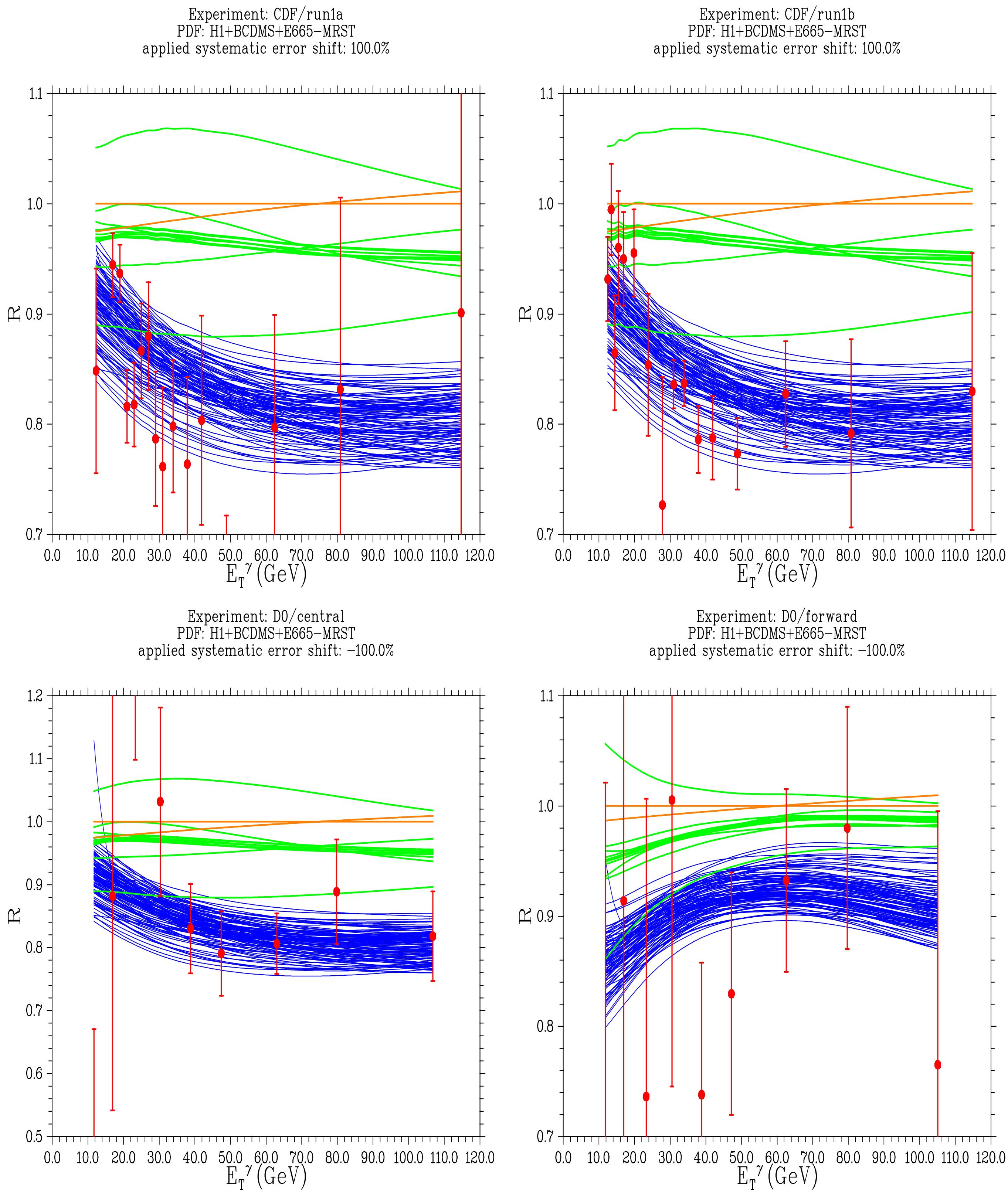
Notes:

- For now we will only show the "scatter plot". A detailed confidence level analysis is in progress.
- We will make predictions for the "loose" `E665-MRST fit` (blue curves).
- We will also make predictions for the "tight" `H1+BCDMS+E665-MRST fit` to see if the DIS proton data constrained PDF's can better explain the photon data (blue curves).
- We use the CDF run 1a/run 1b data and the D0 run 1b central/forward data (red data points). Statistical uncertainties are shown. A systematic shift is applied as indicated.
- We also make the prediction for the 12 `MRS99` PDF's (green) and the 2 `CTEQ5` PDF's (`M` and `L`) (orange).
- We will show the ratio's with respect to the `CTEQ5M` PDF.

The E665-MRST fit........... Anything goes



The H1+BCDMS+E665-MRST fit.......... Uncertainties matter



Outlook: Conclusions and Future

Conclusions:

- The PDF fits to the F_2^P data seem to describe the sampling of the Tevatron data within uncertainties pretty well but not good enough. Probably the priors need to be revisited
- In the final analysis the method is quite simple. A "fit" is merely an efficiency improvement in a monte carlo integration in PDF parameters over experimental response functions.
- All the physics is in the priors:
  - Construction of $P_{exp}(x_t|x_e)$
  - Construction of physics models to predict x_T for the relevant experiments.

  After that everything is reduced to a numerical exercise .
- The "fits" are available at `pdf.fnal.gov` with supporting software. Also examples are/will be available for analysis of data.

Future:

- One of the most urgent improvements is the PDF parameterization. It is important to probe different parameterization priors. Also we need fast predictions for all processes when we change PDF parameters: this means going to complete sets of functions and/or spline interpolated PDF's.

- Another important project is to include shadowing models in the theory prior. This will enable us to include the large set of heavy target experiments. As an additional benefit we can learn something about shadowing.

- Many other avenues of extensions are possible. Which ones will be pursued depend on the response (or lack thereof) from people using these PDF's