“Global” Determination of Parton Distributions

Walter Giele

Outline:

1. Basic properties of PDs:
   What do we need to know about them ?
2. Determination of PDs:
   How does this affect the way we make predictions ?
3. Using PDs:
   How do we use PDs ?

CTEQ summerschool
Madison, June 10 2002

Basic Properties of PDs

We need PDs ($\mathcal{F}$) for making pQCD predictions in Lepton-Hadron and Hadron-Hadron collisions:

$$d\,\sigma_{lH} = \sum_{a \in \{p\}} \int_0^1 d\,x\; \mathcal{F}_a(x, Q_F) d\,\sigma_{la}(x, Q_F)$$

$$d\,\sigma_{H_1 H_2} = \sum_{a,b \in \{p\}} \int_0^1 d\,x_1\, d\,x_2\, \mathcal{F}_a(x_1, Q_F) \mathcal{F}_b(x_2, Q_F) d\,\sigma_{ab}(x_1, x_2, Q_F)$$

- x is the fraction of the hadron momentum carried by the parton.
- Q_F is the factorization scale.
- $\{p\} = \{\bar{b}, \bar{c}, \bar{s}, \bar{u}, \bar{d}, g, d, u, s, c, b\}$ is the list of partons in the Hadron.
- At *Leading Order* the PD $\mathcal{F}$ can be interpreted as a probability density. That is $\int_a^b d\,x \mathcal{F}_p(x, Q_F)$ is the probability to find parton p in the Hadron with momentum fraction $a < x \leq b$ when probed at scale Q_F.
  Beyond leading order only the product of the PDs with the matrix elements have physical meaning.

What are the basic properties of the PD's

- The PD $\mathcal{F}(x)$ is a (analytic) function on the interval $0 < x \leq 1$ with $\mathcal{F}(1) = 0$.
- The momentum density of the PD $x\mathcal{F}(x)$ is an integrable function ($\int_0^1 d\,x\; x\mathcal{F}(x) < \infty$) and the momentum sum rule requires $\sum_{a\in\{p\}} \int_0^1 d\,x\; x\mathcal{F}_a(x) = 1$.
- The electrical charge of the Hadron enforces a charge sum rule. For instance for a proton $P = (u_v, u_v, d_v)$

$$\int_0^1 d\,x\,(\mathcal{F}_u(x) - \mathcal{F}_{\bar{u}}(x)) = 2;\;\; \int_0^1 d\,x\,(\mathcal{F}_d(x) - \mathcal{F}_{\bar{d}}(x)) = 1$$

- However the most important property for the PDs is that we can calculate the evolution order by order in pQCD (if we know $\mathcal{F}(x, Q_0)$ we know $\mathcal{F}(x, Q)$ for any Q).

$$Q\frac{\partial}{\partial Q}\mathcal{F}_a(x,Q) = \sum_{b\in\{p\}} \int_x^1 \frac{d\,z}{z}\frac{\alpha_S(Q)}{2\pi}\mathcal{F}_b(z,Q)\,P_{b\to a}(x/z)$$

This fact makes it possible to determine the PDs using a selection of experiments so we can predict the other experiments.

- Beyond leading order only the product of the PD's and the matrix element have physical meaning. This means:
  - **(a)** We lose the probability density interpretation of the PDs
  - **(b)** The mass factorization scheme in both the PDs and the matrix element calculation have to match (e.g. $\overline{\mathrm{MS}}$ or DIS scheme).
  - **(c)** The PDF evolution should be *at least* at the same order as the matrix element calculation.

Summary:

→ PDs are analytic functions, constrained by sumrules. The resulting momentum densities are integrable functions.

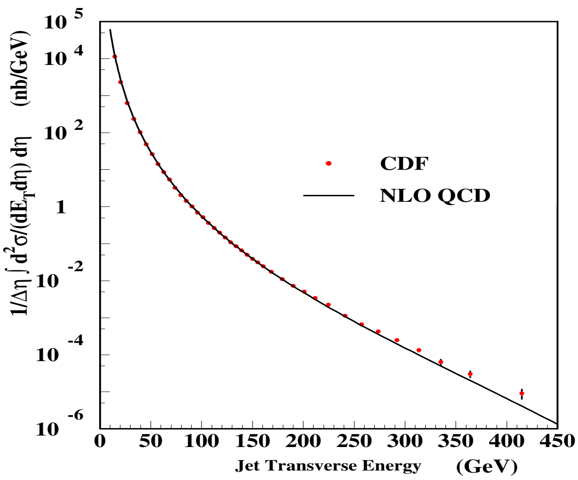
→ Evolution connects all hadronic measurements through PDs. This makes it possible to determine the PDs using only a subset of the hadronic data.

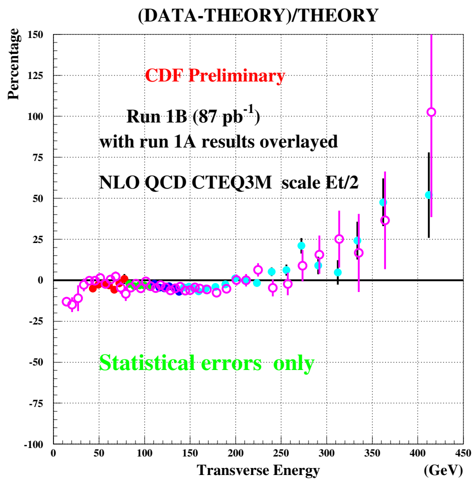
→ Beyond leading order make sure
  - ⋆ Evolution order matches matrix element calculation order
  - ⋆ Factorization scheme matches between PDs and matrix element calculation

Determination of PDs

What is the scope of PD determination ?

- Experimental uncertainties will be reflected in PDs.
- Theory uncertainties will be reflected in PDs.
- Do we need to worry about these uncertainties ?





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

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

Structure Function Subgroup in Snowmass 1996 Workshop:

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 $\cdots$

- Given the ever increasing precision of hadronic experiments (HERA, TEVATRON, LHC) PD uncertainties have become important and crucial for the final phenomenology.

What is involved in PD determination ?

1. Select observables used in the determination:
   - ⋆ We have the experimental measured values of the observable: x_e.
   - ⋆ For each PD $\mathcal{F}$ we have the theory prediction for the observable: $x_t(\mathcal{F})$.
     Note that the theory prediction has a perturbative part (LO/NLO/...) and a non-perturbative part (e.g. higher twist model/nuclear model)
   - ⋆ The experimental uncertainty in the observable gives the conditional probability density function: $P_{exp}(x_e|x_t(\mathcal{F}))$.
     Given the PD $\mathcal{F}$ what is the likelyhood the experimental measurement x_e is originating from the theory prediction $x_t(\mathcal{F})$.
     Examples:

$$P_{exp}(x_e|x_t(\mathcal{F})) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x_e - x_t(\mathcal{F}))^2/\sigma^2}$$

$$P_{exp}(x_e|x_t(\mathcal{F})) = \frac{1}{(2\pi)^{n/2}|C|} e^{-\frac{1}{2}(x_e - x_t(\mathcal{F}))_i C_{ij}^{-1}(x_e - x_t(\mathcal{F}))_j}$$

2. Define the metric in PD functional space: $P_{prior}(\mathcal{F})$

- $\star$ In its most basic formulation this means chosing a parametrization for the PD:
  Example of a paramtrization of the proton PD at a chosen initial scale of $Q_0 = 1$ GeV:

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

  From these functions we define the individual parton PDs:

$$
\begin{aligned}
\mathcal{F}_{\bar{s}}(Q_0,x) = \mathcal{F}_s(Q_0,x) &= 0.1 \times S \\
\mathcal{F}_{\bar{u}}(Q_0,x) &= 0.2 \times S - 0.5 \times \Delta \\
\mathcal{F}_{\bar{d}}(Q_0,x) &= 0.2 \times S + 0.5 \times \Delta \\
\mathcal{F}_g(Q_0,x) &= g \\
\mathcal{F}_d(Q_0,x) &= d_v + \mathcal{F}_{\bar{d}}(Q_0,x) \\
\mathcal{F}_u(Q_0,x) &= u_v + \mathcal{F}_{\bar{u}}(Q_0,x)
\end{aligned}
$$

  Heavy flavors are generated through evolution from threshold ($m_c^{treshold} = 1.5$ GeV and $m_b^{treshold} = 4.5$ GeV).

- $\star$ Of all possible functionals why this subset ?
- $\star$ Why this parametrization choice ?

3. Formulate the mathematical procedure:

We want to calculate the probability to measure a value x for an observable. To do this we need to calculate the probability density function $P_{pdf}(x)$:

- ⋆ For a given PD we can make a theory prediction for the observable with a theory uncertainty $P_{theory}(x(\mathcal{F}))$:

$$P_{pdf}(x) = P_{theory}(x|x_t(\mathcal{F}))$$

- ⋆ To calculate the PD uncertainty we consider all possible PDs constrainted by the prior assumptions:

$$P_{pdf}(x) = \int d\,\mathcal{F}\ P_{prior}(\mathcal{F}) \times P_{theory}(x|x_t(\mathcal{F}))$$

or

$$P_{pdf}(x) = \int d\,\lambda_i\ P_{theory}(x|x_t(\mathcal{F}(\lambda_i)))$$

- ⋆ However, we want to determine the PDs using another experiment:

  For each PD calculate the likelyhood it describes the experiment:

$$P_{pdf}(x) = \int d\,\mathcal{F}\ P_{prior}(\mathcal{F}) \times P_{exp}(x_e|x_t(\mathcal{F})) \times P_{theory}(x|x_t(\mathcal{F}))$$

or

$$P_{pdf}(x) = \int d\,\lambda_i\ P_{exp}(x_e|x_t(\mathcal{F}(\lambda_i))) \times P_{theory}(x|x_t(\mathcal{F}(\lambda_i)))$$

⋆ Once $P_{pdf}(x)$ is calculated we can calculate the probability of measuring the observable between x_{min} and x_{max}:

$$L(x_{min}, x_{max}) = \int_{x_{min}}^{x_{max}} d\,x\ P_{pdf}(x)$$

⋆ A crucial issue is that by redefining the parameters

$$\lambda_i \longrightarrow \mu_i(\lambda_1, \lambda_2, \ldots)$$

we generates a jacobian

$$d\,\mu_i = \left|\frac{\partial \mu_i}{\partial \lambda_j}\right| d\,\lambda_j$$

This means

$$P_{pdf}(x) = \int d\,\lambda_i\ P_{exp}(x_e|x_t(\mathcal{F}(\lambda_i))) \times P_{theory}(x|x_t(\mathcal{F}(\lambda_i)))$$

and

$$P'_{pdf}(x) = \int d\,\mu_i\ P_{exp}(x_e|x_t(\mathcal{F}(\mu_i))) \times P_{theory}(x|x_t(\mathcal{F}(\mu_i)))$$

generate different predictions for the observable **!**

No current PD set addresses this issue at the moment. The root of the problem is that PDs have infinite degrees of freedom while the data is finite. A metric/regulator has to be defined in the functional space of PDs

4. How to evaluate the formula

$$P_{pdf}(x) = \int d\,\lambda_i \; P_{exp}(x_e|x_t(\mathcal{F}(\lambda_i))) \times P_{theory}(x|x_t(\mathcal{F}(\lambda_i)))$$

such that it can be used for many observables ?

- ⋆ Somehow we have to represent $P_{exp}(x_e|x_t(\mathcal{F}(\lambda_i)))\; d\,\lambda_i$ in the PD determination solution.
- ⋆ If we are not interested in the PD uncertainty, we would determine the "best fit":

$$P_{exp}(x_e|x_t(\mathcal{F}(\lambda_i))) = \prod_i \delta\left(\lambda_i - \lambda_i^{max}\right)$$

  i.e. the most likely PD.
- ⋆ However, we want to take the PD uncertainty into account. The most straightforward way of accomplishing this is using a monte carlo approach:

$$P_{pdf}(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \; P_{exp}(x_e|x_t(\mathcal{F}(\lambda_i^{(j)}))) \times P_{theory}(x|x_t(\mathcal{F}(\lambda_i^{(j)})))$$

- ⋆ This would be inefficient and we would need a large sample of PDs. The standard procedure in this case is to draw the PDs from the probability density function $P_{exp}(x_e|x_t(\mathcal{F}(\lambda_i^{(j)})))$ using sequential monte carlo techniques (e.g. metropolis algorithm). Such a method is often called random sampling.
- ⋆ Note that formally we have redefined the PD parameters such that $d\,\mu_i = P_{exp}(x_e|x_t(\mathcal{F}(\lambda_i)))\; d\,\lambda_i$ giving

$$P_{pdf}(x) = \int d\,\mu_i \; P_{theory}(x|x_t(\mathcal{F}(\mu_i)))$$

⋆ Now we get:

$$P_{pdf}(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} P_{theory}(x|x_t(\mathcal{F}(\mu_i^{(j)}))$$

where the PD parameter sets $\{\mu_i^{(j)}\}$ are distributed according to $P_{exp}(x_e|x_t(\mathcal{F}(\mu_i^{(j)})))$.

⋆ This formula simply means we evaluate the observable for the N PD sets and take the "average".

⋆ By making the approximation that

$$P_{exp}(x_e|x_t(\mathcal{F}(\mu_i))) \sim e^{-\frac{1}{2}\chi^2(\mu_i)}$$

with

$$\chi^2 = \chi^2_{min} + (\chi^2(\mu_i) - \chi^2_{min})C_{ij}(\chi^2(\mu_j) - \chi^2_{min})$$

we get a more traditional approach:

→ Find the best fit, i.e. χ^2_{min}

→ Calculate the local neighborhood around the minimum using a Taylor expansion, i.e. how fast is χ^2 changing if we move the PD parameters away from the best fit.

This way we can calculate the gaussian approximation to $P_{exp}(x_e|x_t(\mathcal{F}(\mu_i)))$ simply using a minimization software package avoiding the need of sequential monte carlo techniques.

Summary:

→ Hadron collider experiments have reached a level of accuracy that requires to take the PD uncertainties into account for the theory predictions.

→ We formulated the PD determination in an integration equation (ie non-algorithmic) allowing precise definitions of uncertainties. A mathematical formulation opens the way to further refinement in methodology.

→ A currently unresolved issue is the parametrization dependence of the PD determination.

→ Random sampling is an intuitive, flexible and easy way to propagate the PD uncertainties to observables.

→ We currently can implement the integration over PD-space using either the gaussian approximation or addaptive monte carlo techniques.

Using PDs

Let us now determine PDs so we can predict the W and Z cross sections at the Tevatron:

- Evaluate all matrix elements at NLO.
- Select the appropriate DIS experiments to determine the PDs.
- Compare to D0 run 1 results.
- We'll use adaptive MC approach and random sampling propagation of PD uncertainties.
- we use the parametrization of previous section.

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 with the other experiments:

	H1		ZEUS		BCDMS		E665		LEP	
	L^2	CL	L^2	CL	L^2	CL	L^2	CL	L^2	CL
H1-fit	-	-	609	0.5%	364	67%	68	21%	1.9	31%
ZEUS-fit	219	22%	-	-	568	0.1%	77	5%	1.6	24%
BCDMS-fit	188	85%	556	1.5%	-	-	70	23%	10	0.5%
E665-fit	236	30%	783	1.6%	398	82%	-	-	1.8	99%

Notes:

- Average log likelihoods ($L^2 = -2\log(Probability)$) do not measure goodness of fit, but confidence levels do (CL: given the theory how likely is it a repeat of the measurement will give a worse agreement).
- ZEUS does not want to be combined with other experiments.
- BCDMS has a problem with the LEP α_S value.

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.

- We will use the D0 run 1a and run 1b results:

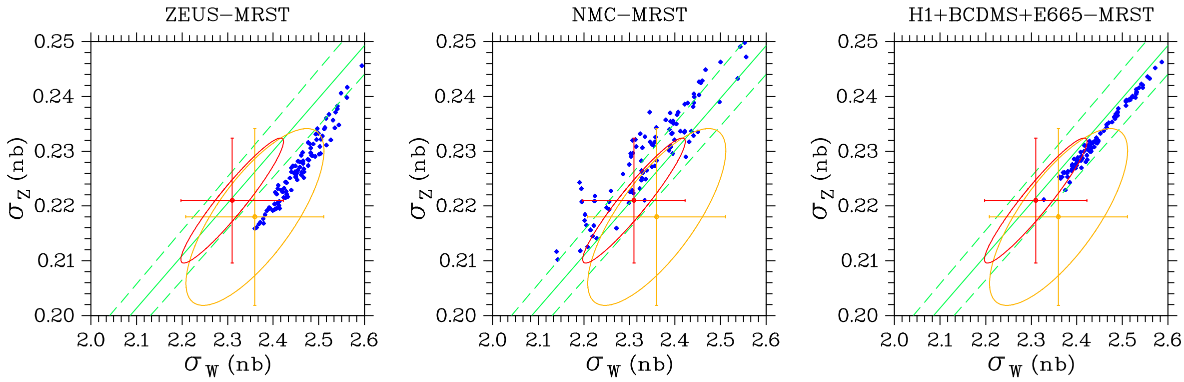
	D0 1a	D0 1b
σ_W (nb)	2.36±0.02±0.08±0.13	2.31±0.01±0.05±0.10
σ_Z (nb)	0.218±0.008±0.008±0.011	0.221±0.003±0.004±0.010
R	10.82±0.41±0.35	10.43±0.20±0.10
$\mathcal{L}^{exp}$ (pb^{-1})	12.8±0.7	84.5±3.7

- For the measured D0 run 1a and run 1b cross section values we find for $CL(\sigma_V^{measured})$ (in percentages):

	D0 1a		D0 1b	
	$\sigma(W)$	$\sigma(Z)$	$\sigma(W)$	$\sigma(Z)$
ZEUS-MRST	51.4	57.6	23.4	64.1
NMC-MRST	99.9	51.7	75.1	53.0
H1+BCDMS+E665-MRST	55.7	37.4	26.1	33.9

- Conclusions: The individual W and Z cross sections are very well predicted by the NLO theory using the three chosen PDF sets. This gives us confidence to use these two cross sections to extract the D0 run 1a and run 1b integrated luminosities.

- We want to extract the D0 run 1a and run 1b integrated luminosity:



- The luminosity experimental response function is given by

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

where

$$D = (\mathcal{L} \times \sigma_W - N_W, \mathcal{L} \times \sigma_Z - N_Z)$$

and C_{ij} the error correlation matrix determined from the W boson, Z boson and ratio R measurements.

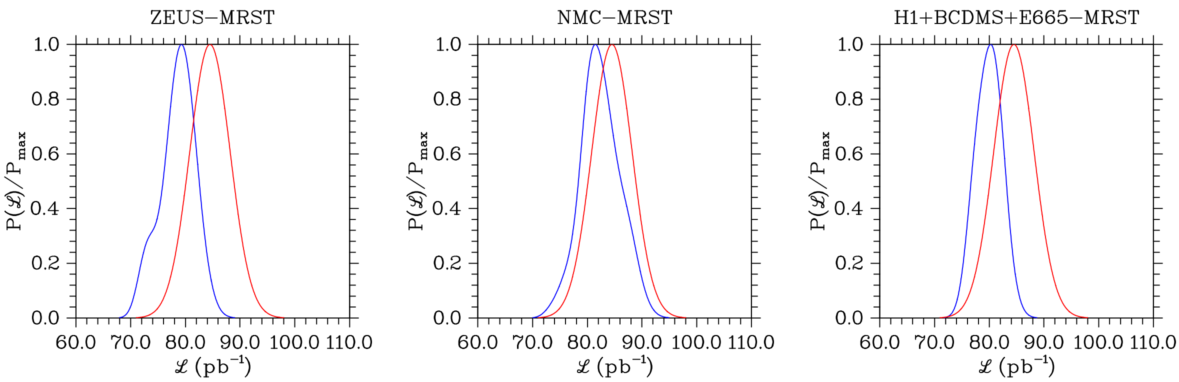
- The theoretical scatter plot representation is given by:

$$P_{pdf}(\sigma_W, \sigma_Z) = \frac{1}{N} \sum_{i=1}^{N} \delta\left(\sigma_W - \sigma_W(\mathcal{F}_i)\right) \delta\left(\sigma_Z - \sigma_Z(\mathcal{F}_i)\right)$$

- Combining this gives us:

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

- Now we can determine the integrated luminosity.....



Red is D0 run 1b quoted integrated luminosity, blue is $P_{pdf}(\mathcal{L}|N_W, N_Z)$

- The confidence lever for agreement between data and theory is (the CL in percentages):

	D0 1a		D0 1b	
	$\mathcal{L}$ (pb^{-1})	CL^{exp}	$\mathcal{L}$ (pb^{-1})	CL^{exp}
$\sigma_{tot}(P\bar{P})$	12.8+0.7-0.7	-	81.5+3.7-3.7	-
ZEUS-MRST	12.3+0.5-0.5	96	79.3+3.1-3.3	4.6 (8.6)
NMC-MRST	12.6+0.7-0.6	26	81.5+4.3-2.9	36 (29)
H1+BCDMS+E665-MRST	12.3+0.5-0.5	54	80.3+2.3-2.8	67 (81)

where $CL^{exp} = CL(N_W, N_Z)$ and in for the run 1b results in brackets the CL^{exp} value using the NNLO matrix element predictions.

- To conclude we see the hard scattering based integrated luminosity measurement outdoing the traditional luminosity measurement. The results are consistent with published D0 integrated luminosities.

A fit including PDF uncertainties is composed of a set of individual PDF's. The LHAPDF interface is designed to deal with such sets:

- Each set is defined through an individual external file (in parametric form)
- Simple to download, setup and use
- Currently evolution code QCDNUM and/or EVLCTEQ included
- Downloadable from **pdf.fnal.gov**

Current available PDF sets:

- Alekhin set (hep-ph/0011002):
  DIS fit, gaussian approx, variable α_S, higher twist model, restricted functional forms.
- Botje set (hep-ph/9912439):
  DIS fit, gaussian approx, fixed α_S, higher twist model, diagonalized experimental uncertainties.
- CTEQ6 set (hep-ph/0201195):
  "Global fit", gaussian approx, fixed α_S, higher "tolerance", no random sampling error propagation.
- Fermi2001 set (hep-ph/0104052):
  Restricted DIS fit, variable α_S
- (MRST98 (hep-ph/9803445), MRST2001 (hep-ph/0110215):
  These sets are best "global fit" sets, no uncertainties.)

Summary:

→ Limited sets of experiments can be easily combined in a single PD determination and used for interesting phonomenology.

→ More global fits such as CTEQ6 require a method to include different experiments which can disagree:
CTEQ6 uses "tolerance" (ie multiply χ^2)

→ The LHAPDF fortran code enables easy access to PD sets with an uniform interface.

This concludes the lecture and should have prepared you to use modern PDs and critically evaluate them.

→ We formulated the problem of PD determination.

→ We derived a mathematical formula describing the PD determination.

→ We determined a PD set based on a few DIS experiments and successfully used them to predict the D0 W and Z cross sections.

→ For "global fits" one has to deal with the reality of sets of experiments which does not allow a single consistent PD determination.

→ A fortran code can be downloaded from **pdf.fnal.gov** to use current PDs.