About PROFFIT

A. Knutsson (DESY)

24th January 2011, DESY

Outline

- Short intro to Automized Tuning
- About PROFFIT
- Application 1: Fits of unintegrated PDFs to HERA data
- Application 2: A fragmentation tune to HERA data
- Summary

Short intro to Automized Tuning

An alterntive approach for fitting

Former fitting method:

Based on running the generator in an **iterative procedure** in parameter space.

→ **Time consuming for exclusive final states. A high statistics MC run can take more than 24h, and *O*(100) iterations needed to find minimum.**

An alterntive approach for fitting

Former fitting method:

Based on running the generator in an **iterative procedure** in parameter space.

→ **Time consuming for exclusive final states. A high statistics MC run can take more than 24h, and *O*(100) iterations needed to find minimum.**

Automatized Tuning Approach:

Describe **parameter dependence analytically before the parameter fitting**, by building up a ***grid of MC predictions in parameter space***.

The MC grid points can be calculated simultaneously.

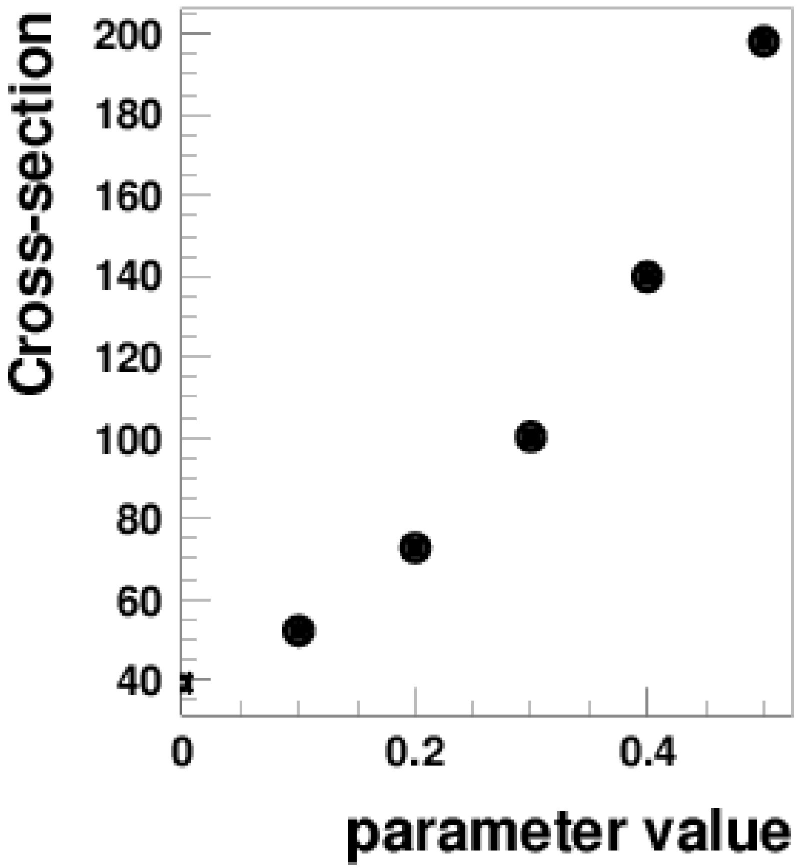
The fitting itself then takes a few seconds.

Automated Tuning – Simple Example

Simplest possible example

1 parameter, 1 data observable

1. Build up the MC grid

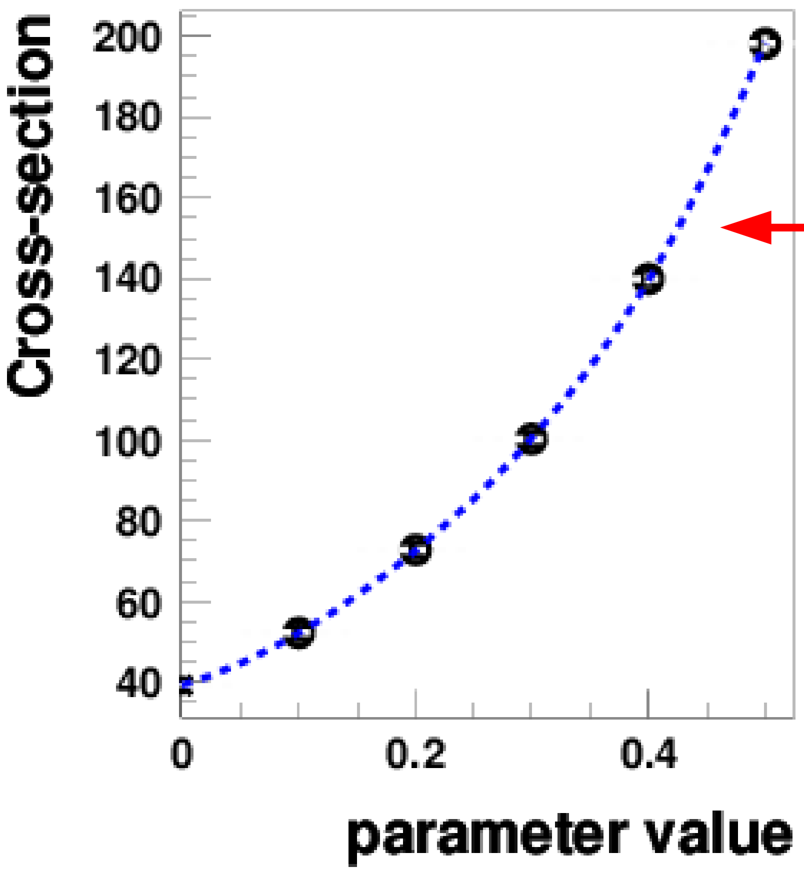


Monte Carlo predictions

Equidistant or random grid points.

Simplest possible example
1 parameter, 1 data observable

2. Determine polynomial using SVD



Monte Carlo predictions

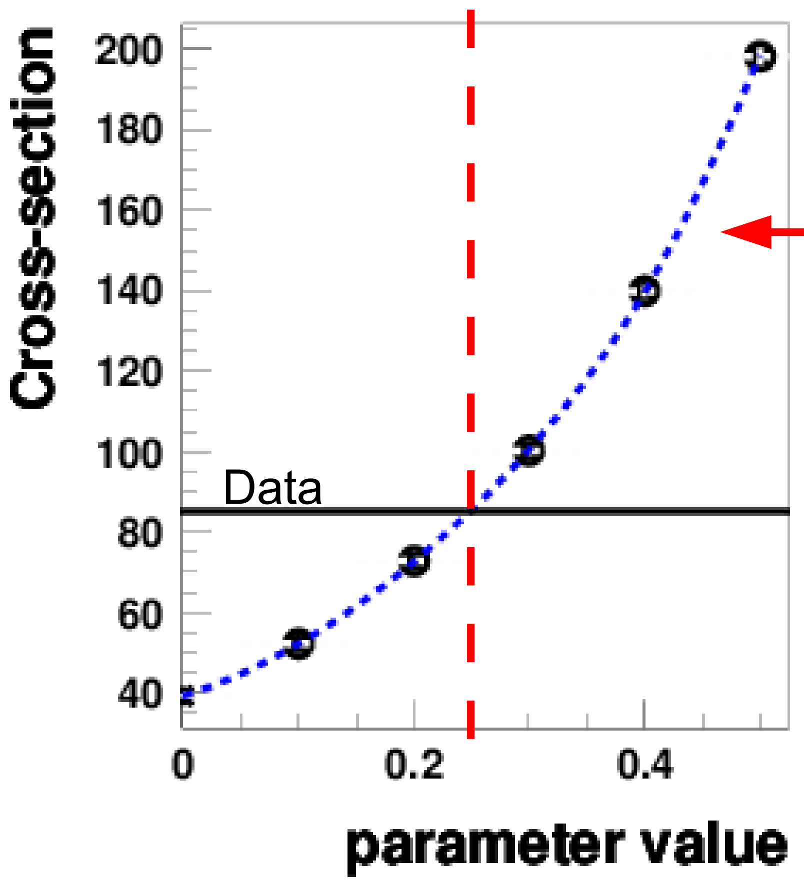
Polynomial parameterisation of the MC

Automated Tuning – Simple Example

Simplest possible example

1 parameter, 1 data observable

3. Minimize Chi2 to data



Monte Carlo predictions

Polynomial fit

Fit polynomial to data cross-section

Very fast!
Getting *O*(100) predictions from a polynomial clearly much faster than running the MC 100 times iteratively.

About PROFFIT

About PROFFIT

- A couple a years ago Hendrik Hoeth (PROFESSOR) gave a talk about MC tuning at the MCNet school in Durham

- We wanted to try the method for fits of the unintegrated PDFs for which a standard iterative fitting method is too time consuming, in particular when determining the kt-dependence in the uPDF. (Need *O*(100) iterations (MC runs) to find minimum. Need decent statistics: 1 MC run *O*(12h).)

- In addition we needed a proper error treatment for the PDFs.
  Error treatment based on CTEQ method, with modified Chi2 calculation to take care of correlated and uncorrelated errors.

- The data we wanted to use existed already in analyses routines in the fortran based HZTOOL framework.

- Takes data and MC predictions from: ascii-tables, hbooks or root-files NEW

PROFFIT

- Publication last autumn:
  Bachetta, Jung, Knutsson, Kutak
  "A method for tuning parameters of MC generators and a determination of the unintegrated gluon desnity"
  EPJC,70 (2010) 503

- Currently a (slow) solo project, but not forgotten. Need to be stream lined and made more user friendly.

- Available on request.

Fits of uPDFs

- *PoS DIS2010:043,2010*
- *Eur Phys J, C70:503-515, 2010*
- *Eur.Phys.J.C70:1237-1249,2010*

Determination of the parameters in the gluon density

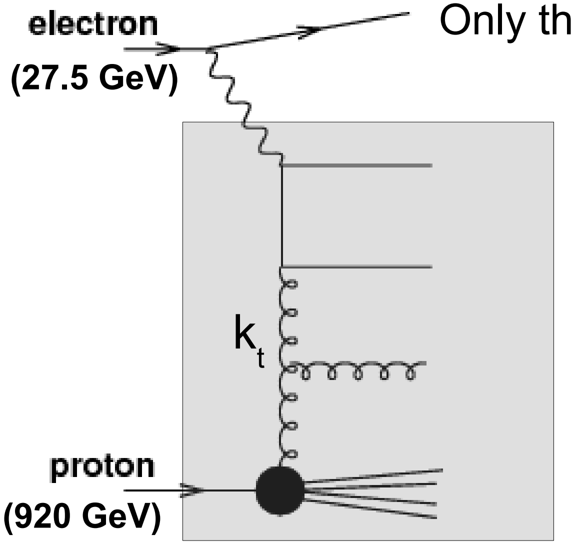
$$xA_0(x, k_T, \bar{q}_0) = N \cdot x^{-B} \cdot (1-x)^C \cdot (1-Dx) \cdot \exp\left(-\frac{(k_T-\mu)^2}{2\sigma^2}\right)$$

Used in the CASCADE MC generator:
Evolve uPDF according to the CCFM equation. Only gluons.

•First goal determine the x-dependence.

•Use the proton structure function (sigma reduced for positrons).
High precision combined measurement from H1 and ZEUS. (JHEP 1001:109 (2010))

Should be fairly insensitive to the kt-dependent part of the gluon. Inclusive measurement with minimum restrictions on the hadronic final state.



Only the electron detected

“Dont care” about the rest

Fitting the x dependence to the proton structure - F2

The previous fits...

$$xA_0(x, k_T, \bar{q}_0) = N \cdot x^{-B} \cdot (1-x)^C \cdot \exp\left(- \frac{(k_T - \mu)^2}{2\sigma^2} \right)$$

Previous fits to F2

The previous fits...

$$xA_0(x, k_T, \bar{q}_0) = N \cdot x^{-B} \cdot (1-x)^C \cdot \exp\left(-\frac{(k_T-\mu)^2}{2\sigma^2}\right)$$

...was good enough for the “low” precision F2-data:

Fitting F2 in the range x < 0.005, 5 < Q^2 < 100 GeV^2, to the “old” structure function measured by H1 (Eur.Phys.J.C21:33-61,2001)

<u>Minimum</u>
N = 0.81 ± 0.02
B = 0.029 ± 0.004
C = 4 (fixed)
σ = 1 (fixed)
μ = 0 (fixed)
χ^2/ndf=1.2

Bacchetta, Jung, Knutsson, Kutak, Himmelstjerna, arXiv:1001-4675 DESY 10-013

This is a **good fit** which **reconstructs** the parameter values in a **former official PDF** *(Jung, Kotikov, Lipatov, Zotov, hep-ph/0611093)* fitted to the same data with the previous fitting approach.
Good validation of the new fitting approach.

New fits

$$xA_0(x, k_T, \bar{q}_0) = N \cdot x^{-B} \cdot (1-x)^C \cdot \exp\left(-\frac{(k_T-\mu)^2}{2\sigma^2}\right)$$

- Fit to new ***high precision*** **combined F2 data from H1 and ZEUS (JHEP 1001:109 (2010))**

- **$x < 0.005$, $5.0 < Q^2 < 100.0$ GeV^2**

Minimum (old fit)
N = 0.81 ± 0.02
B = 0.029 ± 0.004
C = 4 (fixed)
σ = 1 (fixed)
μ = 0 (fixed)
χ^2/ndf=1.2

Bacchetta, Jung, Knutsson, Kutak, Himmelstjerna, arXiv:1001-4675 DESY 10-013

Minimum (fit to new data)
N = 0.83 ± 0.02
B = 0.017 ± 0.003
C = 4 (fixed)
σ = 1 (fixed)
μ = 0 (fixed)
χ^2/ndf=5.1

Roughly the **same minimum**, but a significantly higher χ^2.

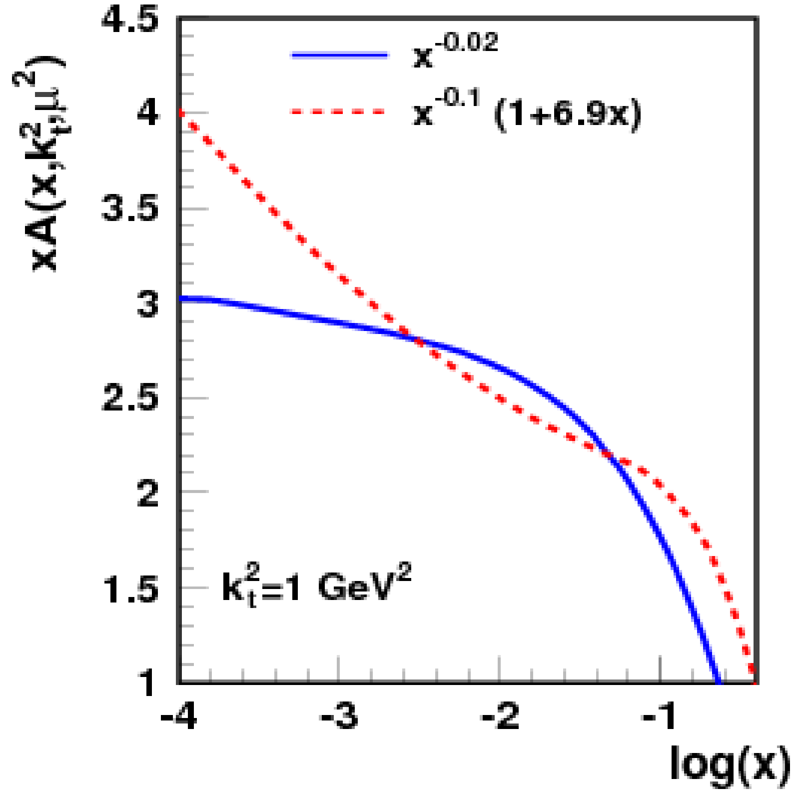
High precision data requires more from the model.

New parameterization

$$xA_0(x, k_T, \bar{q}_0) = N \cdot x^{-B} \cdot (1 - Dx) \cdot (1 - x)^C \cdot \exp\left(- \frac{(k_T - \mu)^2}{2\sigma^2} \right)$$

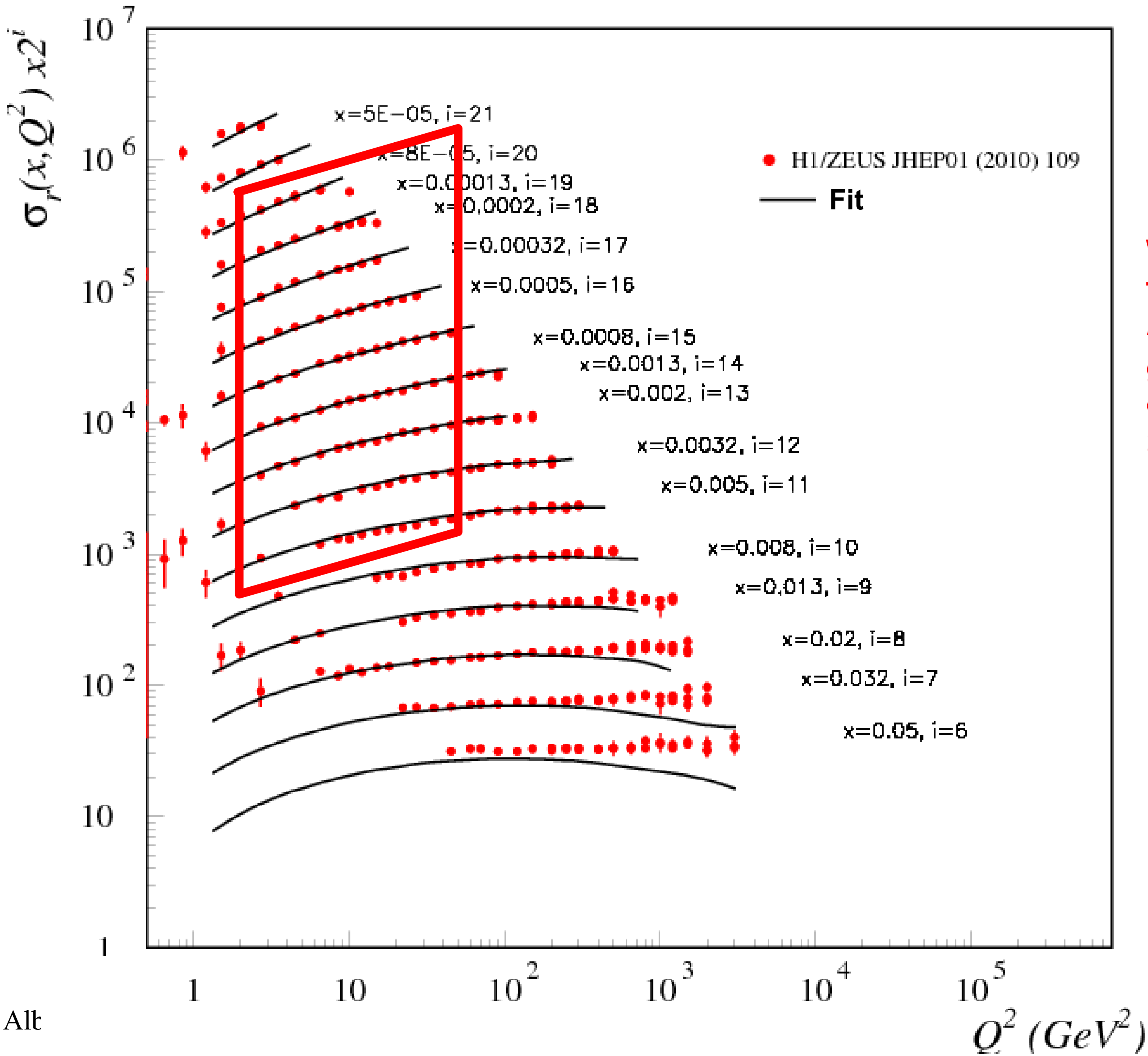
- *(1-Dx)* gives additional freedom to the gluon
- **$0.0001 < x < 0.005$, $2.0 \leq Q^2 < 50$ GeV^2**
- Significant improvement of the fit

Minimum
N = 0.47 ± 0.03
B = 0.11 ± 0.01
D = -6.9 ± 0.9
C = 4 (fixed)
σ = 1 (fixed)
μ = 0 (fixed)
χ^2/ndf=186.8/85=2.2



The new gluon is more pronounced at low and high x.

Results



= *included in fit*

χ^2/ndf=2.2

With only *3 parameters* fitted, the *gluon* in the *k_t-factorization* scheme gives a decent description of the data *within the fitted range.*

...but room for improvements.

Parameterization of PDF. (More freedom in fit...)

Need contribution from quarks?

Sensitivity to other MC parameters:
E.g light quark mass.

Tuning of Hadronization parameters to HERA data

Motivation:

- **Test of factorization ansatz of hadronisation**
- **Does a tune to HERA data give the same result as the PROFESSOR tunes to LEP data?**
- ***Test PROFFIT***

The data

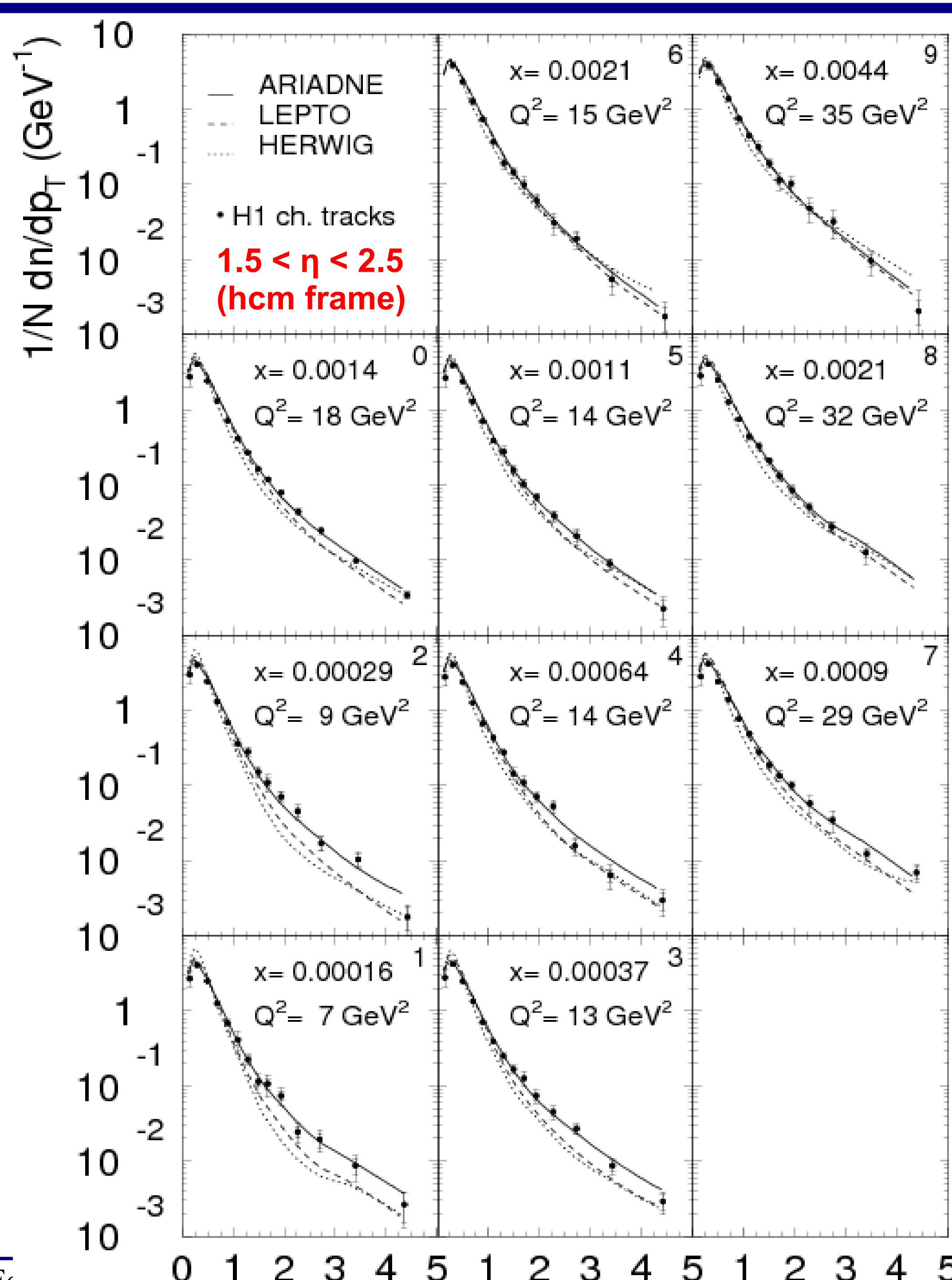
Transverse momentum spectra of charged particles in deep inelastic scattering *ep*-collisions at HERA.

- Figure from publication: Average charge particle multiplicity as a function of the the transverse momenta of the particles.

- Non-DGLAP based model is expected to produce more hard particles.

- ARIADNE with the Color Dipole Model (CDM), descibes the data better at high p_T

→ $p_T > 1.25$ GeV region not included in the tuning.



H1 Collaboration, Nucl.Phys.B485:3-24,1997, hep-ex/9610006

The data

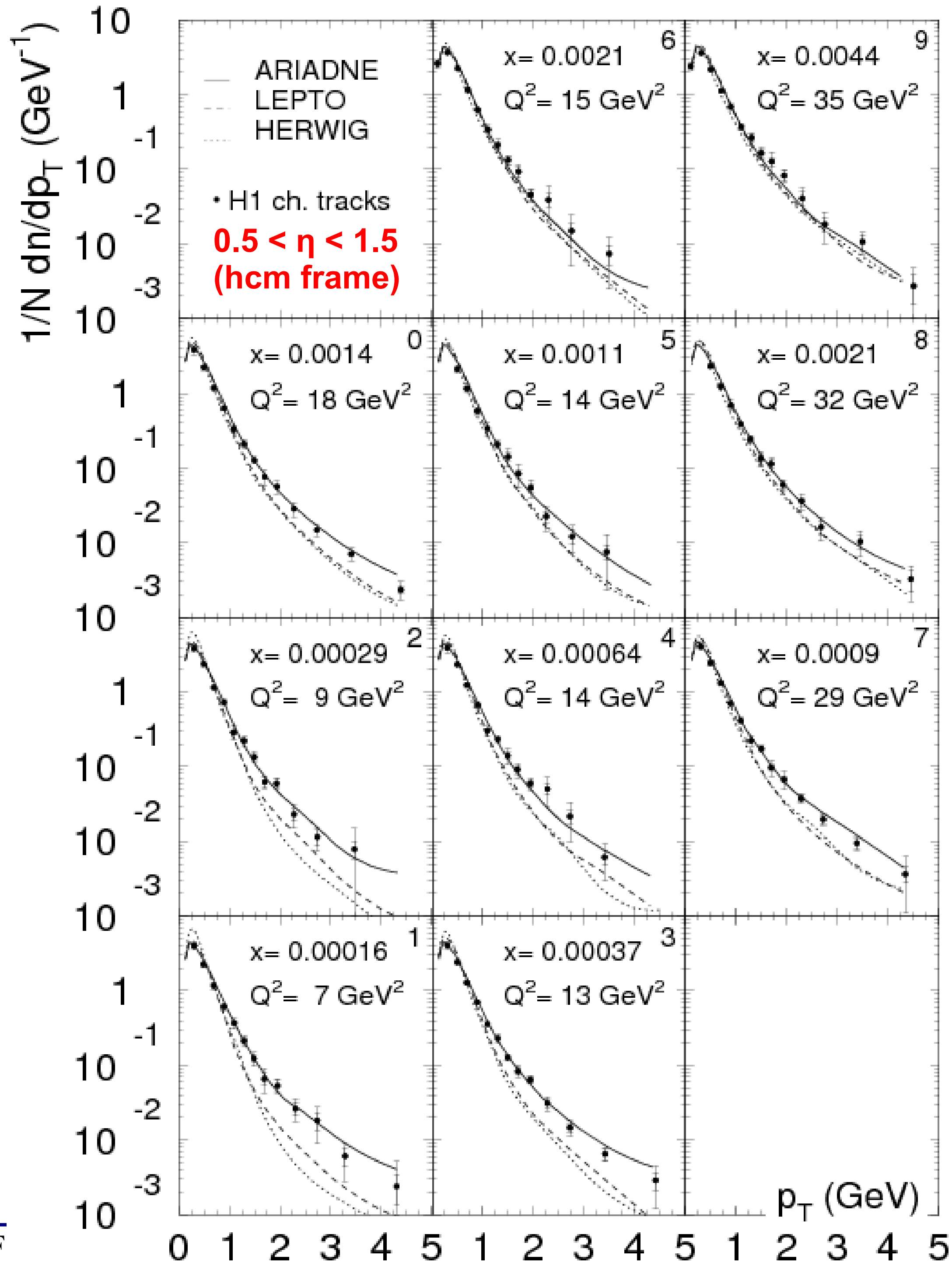
Transverse momentum spectra of charged particles in deep in-elastic scattering *ep*-collisions at HERA.

- Figure from publication: Average charge particle multiplicity as a function of the the transverse momenta of the particles.

- Non-DGLAP based model is expected to produce more hard particles.

- ARIADNE with the Color Dipole Model (CDM), descibes the data better at high p_T

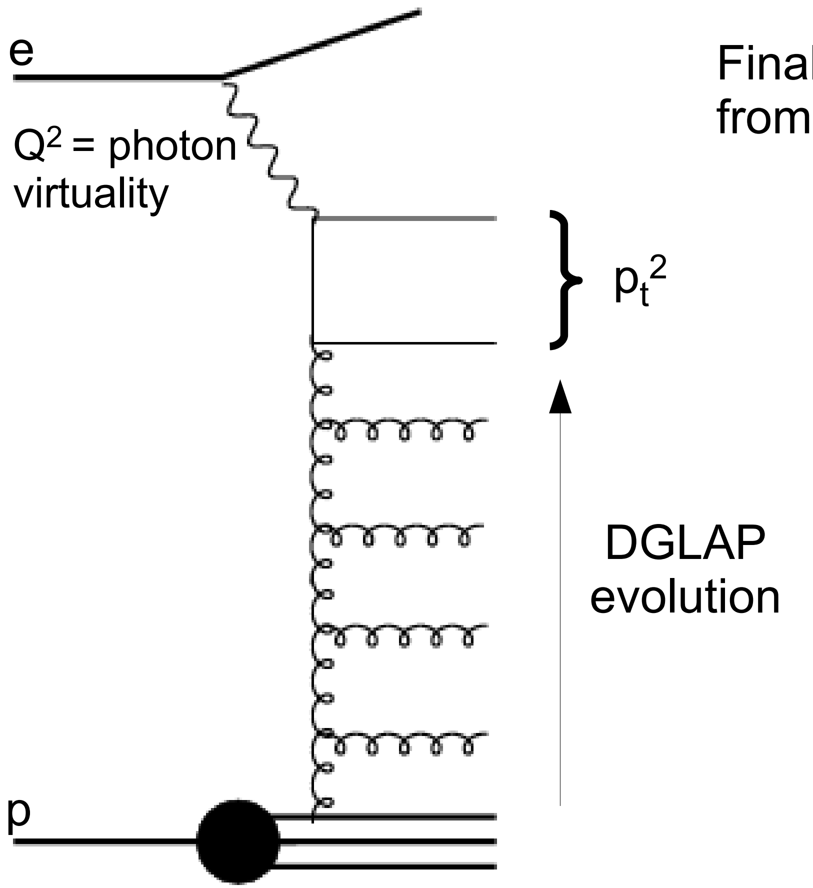
→ **p_T > 1.25 GeV region not included in the tune. Also forward (lab) region excluded.**



H1 Collaboration, Nucl.Phys.B485:3-24,1997, hep-ex/9610006

RAPGAP

Monte Carlo event generator for *ep*-scattering with LO ME and DGLAP intial and final state parton showers.
(H. Jung, Comput.Phys.Commun.86:147-161,1995)



Final state parton showers and hadronization from PYTHIA.

Settings for the hadronization tune:

PDF: CTEQ6.1L
Scales: $\mu_F = \mu_R = Q^2 + p_t^2$

Default parameters, but
flavour parameters:
The PROFESSOR tune to LEP data

The Tune

PARJ(21) σ_q – width of Gaussian for px and py of primary hadrons

PARJ(41) a

PARJ(42) b $\}$ parameters in the Lund fragmentation function

PARJ(47) r_b – interpolation between Bowler and Lund fragmentation. (1=pure Bowler shape)

PARJ(81) Λ_{QCD} for α_s in parton showers

PARJ(82) Invariant mass cut-off for PS. Partons below this value do not radiate.

Parameter	Default	Professor Tune	HERA Tune
PARJ(21)	0.36	0.325	0.43 ± 0.01
PARJ(41)	0.3	0.5	1.07 ± 0.18
PARJ(42)	0.58	0.6	0.77 ± 0.17
PARJ(47)	1.0	0.67	0.45 (no sensitivity)
PARJ(81)	0.29	0.29	0.2 ± 0.02
PARJ(82)	1.0	1.65	2.97 ± 0.96
χ^2/ndf *	245/94=**2.59**	417/94 = **4.4**	69.7/94=**0.74**

* χ^2 values are calculated for the HERA data by running the generator with the different parameter sets (errors of parameters are not considered)

The Tune

Parameter	Default	Professor Tune	HERA Tune
PARJ(21)	0.36	0.325	0.43 ± 0.01
PARJ(41)	0.3	0.5	1.07 ± 0.18
PARJ(42)	0.58	0.6	0.77 ± 0.17
PARJ(47)	1.0	0.67	0.45 (no sensitivity)
PARJ(81)	0.29	0.29	0.2 ± 0.02
PARJ(82)	1.0	1.65	2.97 ± 0.96
χ^2/ndf (p_t<1.25 GeV)	245/94=2.59	417/94 = 4.4	69.7/94=0.74
χ^2/ndf (p_t<0.8 GeV)	65.6/67=0.98	102/67= 1.5	36.3/67=0.54

Note: **At lower pt all parameters sets work!**

Warning!!! Is there non-DGLAP physics at pt>0.8 which is “lost” in the hadronization tune?

How to disentangle hadronization effects and small x effects?
How can we identify the different contributions?

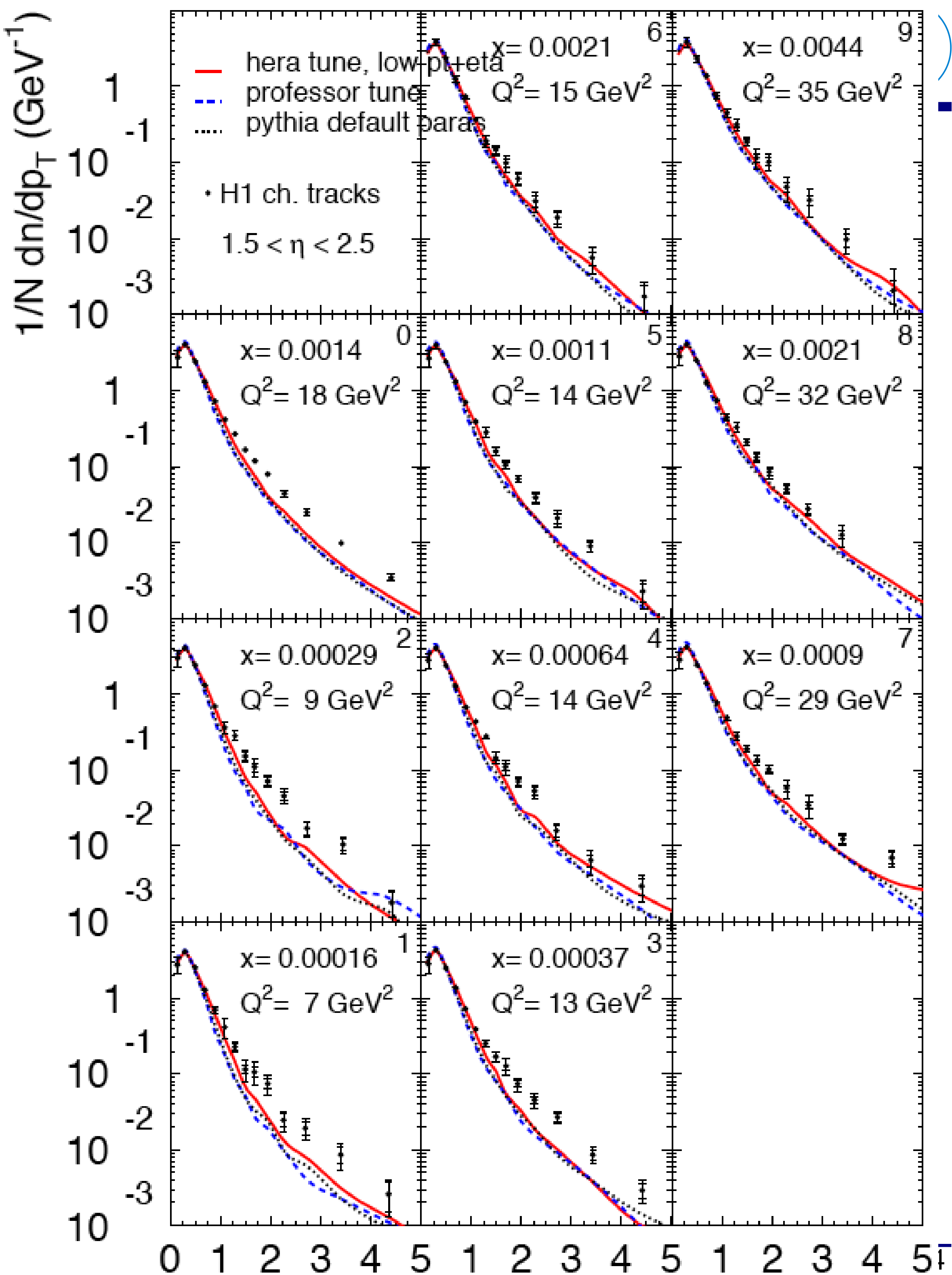
The question is also very relevant for LHC tunes: For example how do we distinguish different contributions to the UE, e.g. MI and parton showers

The Results

Pt -spectra.

Visually no big difference between the tunes.

Red line – HERA tune
Blue dashed – PROFESSOR tune
Black dotted – Default parameters



Summary

Summary

- **PROFFIT** – Fully functionally for multidimensional tunes.

- PROFFIT has the possibility to correctly treat correlated systematic errors. Used in the **fits of uPDFs to high precision HERA data**.

- Has been used for a **first hadronization tune to HERA data**.
  - Tune influenced by small x effects? The sensitivity to perturbative physics is a matter of investigation. Similar issues important for pp: UE/MI etc

albert.knutsson@cern.ch

Backups

The Tune

Tune the following hadronisation parameters to the HERA data:
Same parameters as tuned in the PROFESSOR tune to LEP data

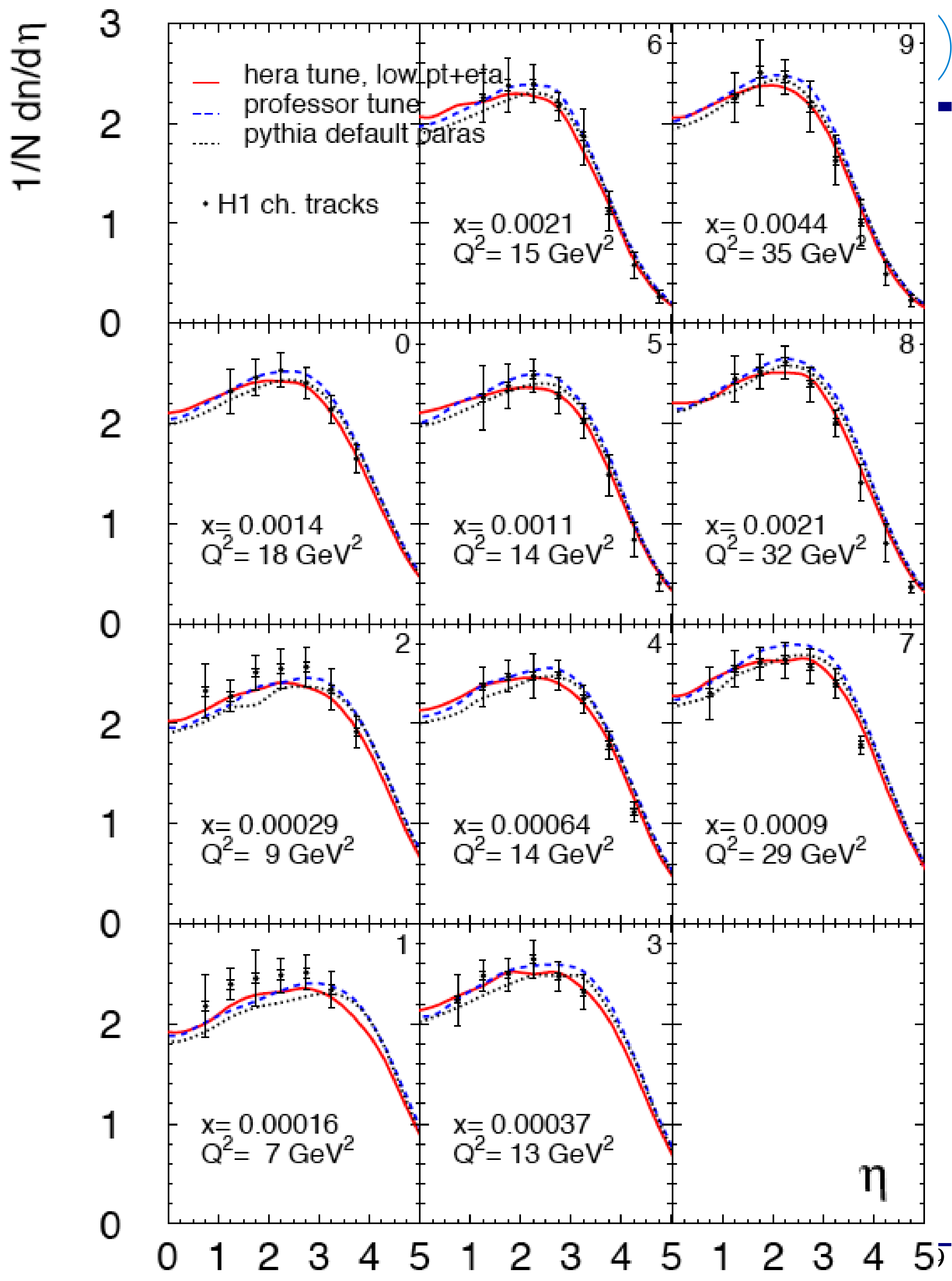
PARJ(21)	σ_q – width of Gaussian for px and py of primary hadrons
PARJ(41)	a } parameters in the Lund fragmentation function
PARJ(42)	b } parameters in the Lund fragmentation function
PARJ(47)	r_b – interpolation between Bowler and Lund fragmentation. (1=pure Bowler shape)
PARJ(81)	Λ_{QCD} for α_s in parton showers
PARJ(82)	Invariant mass cut-off for PS. Partons below this value do not radiate.

Only statistic and total systematic errors provided in the publication. Systematic error is used uncorrelated. Statistical error in MC considered.

The Results

Rapdity spectra. Central region included in the tuning.

Red line – HERA tune
Blue dashed – PROFESSOR tune
Black dotted – Default parameters



About PROFFIT

- **Singular Value Decomposition** used to determine the polynomial describing the MC grid. 2^{nd}, 3^{rd} or 4^{th} order polynomial can be used to described the MC grid.

- The fit of the MC parameters (in the polynomial) to the data is done by Minuit (**MIGRAD**)

- Equidistant MC grids has been used for the uPDF fits, but this is not possible for tunes with many parameter since the number of points needed for parameterization is at least 4^N for N parameters. → Use randomized MC grid.

Error treatment

- The statistical errors of the MC is propagated to the coefficients of the polynomial. A co-variance matrix for the coefficients are calculated.

- The CTEQ Chi2 calculation (hep/ph/0101051) is used to take the correlated errors in the data into consideration. Basically the χ^2 is differently calculated.

In the fit of the MC parameters to the data the uncorrelated errors and the different correlated errors can be treated separately according to:

$$\chi^2 = \Sigma \frac{(X_{Data} - X_{Polynomial})^2}{\alpha^2} - \Sigma_j \Sigma_{j'} B_j (A^{-1})_{jj'} B_{j'}$$

$\alpha^2 =$ Sum of uncorrelated errors (data and polynomial)

$\Sigma_j \Sigma_{j'} B_j (A^{-1})_{jj'} B_{j'} =$ Term related to the correlated systematic errors (vector *B*), and their correlations (matrix *A*)