Extracting PDF from LHC data: The ApplGrid project

Tancredi Carli, Dan Clements, Amanda Cooper-Sarkar, Claire Gwenlan, Gavin Salam, Pavel Starovoitov, Mark Sutton

CERN, Glasgow, LPTHE, NCPHEP, Oxford, UCL

February 22, 2008

Outline

1. Motivation
2. Formalism
3. Jet production
4. W^+-production
5. Z^0-production
6. Accuracy studies
7. Conclusions & Next Steps

1. Motivation.

- "New Physics" might be discovered via deviation of precise cross section measurements at LHC from NLO QCD predictions.
- Calculating NLO Cross-Sections takes a long time ($\sim$ *days*).
- Since perturbative coefficients are essentially independent from PDF functions due to factorization theorem, we can split calculation into two parts
  - Step 1: We collect perturbative weights to grids (long run).
  - Step 2: We convolute grid with PDF's ($\sim$1ms).
- This technique allows
  - fast study of impact of PDF uncertainty on cross section measurements
  - fitting of PDF from observable cross section in a consistent way

2. Formalism. Variables

Instead of using the parton momentum fraction x and the factorisation scale Q^2, we use a variable transformation that provides good coverage of the full x and Q^2 range

$$
\begin{aligned}
y(x) &= \ln \frac{1}{x} + a(1-x) \\
\tau(Q^2) &= \ln\left(\ln \frac{Q^2}{\Lambda^2}\right)
\end{aligned}
$$

the a-parameter serves to increase the density of points in the large x region.

$\Lambda \sim \Lambda_{\mathrm{QCD}}$, but not necessarily identical.

2. Formalism. DIS case

We have the set of events $m = 1 \dots N$ produced by a NLO MC generator. Each event is characterized by the momentum fraction x_m, the factorization scale Q_m^2, the weight w_m and the pQCD order p_m.
The result W of the standard Monte Carlo integration:

$$W = \sum_{m=1}^{N} w_m \left(\frac{\alpha_s(Q_m^2)}{2\pi} \right)^{p_m} q(x_m, Q_m^2).$$

Instead one can introduce a weight grid updated on event-by-event basis

$$W_{k+i,\kappa+\iota}^{(p_m)} \to W_{k+i,\kappa+\iota}^{(p_m)} + w_m \, I_i^{(n)} \left(\frac{y(x_m)}{\delta y} - k \right) I_\iota^{(n')} \left(\frac{\tau(Q_m^2)}{\delta \tau} - \kappa \right),$$

where $I_i^{(n)}(\dots)$ – interpolation functions

2. Formalism. DIS case. Master formula

.. and saved to the ROOT-file at the end of the MC run

$$W^{(p)}_{i_y,i_\tau} \rightarrow TH2D[p][observable\ bins]\left(y \equiv x, \tau \equiv Q^2\right)$$

Then, the observable could be reconstructed 'a posteriori'

$$W = \sum_p \sum_{i_y} \sum_{i_\tau} W^{(p)}_{i_y,i_\tau} \left(\frac{\alpha_s\left(Q^{2\,(i_\tau)}\right)}{2\pi}\right)^p q\left(x^{(i_y)}, Q^{2\,(i_\tau)}\right)$$

for any PDF set

$$ApplGrid \quad \Rightarrow \quad W^{(p)}_{i_y,i_\tau} \quad \Rightarrow \quad W \quad \Leftarrow \quad q\left(x, Q^2\right) \quad \Leftarrow \quad LHAPDF\ or\ fit$$

technique developed at HERA: H1 EPJ 19(2001)289, ZEUS EPJ C42(2005)1, M.Wobbisch PhD Thesis

2. Formalism. *pp* collisions

In *pp* scattering one can use analogous procedures with one more dimension. Besides Q^2, the weight grid depends on the momentum fraction of the first x_1 and second x_2 hadron.

$$W = \sum_{m=1}^{N} w_m \left(\frac{\alpha_s(Q_m^2)}{2\pi} \right)^{p_m} q_1\left(x_{1\,m}, Q_m^2\right) q_2\left(x_{2\,m}, Q_m^2\right).$$

Since the w_m weight is the same for some parton flavor combinations, it is possible to reorganize cross section formula to the sum of contributions from different subprocesses specific to the process under investigation.

$$W = \sum_{l=0}^{L} \sum_{n_l=1}^{M} w_{n_l} \left(\frac{\alpha_s(Q_{n_l}^2)}{2\pi} \right)^{p_{n_l}} F^{(l)}\left(x_{1\,n_l}, x_{2\,n_l}, Q_{n_l}^2\right),$$

where $F^{(l)}(\ldots)$ – generalized PDF's

2. Formalism. Summary

Look-up tables of perturbative coefficients could be produced for a posteriori cross section calculation.

Depending on what should be calculated one can choose

- pQCD order
- number of subprocesses

Accuracy can be optimized using grid parameters

- number of bins in $x_{1,2}$ and Q^2
- interpolation order in $x_{1,2}$ and Q^2
- bin spacing in high x(a parameter)
- bin spacing in Q^2 (Λ)
- weighting by the factor approximating PDF in a region with steep gradient

3. Jet production. Sub-processes

In the case of jet production in proton-proton collisions the weights generated by the Monte Carlo program as well as the PDFs can be organised in seven possible initial state combinations of partons:

$$
\begin{aligned}
\mathrm{gg}: \quad F^{(0)}(x_1, x_2; Q^2) &= G_1(x_1) G_2(x_2) \\
\mathrm{qg}: \quad F^{(1)}(x_1, x_2; Q^2) &= \left(Q_1(x_1) + \overline{Q}_1(x_1)\right) G_2(x_2) \\
\mathrm{gq}: \quad F^{(2)}(x_1, x_2; Q^2) &= G_1(x_1) \left(Q_2(x_2) + \overline{Q}_2(x_2)\right) \\
\mathrm{qr}: \quad F^{(3)}(x_1, x_2; Q^2) &= Q_1(x_1) Q_2(x_2) + \overline{Q}_1(x_1) \overline{Q}_2(x_2) - D(x_1, x_2) \\
\mathrm{qq}: \quad F^{(4)}(x_1, x_2; Q^2) &= D(x_1, x_2) \\
\mathrm{q\bar{q}}: \quad F^{(5)}(x_1, x_2; Q^2) &= \overline{D}(x_1, x_2) \\
\mathrm{q\bar{r}}: \quad F^{(6)}(x_1, x_2; Q^2) &= Q_1(x_1) \overline{Q}_2(x_2) + \overline{Q}_1(x_1) Q_2(x_2) - \overline{D}(x_1, x_2),
\end{aligned}
$$

where g denotes gluons, q quarks and r quarks of different flavour $q \neq r$

3. Jet production. Master formula

Weights are saved to a ROOT-file (weight-grids)

$$W^{(p)(l)}_{i_{y_1},i_{y_2},i_\tau} \rightarrow TH3D[p][l][observable\ bins]\left(y_1 \equiv x_1, y_2 \equiv x_2, \tau \equiv Q^2\right)$$

and the final result

$$W = \sum_p \sum_{l=0}^{6} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} W^{(p)(l)}_{i_{y_1},i_{y_2},i_\tau} \left(\frac{\alpha_s\left(Q^{2\,(i_\tau)}\right)}{2\pi}\right)^p F^{(l)}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, Q^{2\,(i_\tau)}\right)$$

is reconstructed from the weight-grids and PDF sets.

3. Jet production. Scale dependence I

Having the weights $W^{(p)(l)}_{i_{y_1},i_{y_2},i_\tau}$ determined separately order by order in α_s, it is straightforward to vary the renormalization μ_R and factorization μ_F scales a posteriori.
We assume scales to be equal

$$\mu_R = \mu_F = Q$$

in the original calculation.
Let we introduce ξ_R and ξ_F corresponding to the factors by which one varies μ_R and μ_F respectively,

$$\mu_R = \xi_R \times Q$$
$$\mu_F = \xi_F \times Q$$

3. Jet production. Scale dependence II

Then for arbitrary ξ_R and ξ_F we may write:

$$W \;(\xi_R, \xi_F) = \sum_{l=0}^{6} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} \left\{ \left(\frac{\alpha_s\left(\xi_R^2 Q^{2(i_\tau)}\right)}{2\pi} \right)^{p_{\rm LO}} \right.$$

$$\times W^{(p_{\rm LO})(l)}_{i_{y_1}, i_{y_2}, i_\tau} F^{(l)}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}\right) + \left(\frac{\alpha_s\left(\xi_R^2 Q^{2(i_\tau)}\right)}{2\pi} \right)^{p_{\rm NLO}}$$

$$\times \left[\left(W^{(p_{\rm NLO})(l)}_{i_{y_1}, i_{y_2}, i_\tau} + 2\pi\beta_0 p_{\rm LO} \ln \xi_R^2 \, W^{(p_{\rm LO})(l)}_{i_{y_1}, i_{y_2}, i_\tau} \right) F^{(l)}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}\right) \right.$$

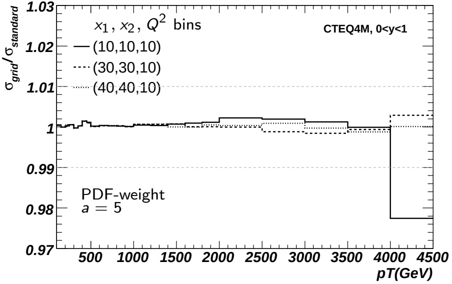
$$- \ln \xi_F^2 \, W^{(p_{\rm LO})(l)}_{i_{y_1}, i_{y_2}, i_\tau}$$

$$\left. \left. \left(F^{(l)}_{q_1 \to P_0 \otimes q_1}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}\right) + F^{(l)}_{q_2 \to P_0 \otimes q_2}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}\right)\right) \right] \right\},$$

where $F^{(l)}_{q_1 \to P_0 \otimes q_1}$ is calculated as $F^{(l)}$, but with q_1 replaced with $P_0 \otimes q_1$(LO splitting function convoluted with PDF), and analogously for $F^{(l)}_{q_2 \to P_0 \otimes q_2}$.

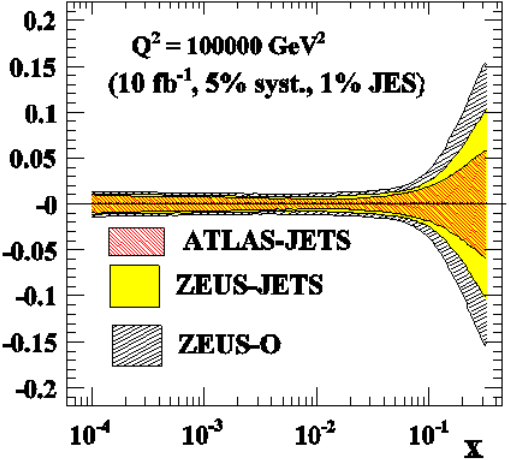
(see hep-ph/0510324)

3. Jet production. Performance studies



Grids can be kept small except regions where PDF is steeply falling

3. Jet production. Gluon uncertainty



Results of ZEUS-fit to extract PDF's from
1) F_2-data; 2) F_2-data + DIS-jets; 3) F_2-data + DIS-jets + pp-jets
using high statistics weight-grids

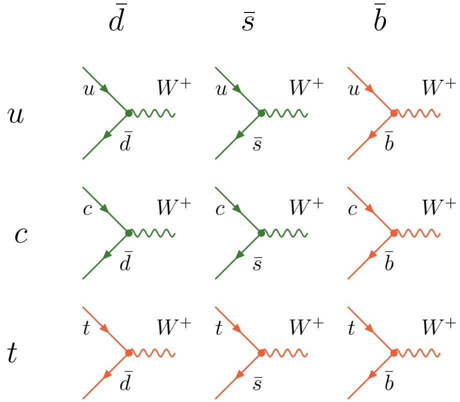
3. Jet production. Summary

- We are able to produce grids of perturbative coefficients for jet production with precision better than 0.02% (systematic uncertainty)
- To get grid with small statistical uncertainty one has to make long calculations
  - parallel jobs running LCG-grid
- Cross section is calculated in short time
  - allows study PDF uncertainty
  - allows study renormalisation/factorisation scale dependance
- Grids are used in a fit of PDF of HERA+LHC jet data

Recently: Extension of this technique for W/Z-production using MCFM

4. W^+-production. Sub-Processes @LO

Organisation of sub-processes for electro-weak production(massless calculation):

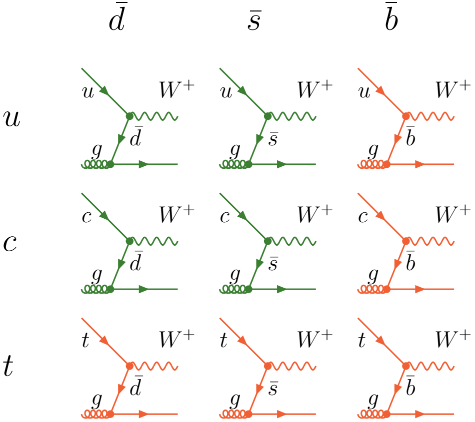


The LO processes can be treated in the same way using CKM matrix elements V_{ij}

$$F^{(l)}(x_1, x_2) = \sum_{i=u,c,t} f_{i/H_1}(x_1) \sum_{j=d,s,b} f_{\bar{j}/H_2}(x_2) V_{ij}^2$$

4. W^+-production. Sub-Processes @NLO

In NLO the situation is a bit more complicated. As example:
Ug-subprocess



The generalized PDF for this subprocess is

$$F^{(I)}(x_1, x_2) = \sum_{i=u,c,t} f_{i/H_1}(x_1)\left(V_{id}^2 + V_{is}^2 + V_{ib}^2\right) f_{g/H_2}(x_2)$$

4. W^+-production. Sub-Processes

The weights for W^+-production can be organized in six possible initial state combinations:

$$\bar{D}U: \quad F^{(0)}\left(x_1, x_2, Q^2\right) = \sum_{j=1,3,5} f_{-j/H_1}\left(x_1\right) \sum_{i=2,4,6} f_{i/H_2}\left(x_2\right) V_{ij}^2$$

$$U\bar{D}: \quad F^{(1)}\left(x_1, x_2, Q^2\right) = \sum_{i=2,4,6} f_{i/H_1}\left(x_1\right) \sum_{j=1,3,5} f_{-j/H_2}\left(x_2\right) V_{ij}^2$$

$$Ug: \quad F^{(3)}\left(x_1, x_2, Q^2\right) = \sum_{i=2,4,6} f_{i/H_1}\left(x_1\right) \left(V_{id}^2 + V_{is}^2 + V_{ib}^2\right) f_{0/H_2}\left(x_2\right)$$

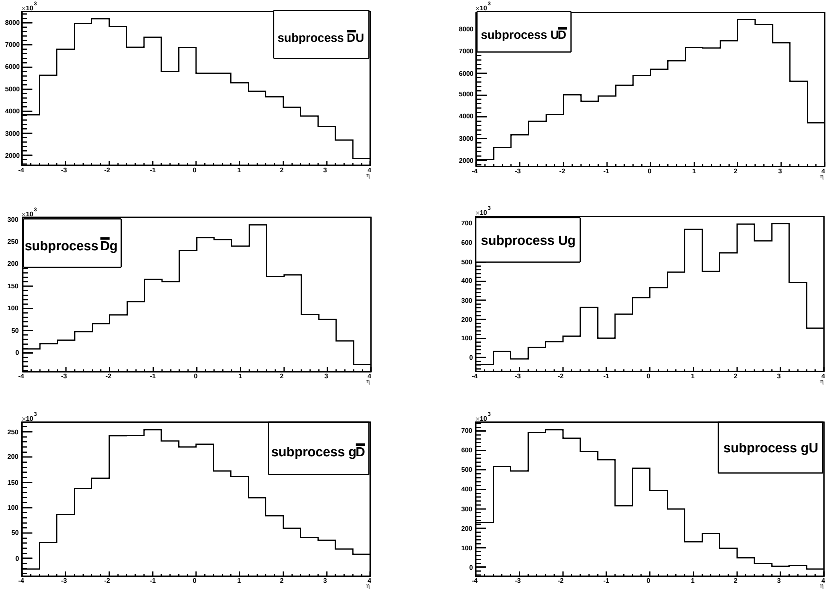
$$gU: \quad F^{(5)}\left(x_1, x_2, Q^2\right) = \sum_{i=2,4,6} f_{0/H_1}\left(x_1\right) \left(V_{id}^2 + V_{is}^2 + V_{ib}^2\right) f_{i/H_2}\left(x_2\right)$$

$$g\bar{D}: \quad F^{(4)}\left(x_1, x_2, Q^2\right) = \sum_{i=1,3,5} f_{0/H_1}\left(x_1\right) \left(V_{iu}^2 + V_{ic}^2 + V_{it}^2\right) f_{-i/H_2}\left(x_2\right)$$

$$\bar{D}g: \quad F^{(2)}\left(x_1, x_2, Q^2\right) = \sum_{i=1,3,5} f_{-i/H_1}\left(x_1\right) \left(V_{iu}^2 + V_{ic}^2 + V_{it}^2\right) f_{0/H_2}\left(x_2\right)$$

W separate $u\bar{d}$ from $\bar{d}u$ in order to get the right rapidity distribution for the electron, because of the chiral nature of the $W^\pm$ couplings

4. W^+-production. SubProcesses Contributions



5. Z^0-production. Sub-Processes

We can introduce 12 sub-processes in Z production

γ/Z interference is included

$$
\begin{aligned}
U\bar{U}: &\quad F^{(0)}\left(x_1,x_2,Q^2\right) = U_{12}(x_1,x_2) \\
D\bar{D}: &\quad F^{(1)}\left(x_1,x_2,Q^2\right) = D_{12}(x_1,x_2) \\
\bar{U}U: &\quad F^{(2)}\left(x_1,x_2,Q^2\right) = U_{21}(x_1,x_2) \\
\bar{D}D: &\quad F^{(3)}\left(x_1,x_2,Q^2\right) = D_{21}(x_1,x_2) \\
gU: &\quad F^{(4)}\left(x_1,x_2,Q^2\right) = G_1(x_1)U_2(x_2) \\
g\bar{U}: &\quad F^{(5)}\left(x_1,x_2,Q^2\right) = G_1(x_1)\overline{U}_2(x_2) \\
gD: &\quad F^{(6)}\left(x_1,x_2,Q^2\right) = G_1(x_1)D_2(x_2) \\
g\bar{D}: &\quad F^{(7)}\left(x_1,x_2,Q^2\right) = G_1(x_1)\overline{D}_2(x_2) \\
Ug: &\quad F^{(8)}\left(x_1,x_2,Q^2\right) = U_1(x_1)G_2(x_2) \\
\bar{U}g: &\quad F^{(9)}\left(x_1,x_2,Q^2\right) = \overline{U}_1(x_1)G_2(x_2) \\
Dg: &\quad F^{(10)}\left(x_1,x_2,Q^2\right) = D_1(x_1)G_2(x_2) \\
\bar{D}g: &\quad F^{(11)}\left(x_1,x_2,Q^2\right) = \overline{D}_1(x_1)G_2(x_2)
\end{aligned}
$$

We separate $u\bar{u}$ from $\bar{u}u$ contributions

6. Accuracy studies. Observables

In W^+ production we define three observables:

$$\frac{d\sigma}{d\eta_e}, \qquad \left.\frac{d\sigma}{dp_{te}}\right|_{|\eta_e|\leq 0.5}, \qquad \left.\frac{d\sigma}{dp_{te}}\right|_{|\eta_e|\geq 3.0}, \tag{1}$$

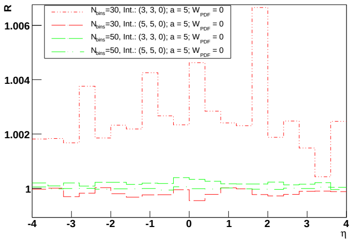
pseudorapidity distribution, p_t distribution in central rapidity bin and p_t in forward region of final state positrons, respectively.

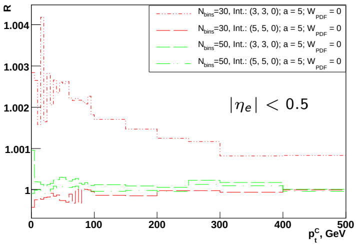
In order to study the influence of grid parameters on a observble we define the ratio

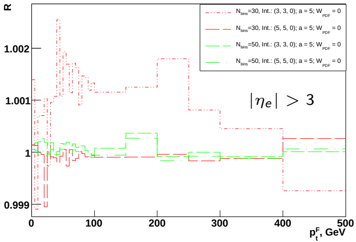
$$R = \frac{d\sigma_{grid}}{d\sigma^{reference}} \tag{2}$$

of the observable, calculated via grid, to the reference observable, calculated by MCFM.

6. Accuracy studies. Binning & Interpolation order

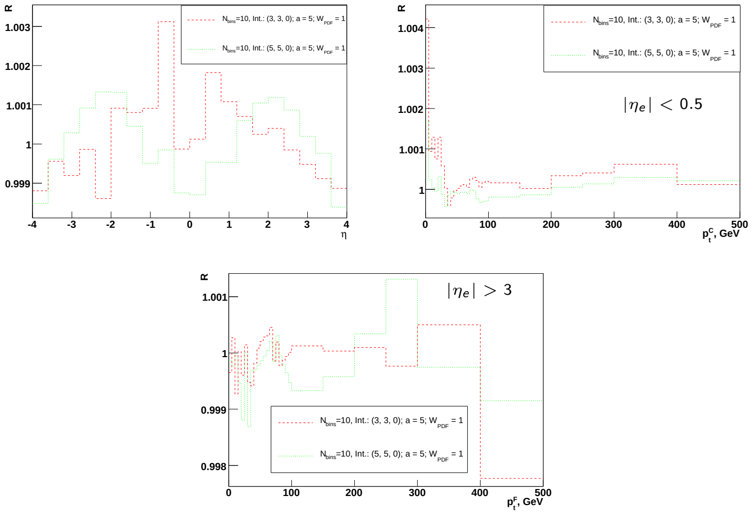






Accuracy is good for all parameters. For more bins and higher interpolation order accuracy is better than 0.001

6. Accuracy studies. Small grid size



Even with small grids accuracy can be attained using PDF weighting in the grid filling step.

7. Conclusions & Next Steps.

- We can produce tables of perturbative coefficients for jet and W/Z production in pp collisions in LO and NLO orders with accuracy better then 0.02%
- Interface to MCFM recently done. Most calculated processes are available now
- Performance of the grids for the process of interest could be optimized by a user via set of parameters
- Still on-going work on software optimisation: store smallest possible grid that takes the least time