An alternative approach to fitting the unintegrated gluon density

(Status report)

HERA-LHC Working Group Week, DESY, 29/10-2/11-2007

Albert Knutsson, Krzysztof Kutak (DESY)

(Thanks to Magnus Hansson (Lund) for help and data!)

Outline

- Intro to the unintegrated gluon
- The Data
- The fitting
- Results
- Summary/Outlook

The unintegrated PDF

The uPDF starting distribution:

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

***N*: Normalization (fitted)**

***B*: Small x behaviour (fitted)**

***C*=4: Large x behaviour (kept fixed)**

μ **=** σ **=1.5: Determines the shape of the intrinsic** k_T **of the gluon below** k_T **= 1.2 GeV (*for simplicity kept fixed at the moment*)**

$\bar{q}_0$ **is the starting scale of the distribution.**

The uPDF is calculated for higher scales by emissions of gluons according to the CCFM evolution scheme.

The parameters N,B,C, μ **,** σ **, are not theoretically calculable.**

→ **We need to fit the uPDF to experimental data.**

The Data – Dijets and azimuthal decorrelations

For testing and development we fit to a double differential dijet cross-section, which is expected to be sensitive to the gluon.

- **Kinematic range**

$$5\ \mathrm{GeV}^2 < Q^2 < 100\ \mathrm{GeV}^2$$

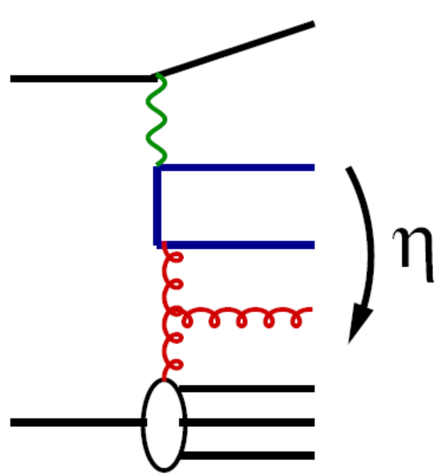
$$0.1 < y < 0.7$$

- **Jet selection**

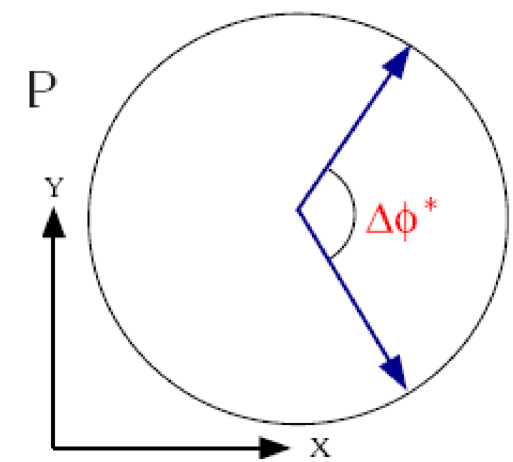
$$-1 < \eta_j < 2.5 \quad \mathrm{(LAB)}$$

$$5\ \mathrm{GeV} < E^*_{\perp j} \quad \mathrm{(HCM)}$$

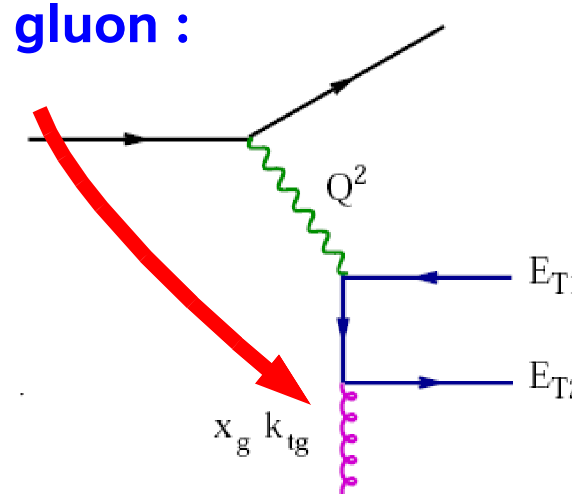
$N_{\mathrm{jets}} > 2 \Rightarrow$ **Pick the two jets closest in η to electron**



Measure $\Delta\phi^*$ between the dijets:



Sensitivity to the gluon :



The Data – Dijets and azimuthal decorrelations

Integrated PDF: DGLAP

LO: **Gluon collinear with proton**

$$k_{t,\text{gluon}} = 0$$

$$\Delta\Phi = 180^o$$ in HCM

Higher orders: $$k_{t,\text{gluon}} \neq 0$$

$$\Delta\Phi < 180^o$$

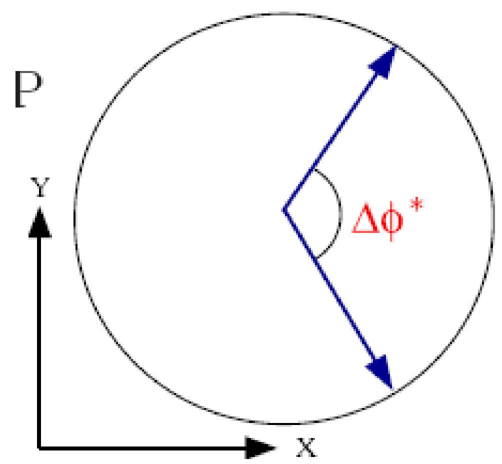
Unintegrated PDF: CCFM or BFKL

$$k_{t,\text{gluon}} \neq 0$$

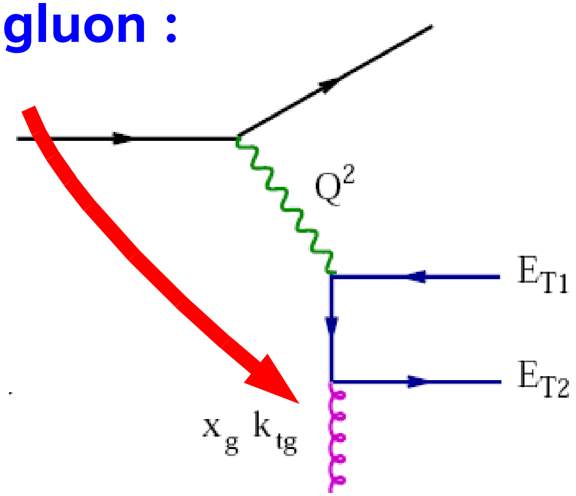
$$\Delta\Phi < 180^o$$

already at LO

Measure $\Delta\phi^*$ between the dijets:

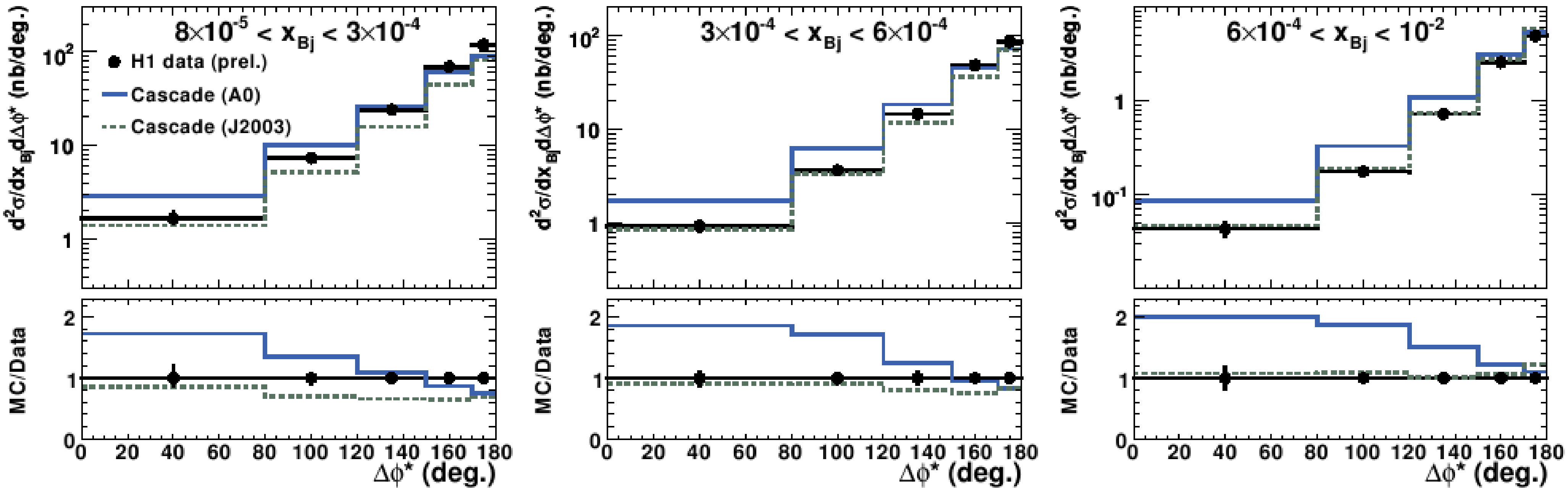


Sensitivity to the gluon :



The Data – Dijets and azimuthal decorrelations

Double differential cross-section (H1 preliminary):



The uPDF has been fitted before (to other, more inclusive, data).

Here compared to the same data as we use,

(*A0: Only non-signular terms in the CCFM gluon splitting function*
J2003: Both singular and non-singular terms in the splitting function)

Fitting (unintegrated) PDFs

The conventional fitting method:

1. Calculate cross-section using CASCADE for a given set of parameter values
2. Compare to data, calculate Chi2 and feed it to MINUIT
3. MINUIT (e.g. the simplex method) estimates new parameter values
4. Iterate 1. - 3. until Chi2 is minimized

This means that if MINUIT needs 100 iterations to minimize Chi2, CASCADE is run 100 times, not simultaneously:

If one CASCADE run takes 1 hour, the minimization takes 100hours.

To fit uPDF one needs exclusive measurements (like the azimuthal dijet measurement)

→

A lot of statistics. Minimization >> 100h.

Acknowledgement

New method!

The method was developed for tuning Monte Carlo models

Parameter Optimisation in Monte Carlo Event Generators

Hendrik Hoeth

(University of Wuppertal)

1st Mcnet School, IPPP Durham, 18-20th April 2007

We try to carry out the same method for fitting uPDFs.

Fitting unintegrated PDFs

1. **Build up a grid in parameter – cross section space using Monte Carlo.**

   If you have a CPU farm (or use the *GRID*) this ultimately takes the time of running CASCADE once.

Fitting unintegrated PDFs

1. **Build up a grid in parameter – cross section space using Monte Carlo.**

   If you have a CPU farm (or use the *GRID*) this ultimately takes the time of running CASCADE once.

2. **Fit polynomials to the Monte Carlo grid.**

$$\sigma_{\mathrm{poly}} = A + \sum_{1}^{N} B_i \cdot p_i + \sum_{1}^{N} C_i \cdot p_i^2 + \sum_{1}^{N} D_i \cdot p_i p_{i-1} + H.O.$$

   A, B, C **and** D **are determined by fitting the polynomial to the parameter-xsection grid.**

   This takes a *few seconds*.

Step 1. and 2. are done for each bin in the measurement.

Fitting unintegrated PDFs

1. Build up a grid in parameter – cross section space using Monte Carlo.

   If you have a CPU farm (or use the ***GRID***) this ultimately takes the time of running CASCADE once.

2. Fit polynomials to the Monte Carlo grid.

$$\sigma_{\text{poly}} = A + \sum_1^N B_i \cdot p_i + \sum_1^N C_i \cdot p_i^2 + \sum_1^N D_i \cdot p_i p_{i-1} + H.O.$$

   A, B, C and D are determined by fitting the polynomial to the parameter-xsection grid.

   This takes a *few seconds*.

Takes care of correlation between parameters

Step 1. and 2. are done for each bin in the measurement.

Fitting unintegrated PDFs

1. Build up a grid in parameter – cross section space using Monte Carlo.

   If you have a CPU farm (or use the ***GRID***) this ultimately takes the time of running CASCADE once.

2. Fit polynomials to the Monte Carlo grid.

$$\sigma_{\text{poly}} = A + \sum_1^N B_i \cdot p_i + \sum_1^N C_i \cdot p_i^2 + \sum_1^N D_i \cdot p_i p_{i-1} + H.O.$$

Takes care of correlation between parameters

A, B, C and D are determined by fitting the polynomial to the parameter-xsection grid.

This takes a *few seconds*.

Step 1. and 2. are done for each bin in the measurement.

3. Determine PDF parameters, p_i , by fitting the polynomials to data

   Also this takes only a few seconds.

Step 2. and 3. are done by Chi2-minimization in for example minuit.

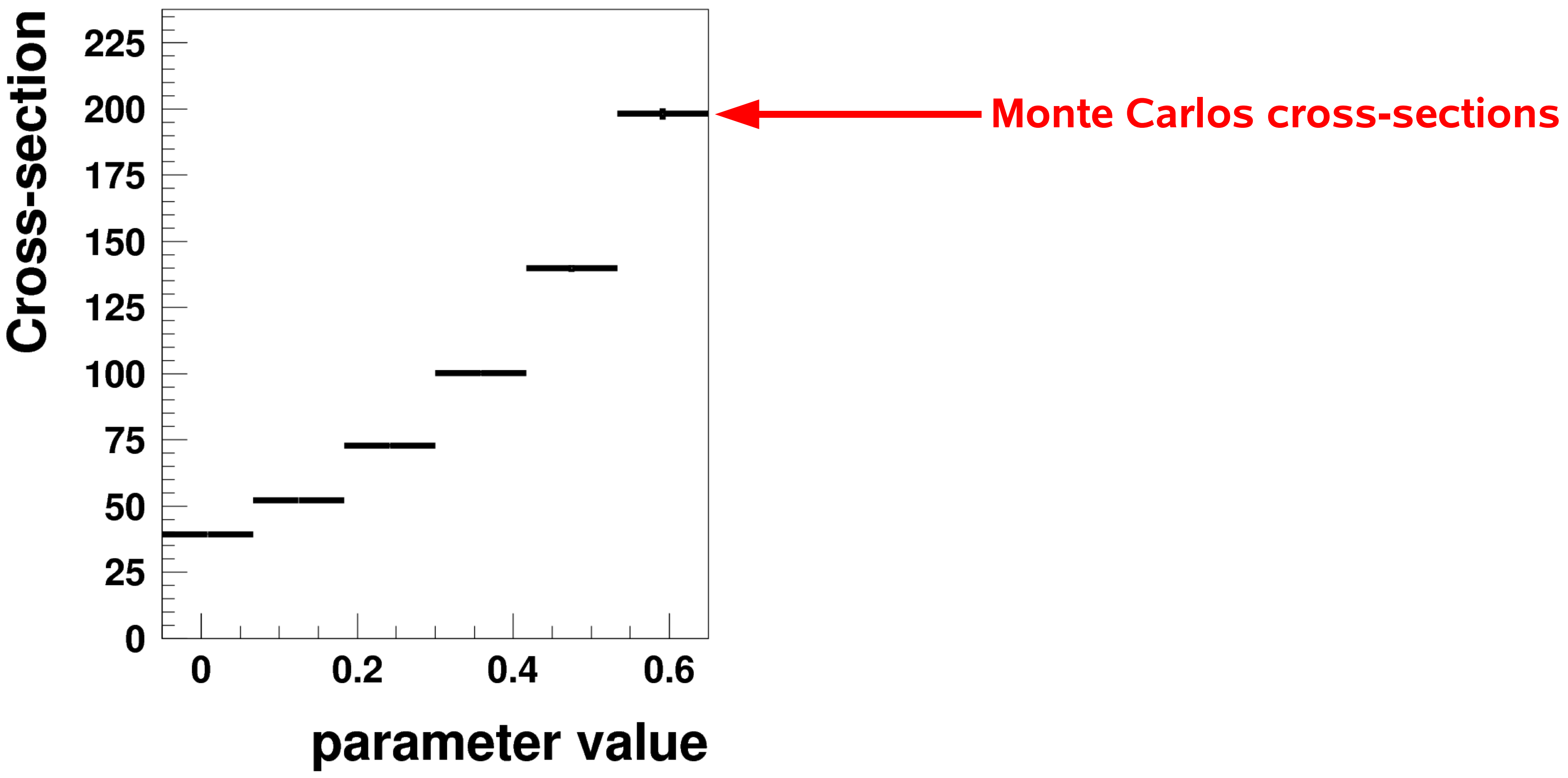
Simple Example

Simplest possible example

1 parameter, 1 data cross-section

(based on a sub-sample of the actual grid.)

1. Build up the grid



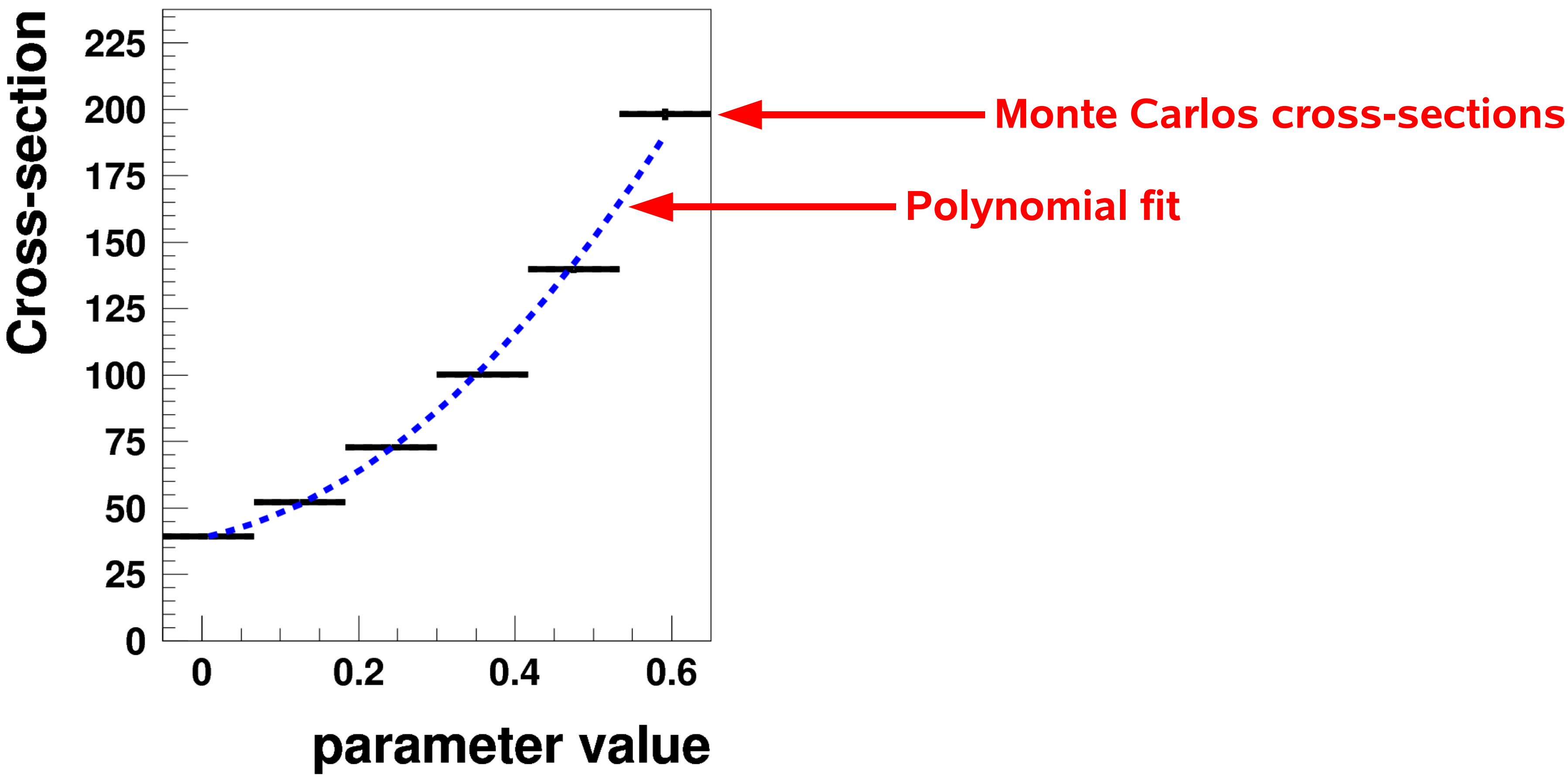
Simple Example

Simplest possible example

1 parameter, 1 data cross-section

(based on a sub-sample of the actual grid.)

2. Fit Ploynomial



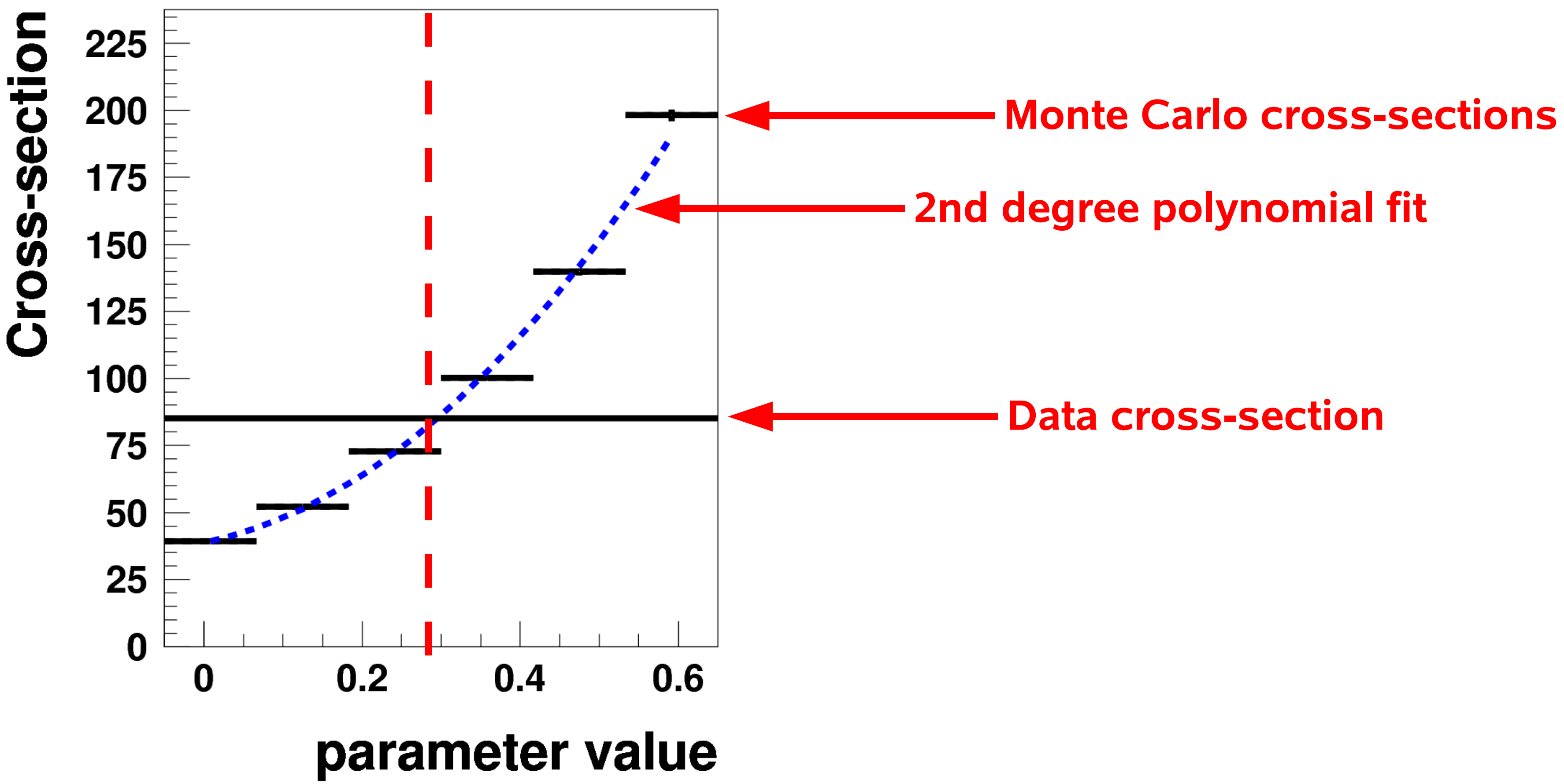
Simple Example

Simplest possible example

1 parameter, 1 data cross-section

(based on a sub-sample of the actual grid.)

3. Minimize Chi2 to data



Fitting unintegrated PDFs

The actual fit: **2 parameters, 5x3 data cross-section**

1. Build up the grid

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

N=[0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7]
B=[0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7]

Fitting unintegrated PDFs

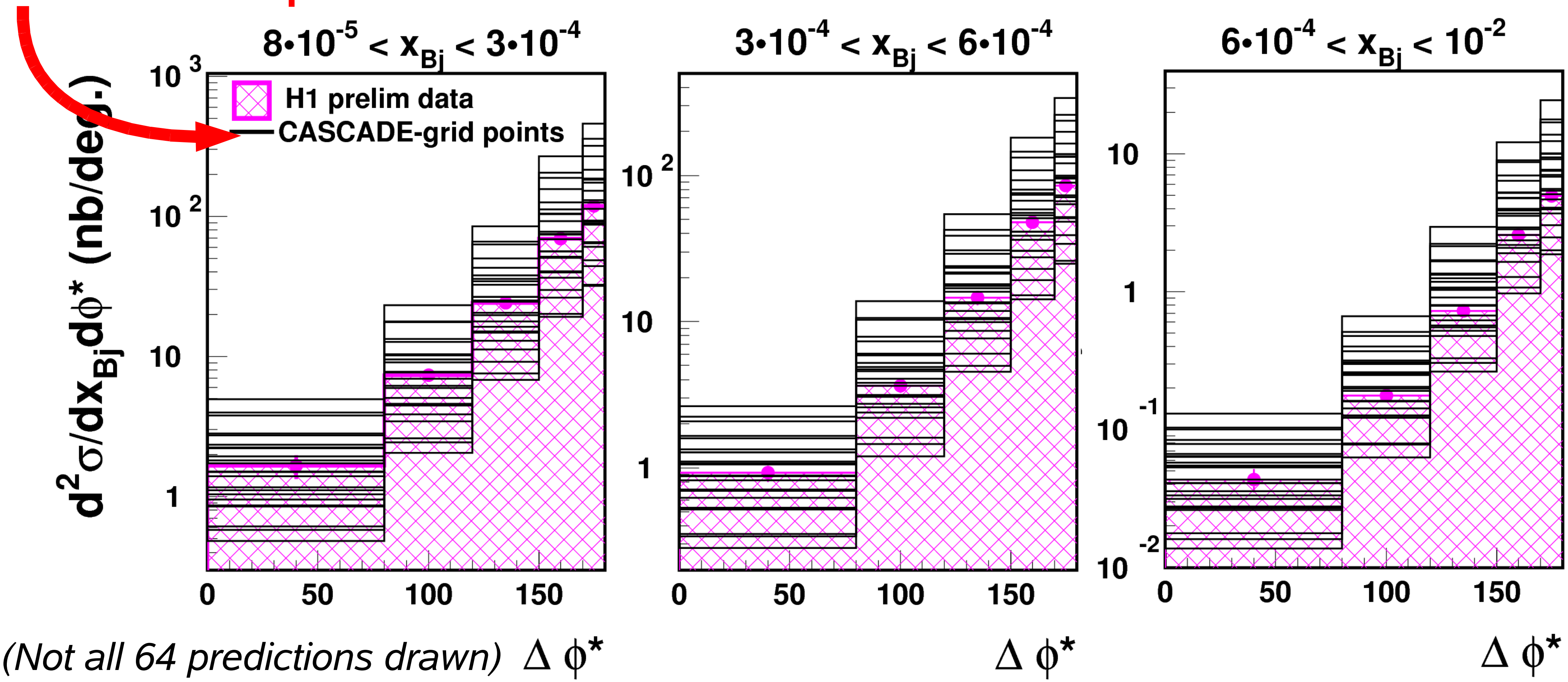
The actual fit: 2 parameters, 5x3 data cross-section

1. Build up the grid

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

Currently 8x8 grid → 64 CASCADE predictions

N=[0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7]
B=[0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7]



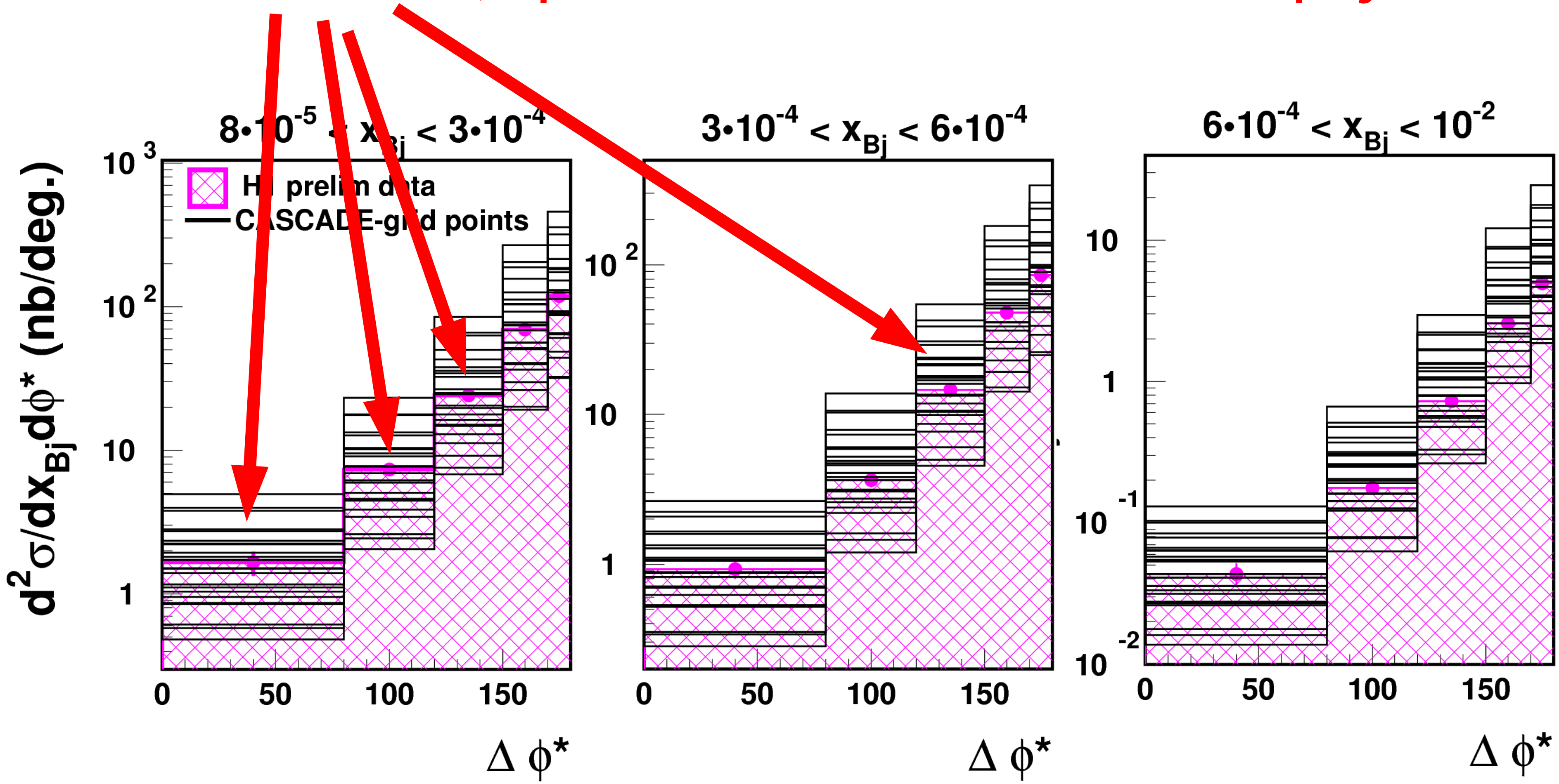
(Not all 64 predictions drawn)

Fitting unintegrated PDFs

The actual fit: **2 parameters, 5x3 data cross-section**

2. Determine polynomials

15 cross-section, 2 parameters ⟶ 15 two dimensional polynomials



Fitting unintegrated PDFs

3. Fit polynomials to Monte Carlo grids

We try different orders of the fitted polynomials

Chi2/NDF for polynomial fits

Order:	2nd	3rd	4th	5th	6th
chi2/ndf for histo 1, bin 1 =	11.8	2.1	1.9	1.9	1.7
chi2/ndf for histo 1, bin 2 =	34.7	3.4	2.0	2.3	2.0
chi2/ndf for histo 1, bin 3 =	129.	10.	3.4	3.8	3.1
chi2/ndf for histo 1, bin 4 =	356.	30.	4.0	4.0	3.4
chi2/ndf for histo 1, bin 5 =	345.	30.	3.4	3.1	2.8
chi2/ndf for histo 2, bin 1 =	8.21	1.9	1.7	1.6	1.5
chi2/ndf for histo 2, bin 2 =	24.1	3.0	1.8	1.7	1.9
chi2/ndf for histo 2, bin 3 =	95.6	7.8	3.1	2.7	2.5
chi2/ndf for histo 2, bin 4 =	293.	23.	4.0	2.7	2.1
chi2/ndf for histo 2, bin 5 =	283.	26.	3.3	2.4	2.3
chi2/ndf for histo 3, bin 1 =	9.33	2.1	2.2	2.1	2.3
chi2/ndf for histo 3, bin 2 =	30.1	5.6	5.7	5.9	6.0
chi2/ndf for histo 3, bin 3 =	130.	16.	14.	14.	14.
chi2/ndf for histo 3, bin 4 =	467.	41.	17.	15.	15.
chi2/ndf for histo 3, bin 5 =	477.	41.	13.	12.	12.

$$\chi^2 = \sum \frac{(\sigma_{i,\mathrm{poly}} - \sigma_{i,\mathrm{MC}})^2}{err^2_{i,\mathrm{MC}}}$$

Improvement up to 4th degree.

4th, 5th and 6th describes MC parameter-xsection space equally well.

Fitting unintegrated PDFs

3. Fit Polynomials to data by varying parameters

Chi2/NDF for polynomial fits

Order:	2nd	3rd	4th	5th	6th
chi2/ndf for histo 1, bin 1 =	11.8	2.1	1.9	1.9	1.7
chi2/ndf for histo 1, bin 2 =	34.7	3.4	2.0	2.3	2.0
chi2/ndf for histo 1, bin 3 =	129.	10.	3.4	3.8	3.1
chi2/ndf for histo 1, bin 4 =	356.	30.	4.0	4.0	3.4
chi2/ndf for histo 1, bin 5 =	345.	30.	3.4	3.1	2.8
chi2/ndf for histo 2, bin 1 =	8.21	1.9	1.7	1.6	1.5
chi2/ndf for histo 2, bin 2 =	24.1	3.0	1.8	1.7	1.9
chi2/ndf for histo 2, bin 3 =	95.6	7.8	3.1	2.7	2.5
chi2/ndf for histo 2, bin 4 =	293.	23.	4.0	2.7	2.1
chi2/ndf for histo 2, bin 5 =	283.	26.	3.3	2.4	2.3
=	9.33	2.1	2.2	2.1	2.3
=	30.1	5.6	5.7	5.9	6.0
=	130.	16.	14.	14.	14.
=	467.	41.	17.	15.	15.
=	477.	41.	13.	12.	12.

$$\chi^2 = \sum \frac{(\sigma_{i,\mathrm{Data}} - \sigma_{i,\mathrm{Poly}})^2}{err^2_{i,\mathrm{Data}}}$$

Order:	2nd	3rd	4th	5th	6th
Chi2/NDF for parameter fit	3.79	3.84	3.67	3.79	3.68
N-value	0.119	0.110	0.115	0.109	0.110
B-value	0.537	0.530	0.520	0.535	0.534

We pick point with lowest Chi2:

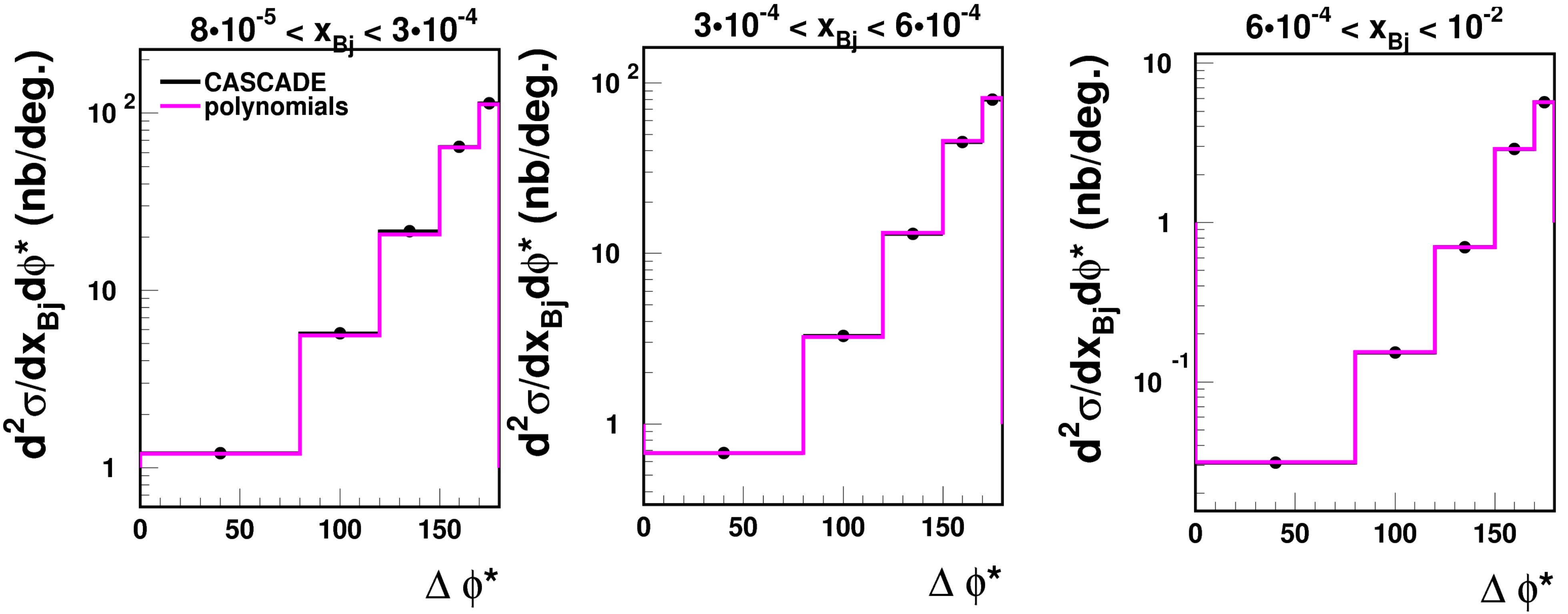
N=0.115

B=0.520

$$xA_0 = N \cdot x^{-B} \cdot (1-x)^C \cdot exp(-\frac{(k_T - \mu)^2}{2\sigma^2})$$

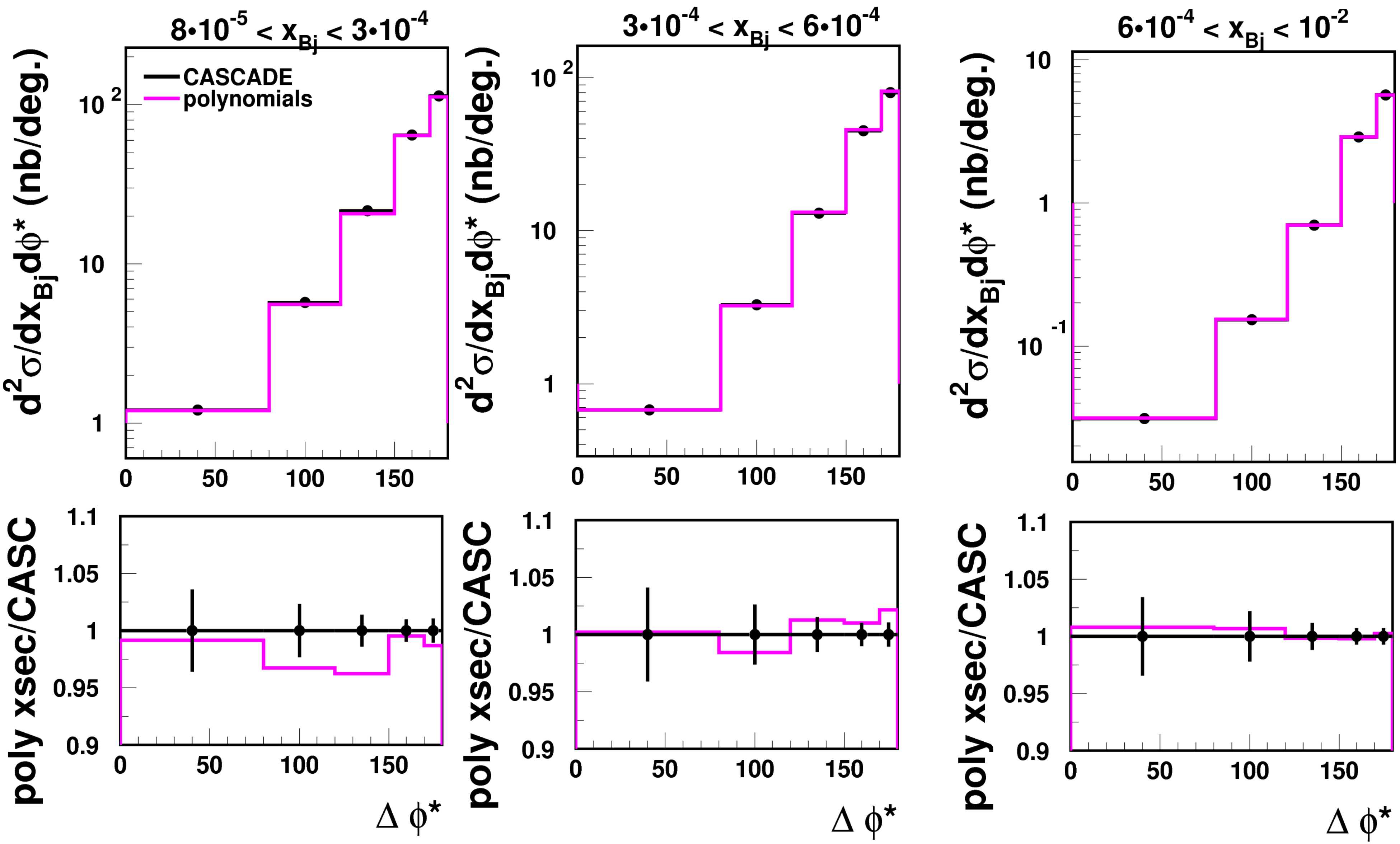
Fitting unintegrated PDFs

Polynomial cross-sections compared to the actual CASCADE prediction



Fitting unintegrated PDFs

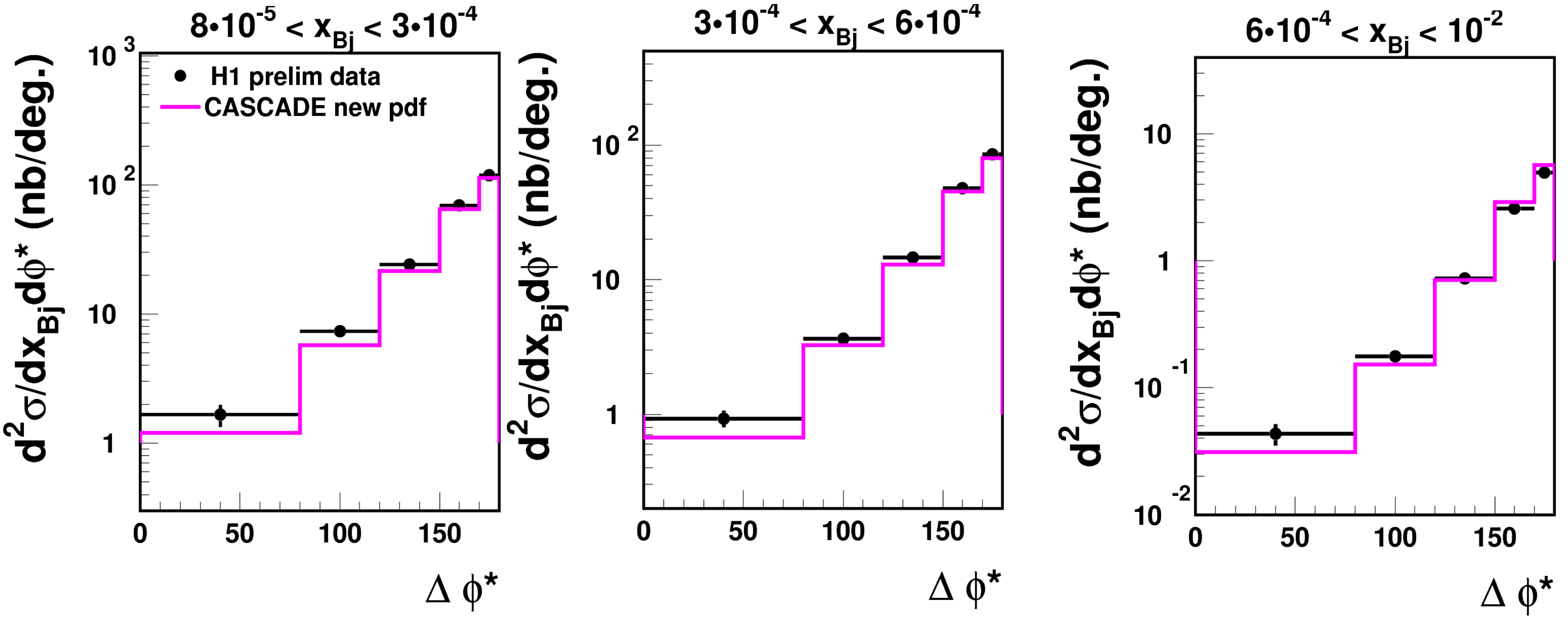
Polynomial cross-sections compared to the actual CASCADE prediction



Room for improvements, but not very bad.

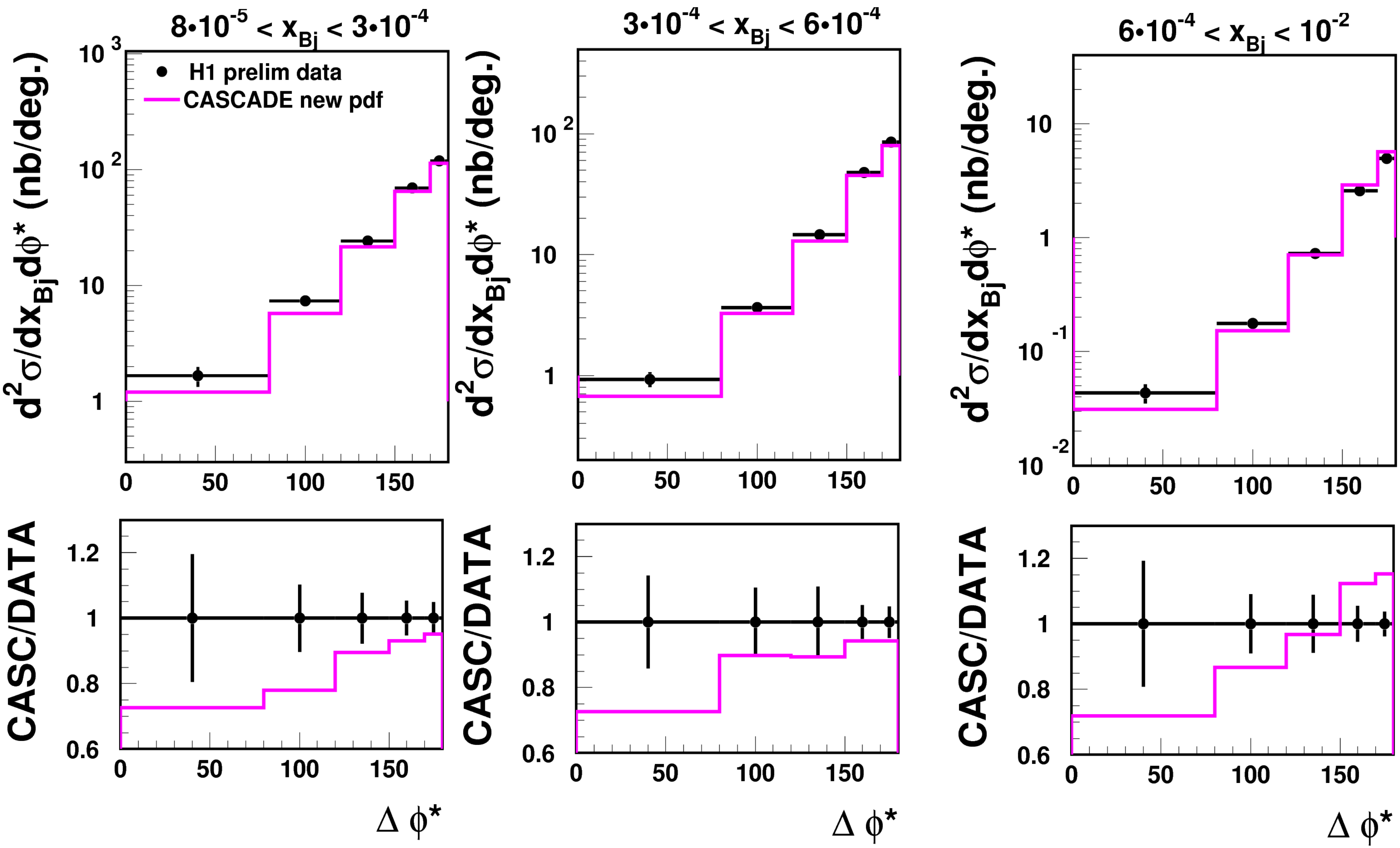
Fitting unintegrated PDFs

The results of the fit compared to data



Fitting unintegrated PDFs

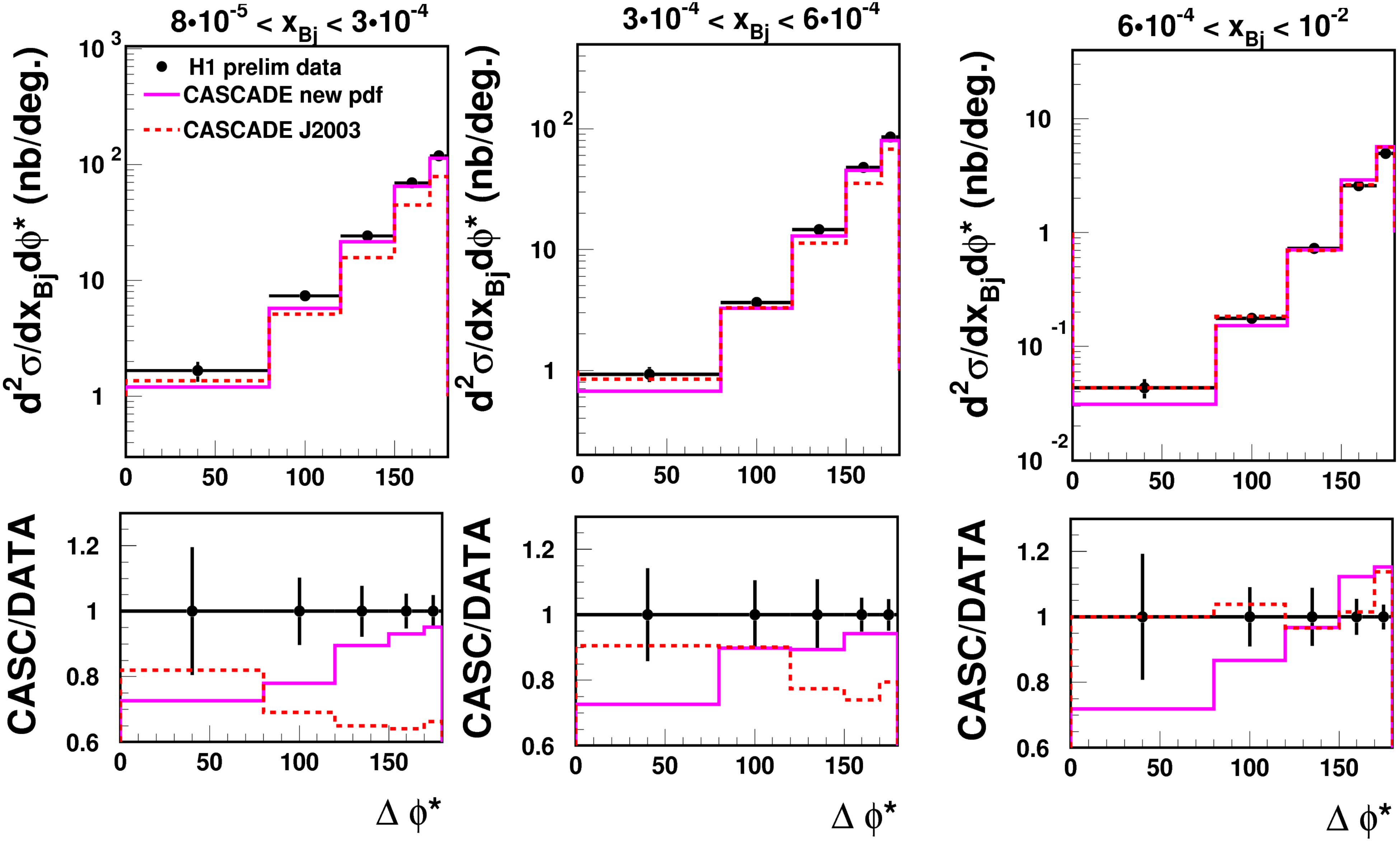
The results of the fit compared to data



Up to 30% deviation

Fitting unintegrated PDFs

The results of the fit compared to data and J2003-pdf



At low *x* and medium *x* in competition with J2003...

That one was fitted to other data.

Bad: We dont reconstruct the data we are fitting to.

Fitting the unintegrated uPDF

Chi2/NDF for polynomial fit

Degree:	2nd	3rd	4th	5th	6th
chi2/ndf for histo 1, bin 1 =	11.8	2.1	1.9	1.9	1.7
chi2/ndf for histo 1, bin 2 =	34.7	3.4	2.0	2.3	2.0
chi2/ndf for histo 1, bin 3 =	129.	10.	3.4	3.8	3.1
chi2/ndf for histo 1, bin 4 =	356.	30.	4.0	4.0	3.4
chi2/ndf for histo 1, bin 5 =	345.	30.	3.4	3.1	2.8
chi2/ndf for histo 2, bin 1 =	8.21	1.9	1.7	1.6	1.5
chi2/ndf for histo 2, bin 2 =	24.1	3.0	1.8	1.7	1.9
chi2/ndf for histo 2, bin 3 =	95.6	7.8	3.1	2.7	2.5
chi2/ndf for histo 2, bin 4 =	293.	23.	4.0	2.7	2.1
chi2/ndf for histo 2, bin 5 =	283.	26.	3.3	2.4	2.3
chi2/ndf for histo 3, bin 1 =	9.33	2.1	2.2	2.1	2.3
chi2/ndf for histo 3, bin 2 =	30.1	5.6	5.7	[illegible]	6.0
chi2/ndf for histo 3, bin 3 =	130.	16.	[illegible]	14.	14.
chi2/ndf for histo 3, bin 4 =	467.	[illegible]	17.	15.	15.
chi2/ndf for histo 3, bin 5 =	[illegible]	41.	13.	12.	12.

What happens if we excludes some bins?

We remove the ones where the polynomial fit is worst.

Fitting the unintegrated uPDF

Chi2/NDF for polynomial fit

	Degree: 2nd	3rd	4th	5th	6th
chi2/ndf for histo 1, bin 1 =	11.8	2.1	1.9	1.9	1.7
chi2/ndf for histo 1, bin 2 =	34.7	3.4	2.0	2.3	2.0
chi2/ndf for histo 1, bin 3 =	129.	10.	3.4	3.8	3.1
chi2/ndf for histo 1, bin 4 =	356.	30.	4.0	4.0	3.4
chi2/ndf for histo 1, bin 5 =	345.	30.	3.4	3.1	2.8
chi2/ndf for histo 2, bin 1 =	8.21	1.9	1.7	1.6	1.5
chi2/ndf for histo 2, bin 2 =	24.1	3.0	1.8	1.7	1.9
chi2/ndf for histo 2, bin 3 =	95.6	7.8	3.1	2.7	2.5
chi2/ndf for histo 2, bin 4 =	293.	23.	4.0	2.7	2.1
chi2/ndf for histo 2, bin 5 =	283.	26.	3.3	2.4	2.3
chi2/ndf for histo 3, bin 1 =	9.33	2.1	2.2	2.1	2.3
chi2/ndf for histo 3, bin 2 =	30.1	5.6	5.7	[illegible]	6.0
chi2/ndf for histo 3, bin 3 =	130.	16.	[illegible]	14.	14.
chi2/ndf for histo 3, bin 4 =	467.	41.	17.	15.	15.
chi2/ndf for histo 3, bin 5 =	177.	41.	13.	12.	12.

What happens if we exclude some polynomial?

We remove the ones where the polynomial fit is worst.

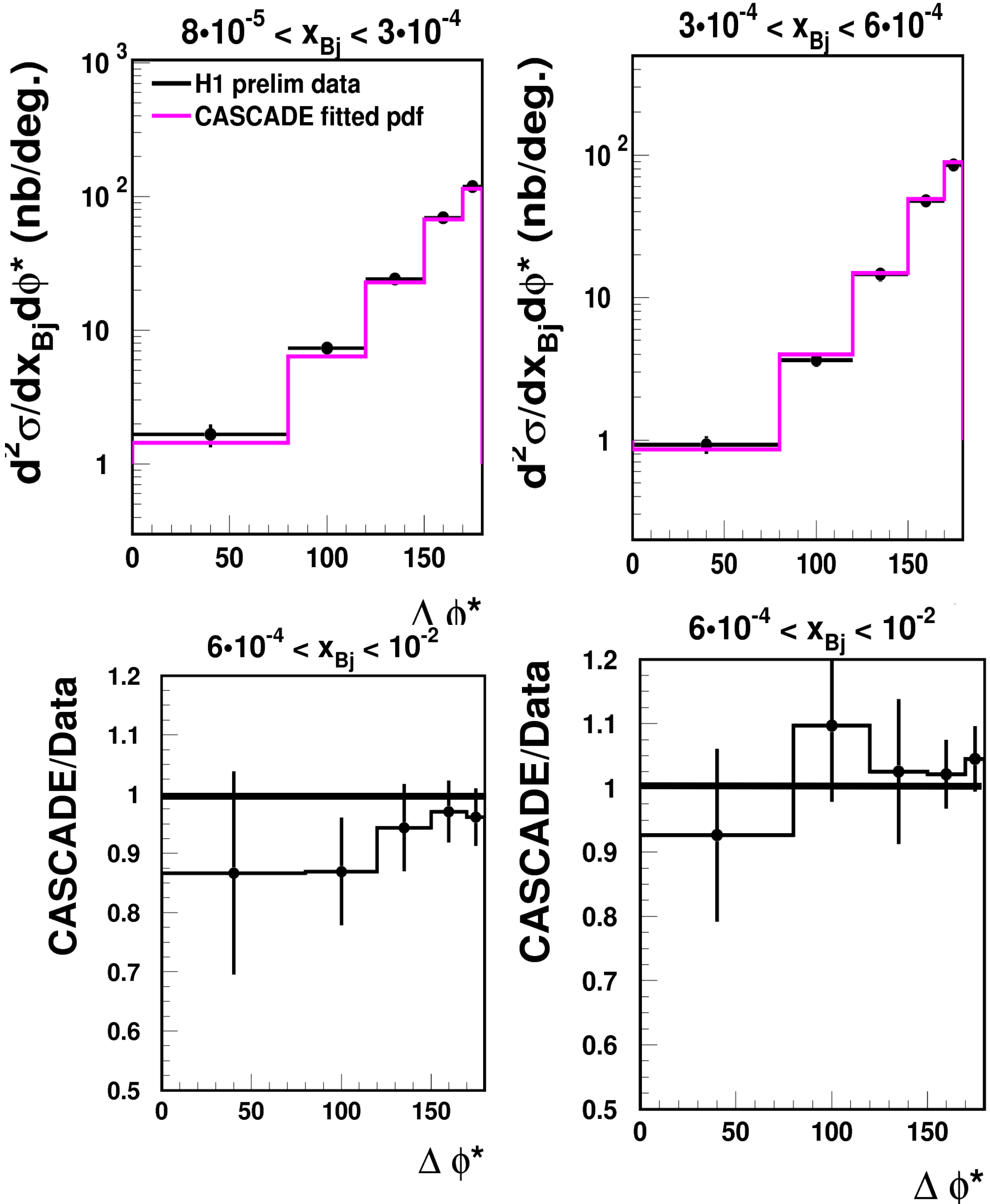
Degree:	2nd	3rd	4th	5th	6th
Chi2/NDF for parameter fit	0.45	0.47	0.52	0.52	0.49
N-value	0.23	0.24	0.22	0.22	0.21
B-value	0.32	0.31	0.35	0.36	0.36

(These parameters are close to Magnus Hanssons'. Conventianl fitting method.)

$$xA_0 = N \cdot x^{-B} \cdot (1-x)^C \cdot exp(-\frac{(k_T-\mu)^2}{2\sigma^2})$$

Fitting unintegrated PDFs

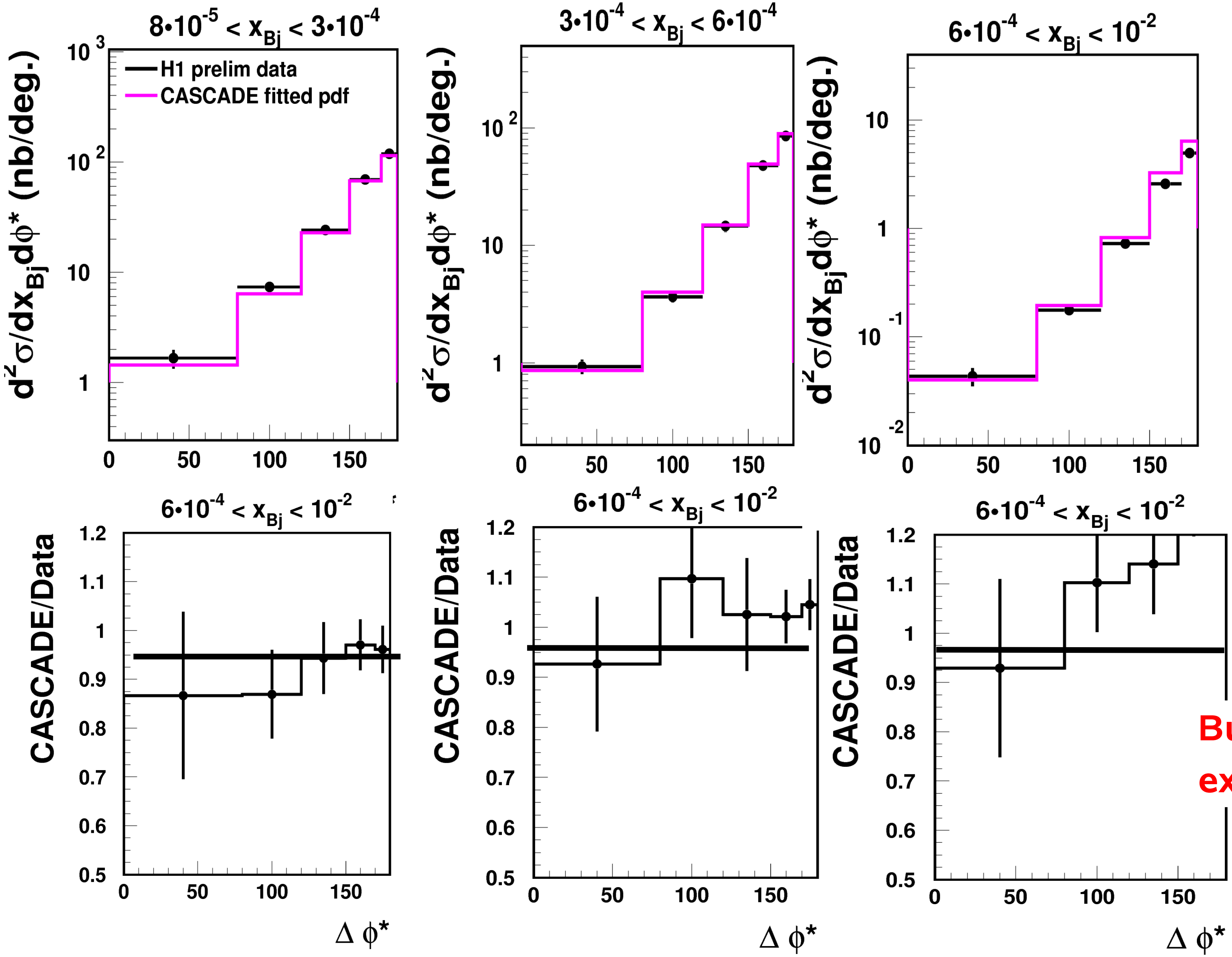
...and compare CASCADE to data with this uPDF



As expected much better description of the data in these bins.

Fitting unintegrated PDFs

...and compare CASCADE to data with this uPDF



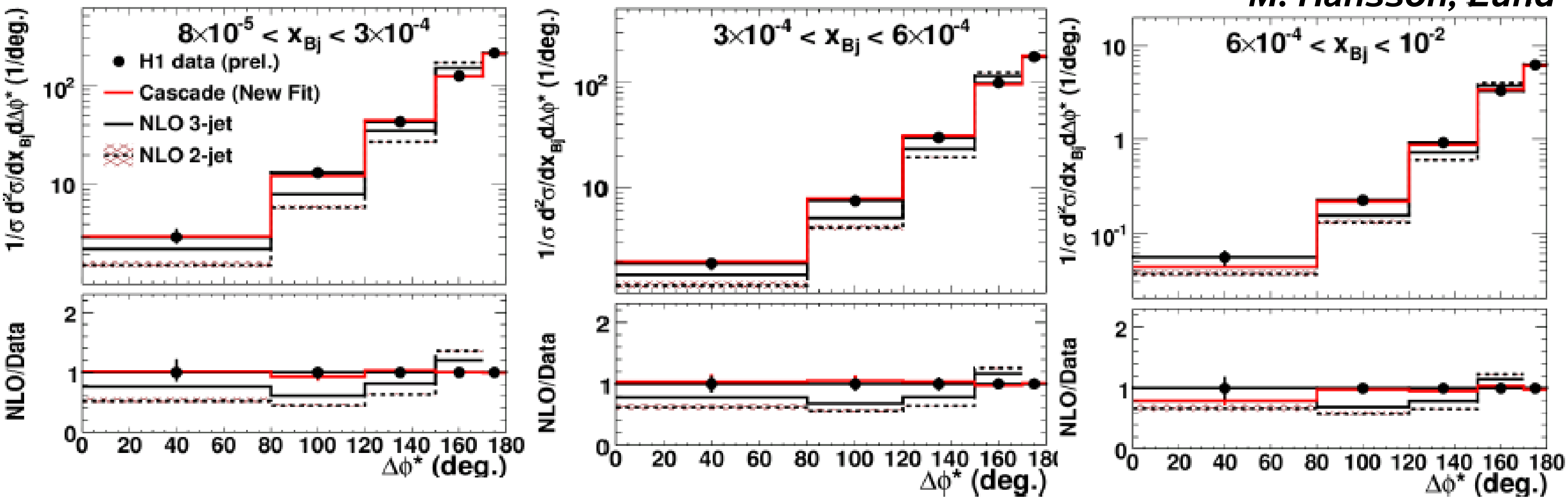
But here worse as expected.

Previous fit

We know that there exists a better minimum.

In the end we should expect something like this...

M. Hansson, Lund



Fit obtained using the conventional fitting method.

Singular Value Decomposition

As before, polynomial

$$\sigma_{\text{poly}}(p_1, p_2) = A + B_1 p_1 + B_2 p_2 + C_1 p_1^2 + C_2 p_2^2 + C_3 p_1 p_2 + H.O.$$

i.e.

$$P_{n,m} X_m = \sigma_{n,poly}$$ **where** $$X_n = (A, B_1, B_2, C_1, C_2, C_3, ...)$$
$$P_n = (1, p_1, p_2, p_1^2, p_1 p_2, p_2^2, ...)$$
$n =$ ***Grid points***

$n > m$ → **Over determined system**

Approach based on SVD algorithm:

To obtain solution we minimize $r = |PX - \sigma|$

where r **is a number as small as the machine accuracy allow us**

Method soon implemented for the PDF fitting.

Tools Used

- **CASCADE – for the physics**
- **MINUIT – for Chi2 minimizations**
- **HZTOOL – basic helper tool**
- **PAW – for visualizations**
- **(Additional Fortran coding)**

Summary

- An alternative method for fitting of uPDF has been presented.
- This method is much less time consuming.
- The method is up and running.
- ...but the results are not satisfactory. More investigations needed.

Project still in testing state...

Outlook

- Singular Value Decomposition!
- Implement error handling of polynomial
- Investigate grid density
- Improve code to N_parameters > 2
- Increase number of data points (other variables/measurements)
- Final fit of uPDF