An analysis of charged-hadron fragmentation functions including Tevatron and LHC data

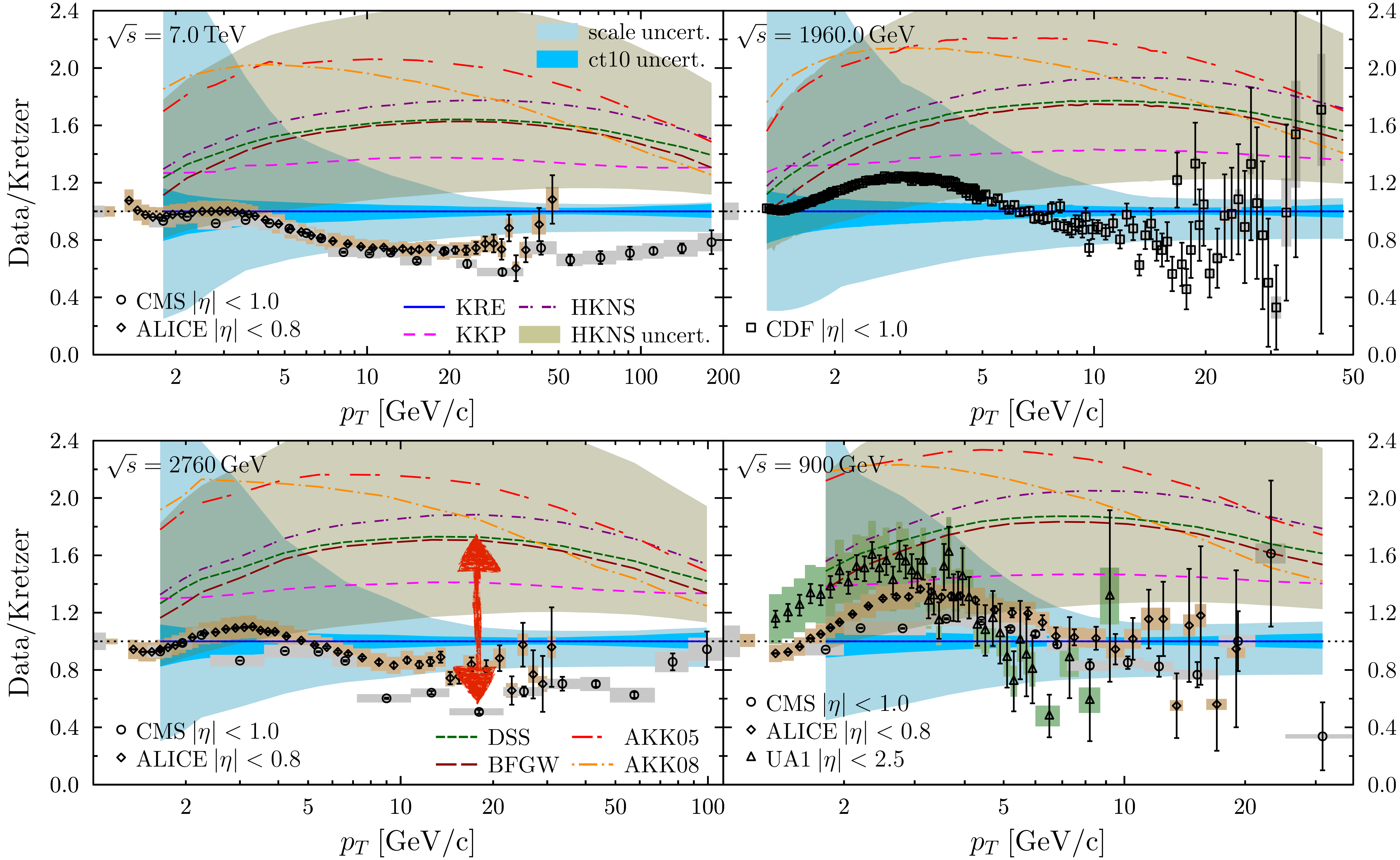
Valerio Bertone
NIKHEF and VU Amsterdam

DIS 2018
April 19, 2018, Kobe (Japan)

Interesting facts

Why bother about fragmentation functions

Comparison data/theory at NLO for inclusive **charged-hadron p_T spectra**:



d'Enterria *et al.* [ArXiv:1311.1415]

- Large energy data tend to be **overshot** by predictions obtained with most of the current FF sets ⇒ too hard gluon FF at large z?
- Not much progress since then.

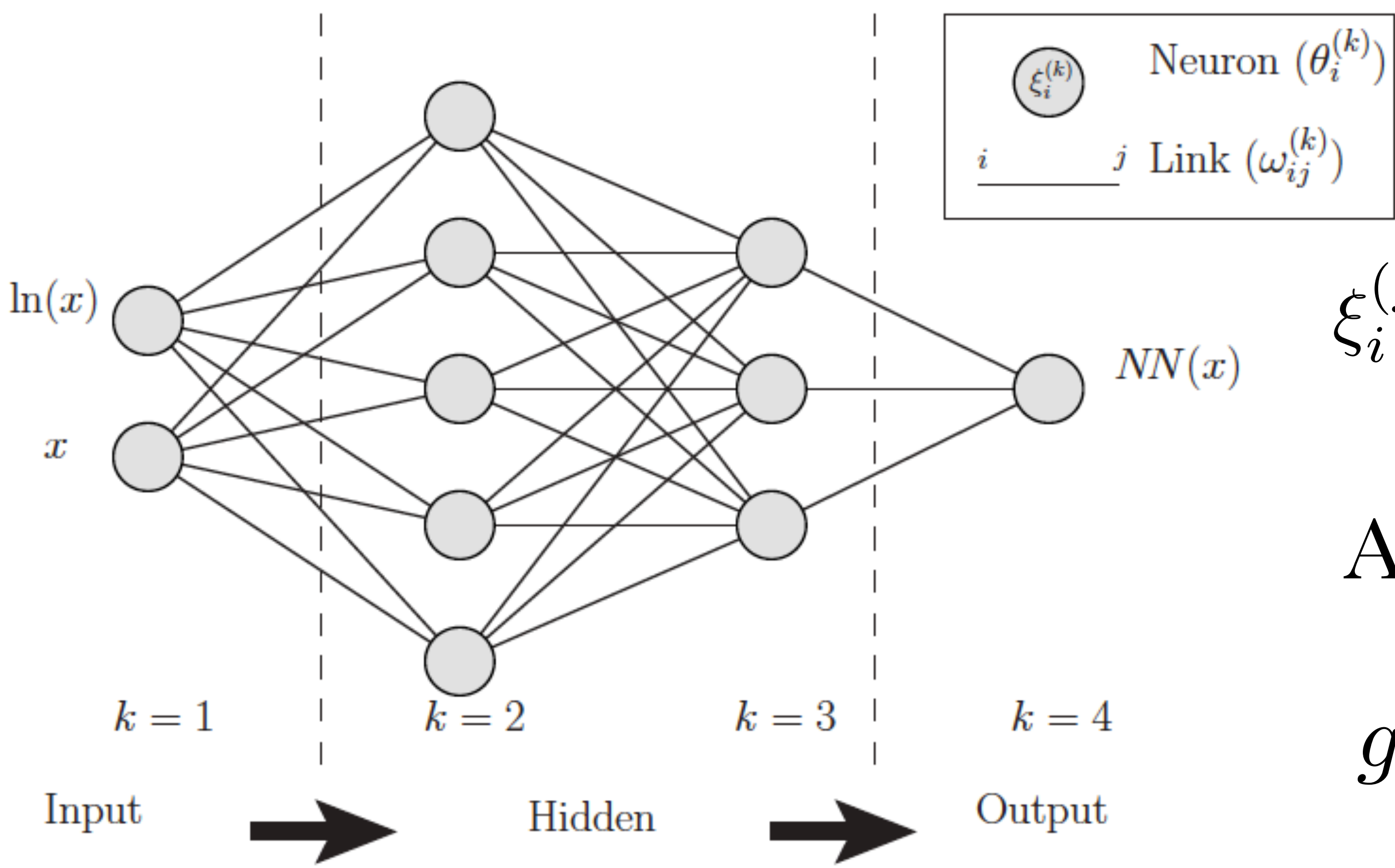
Fitting methodology

The NNPDF approach in a nutshell

- **Monte Carlo** sampling:
  - construct a set of data replicas which reproduces the statistical features of the original dataset,
  - clear statistical interpretation.
- **Neural network (NN)** parameterisation:
  - flexible (redundant) functional form parametrised by a large set of parameters.
- **Genetic algorithm/CMA-ES** for the fit:
  - suitable exploration of the parameter space to avoid to fall into local minima of the figure of merit.
- Determination of the best fit by **cross-validation**:
  - exploit the random distribution of the statistical fluctuations in a given MC replica to avoid over-learning.
- So far, successfully used to extract **PDFs** and **FFs**.

Fitting methodology
A word on the parametrisation

- FFs are parametrised in terms of NNs with architecture (2-5-3-1):



$$\xi_i^{(j)} = g\left(\sum_k^{\text{(j-1)th layer}} \xi_k^{(j-1)} \omega_{ki}^{(j)} - \theta_i^{(j)}\right)$$

Activation function:

$$g(x) = \text{sign}(x)\ln(|x|+1)$$

- Each NN has 37 free parameters.
- FFs are expressed as $f_i(x) = \text{NN}_i(x) - \text{NN}_i(1)$
  - The $\text{NN}_i(1)$ term ensures that $f_i(x) \xrightarrow[x\to 1]{} 0$

DGLAP equations

Time-like evolution

- FFs obey the standard **time-like** DGLAP evolution equations:

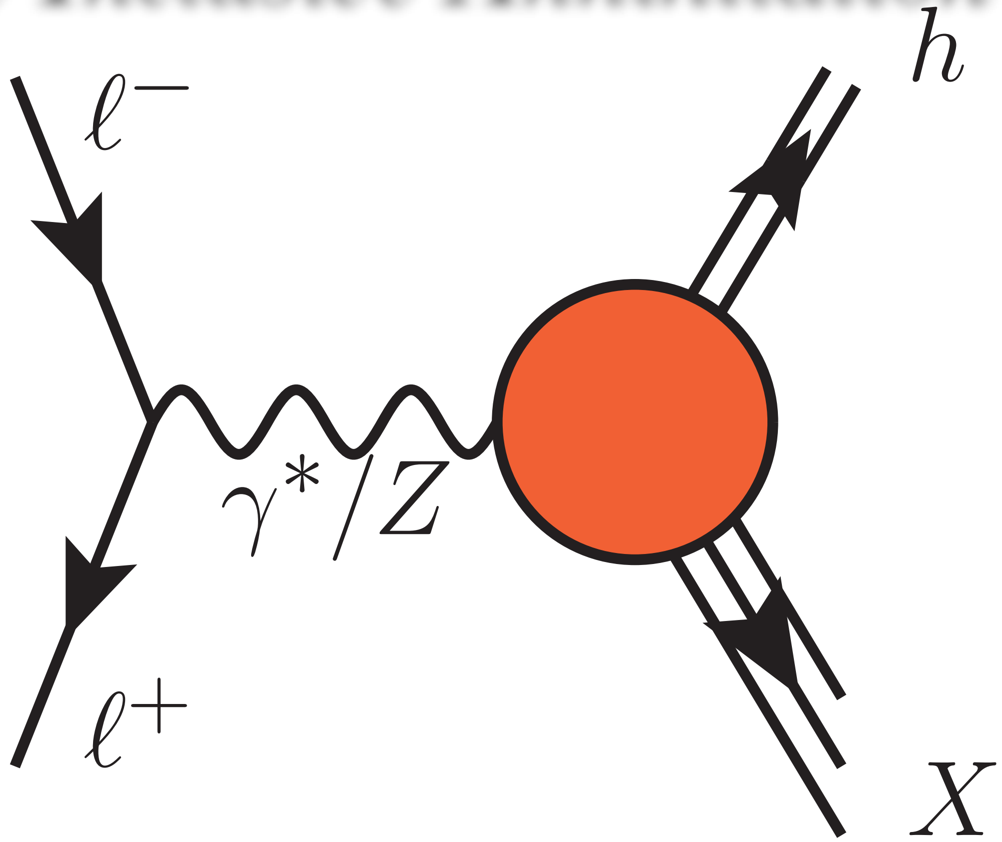
$$\mu^2 \frac{\partial}{\partial \mu^2} D^h_{\rm NS} = P_{\rm NS} \otimes D^h_{\rm NS}$$

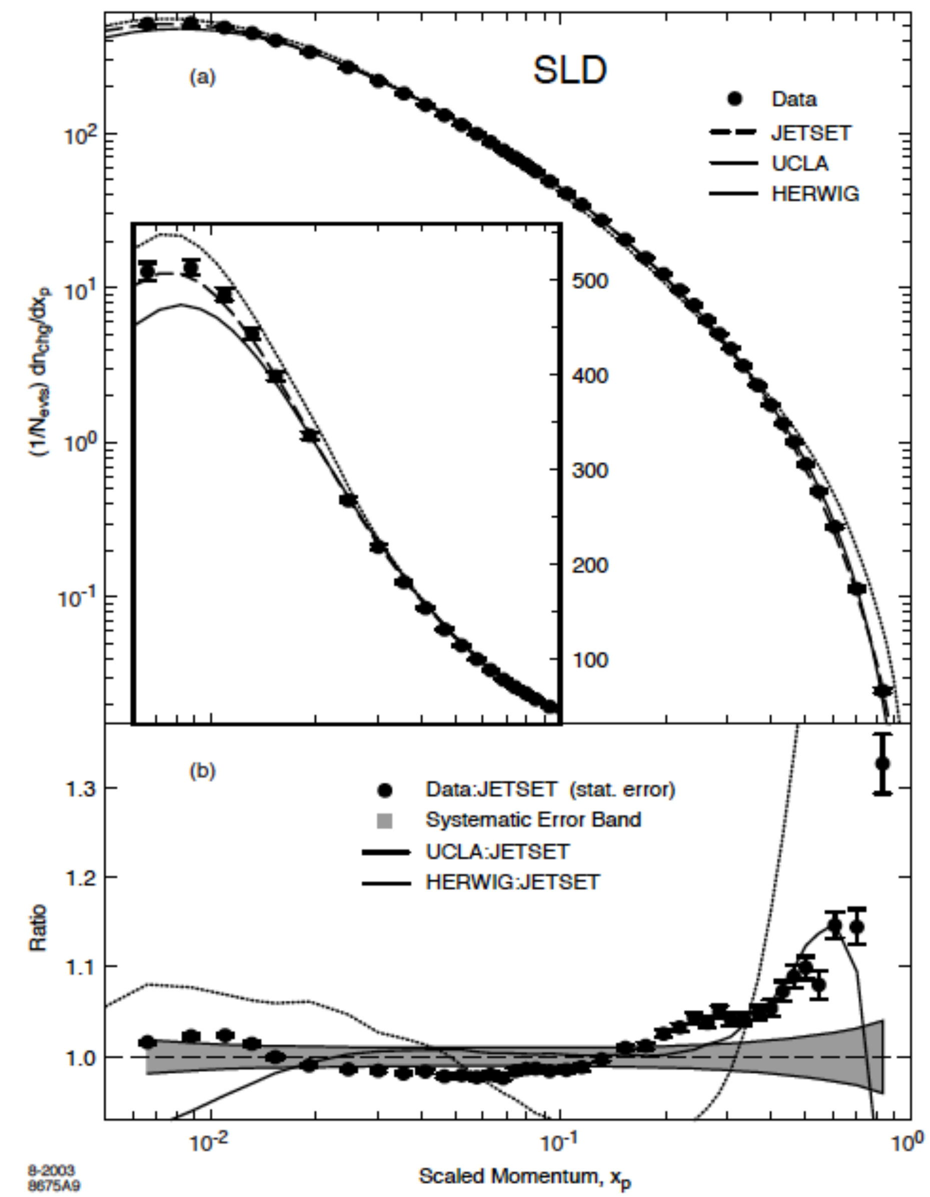
$$\mu^2 \frac{\partial}{\partial \mu^2} \begin{pmatrix} D^h_\Sigma \\ D^h_g \end{pmatrix} = \begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \otimes \begin{pmatrix} D^h_\Sigma \\ D^h_g \end{pmatrix}$$

- Time-like splitting functions known up to NNLO.
  A. Mitov and S. O. Moch [hep-ph/0604160], M. Gluck, E. Reya, and A. Vogt [Phys.Rev. D48 (1993)]
- Numerical implementation in the **APFEL** code:
  V. Bertone, C. Carrazza, J. Rojo [arXiv:1310.1394]
  - careful benchmark against in the the N-space **MELA** code,
    V. Bertone, S. Carrazza, E. R. Nocera [arXiv:1501.00494]
  - perfect agreement with **QCDNUM** (after a correction of a bug in the latter).
    M. Botje [arXiv:1602.08383]

Relevant observables

Single Inclusive Annihilation (SIA)





$$\frac{d\sigma^h}{dz} = \hat{\sigma}_0^h \left[C_q \otimes D_\Sigma^h + C_g \otimes D_g^h + C_{\rm NS} \otimes D_{\rm NS}^h \right]$$

- ✅ **Clean** channel: only FFs involved,
- ✅ **higher-order** corrections to NNLO,
- ✅ **precise data** available.
- ❌ **No flavour separation**,
  - ✅ tagged data for heavy-quark FFs.
- ❌ gluon distribution **suppressed** by α_s.

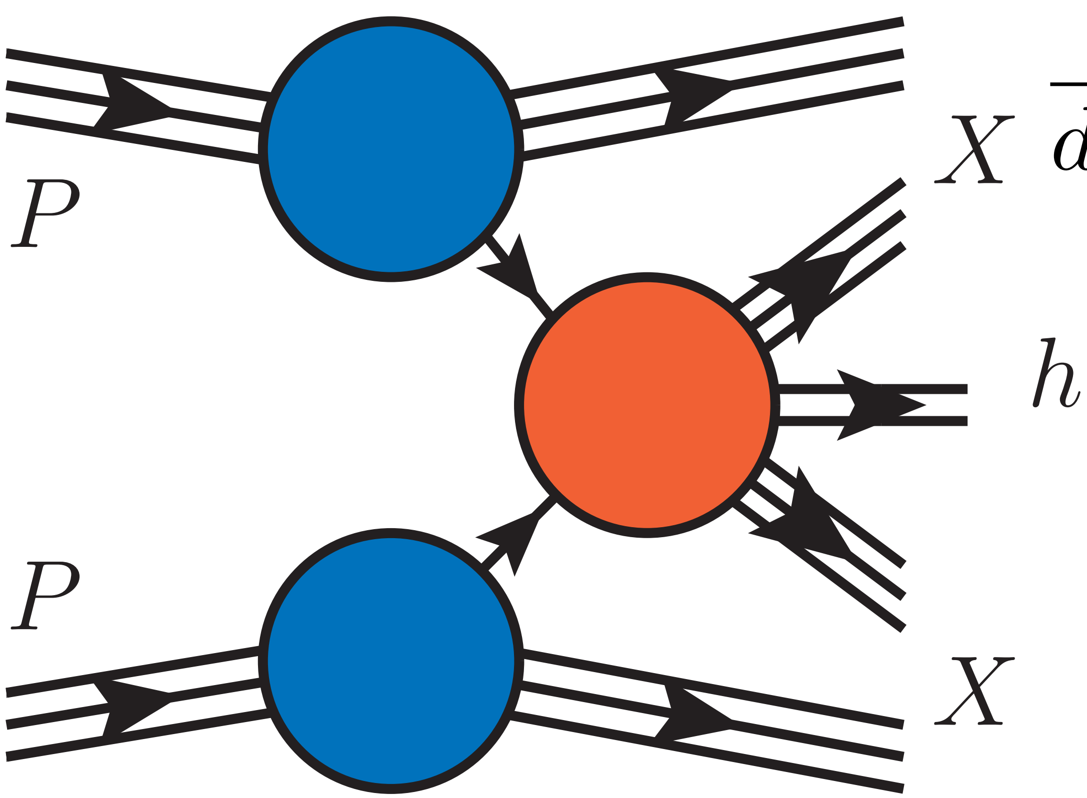
$$D_\Sigma^h = \sum_q \left(D_q^h + D_{\bar{q}}^h \right) = \sum_q D_{q^+}^h$$

$$D_{\rm NS}^h = \sum_q \left(\frac{\hat{e}_q^2}{\langle \hat{e}_q^2 \rangle} - 1 \right) D_{q^+}^h$$

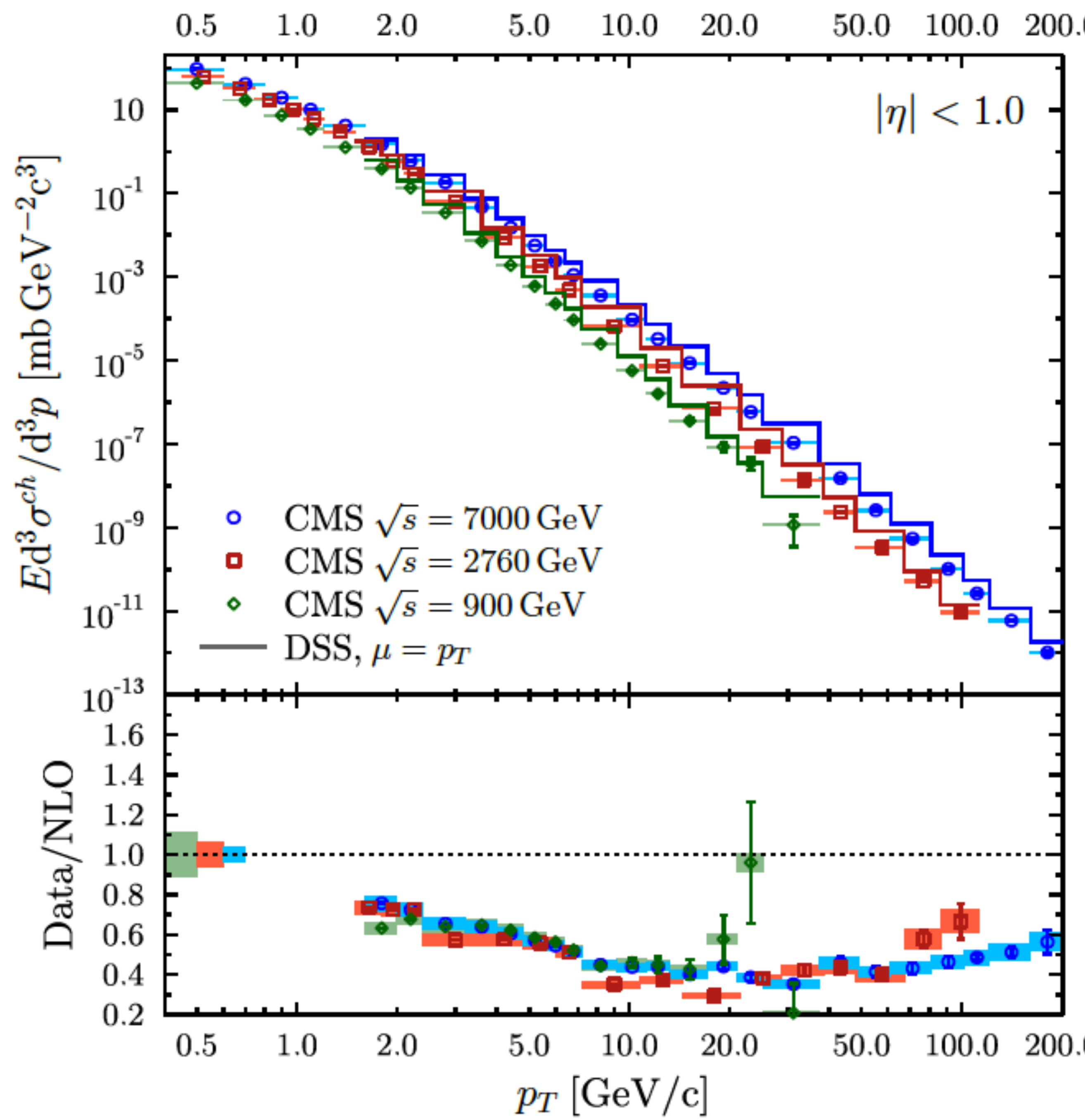
$C_q, C_{\rm NS} \propto \mathcal{O}(1)$ while $C_g \propto \mathcal{O}(\alpha_s)$

Relevant observables

Hadroproduction in proton-proton collisions (pp)



$$\frac{d\sigma^h}{dp_T d\eta} = \sum_{ijk} f_i^{(1)} \otimes f_j^{(2)} \otimes \frac{d\hat{\sigma}_{ijk}}{dp_T d\eta} \otimes D_k^h$$



- ✅ Direct sensitivity to the **gluon FF**,
- ✅ **Scale scan** of FFs ($\mu_F \propto p_T$),
- ✅ Precise data from LHC/Tevatron.
- ❌ Involves both **FFs** and **PDFs**,
- ❌ Known so far up to NLO,
  - ❌ large scale variations at low p_T,
- ❌ cumbersome to compute.

Fit settings

- **Physical parameters**:

$$\alpha_s(M_Z) = 0.118, \quad \alpha_{\rm em}(M_Z) = 1/127, \quad m_c = 1.51 \text{ GeV}, \quad m_b = 4.92 \text{ GeV}$$

- **Parametrisation scale**:

$$Q_0 = 5 \text{ GeV } (> m_c, m_b)$$

  - substantial heavy-quark intrinsic component,
  - heavy-quark FFs parametrised on the same footing as the light FFs.

- **5 independent FFs** for each hadronic species *h*:

$$\{D^h_{u^+}, D^h_{s^+ + d^+}, D^h_{c^+}, D^h_{b^+}, D^h_g\}$$

  - **inclusive SIA data** only constrains three FF combinations,
  - heavy-quark FFs constrained directly by **tagged SIA data**.

- Each FF is parametrised by a **Neural Net** (architecture 2-5-3-1).

- **Kinematic cuts**:

$$z_{\rm min} \leq z \leq z_{\rm max}, \quad z_{\rm min} = \begin{cases} 0.02 & \text{for } \sqrt{s} = M_Z \\ 0.075 & \text{otherwise} \end{cases}, \quad z_{\rm max} = 0.9$$

Fit settings

- **Physical parameters**:

$$\alpha_s(M_Z) = 0.118, \quad \alpha_{\rm em}(M_Z) = 1/127, \quad m_c = 1.51 \text{ GeV}, \quad m_b = 4.92 \text{ GeV}$$

- **Parametrisation scale**:

$$Q_0 = 5 \text{ GeV} \ (> m_c, m_b)$$

  - substantial heavy-quark intrinsic component,
  - heavy-quark FFs parametrised on the same footing as the light FFs.

- **5 independent FFs** for each hadronic species h:

$$\{D^h_{u^+}, D^h_{s^+ + d^+}, D^h_{c^+}, D^h_{b^+}, D^h_g\}$$

  - **inclusive SIA data** only constrains three FF combinations,
  - heavy-quark FFs constrained directly by **tagged SIA data**.

- Each FF is parametrised by a **Neural Net** (architecture 2-5-3-1).

- **Kinematic cuts**:

contributions $\propto M_h/ sz^2$ and $\ln(z)$

$$z_{\rm min} \le z \le z_{\rm max}, \quad z_{\rm min} = \begin{cases} 0.02 & \text{for } \sqrt{s} = M_Z \\ 0.075 & \text{otherwise} \end{cases}, \quad z_{\rm max} = 0.9$$

Fit settings

- **Physical parameters**:

$$\alpha_s(M_Z) = 0.118, \quad \alpha_{\rm em}(M_Z) = 1/127, \quad m_c = 1.51 \text{ GeV}, \quad m_b = 4.92 \text{ GeV}$$

- **Parametrisation scale**:

$$Q_0 = 5 \text{ GeV} \ (> m_c, m_b)$$

  - substantial heavy-quark intrinsic component,
  - heavy-quark FFs parametrised on the same footing as the light FFs.

- **5 independent FFs** for each hadronic species h:

$$\{D^h_{u^+}, D^h_{s^+ + d^+}, D^h_{c^+}, D^h_{b^+}, D^h_g\}$$

  - **inclusive SIA data** only constrains three FF combinations,
  - heavy-quark FFs constrained directly by **tagged SIA data**.

- Each FF is parametrised by a **Neural Net** (architecture 2-5-3-1).

- **Kinematic cuts**:

contributions $\propto \ln(1 - z)$

$$z_{\min} \le z \le z_{\max}, \quad z_{\min} = \begin{cases} 0.02 & \text{for } \sqrt{s} = M_Z \\ 0.075 & \text{otherwise} \end{cases}, \quad z_{\max} = 0.9$$

Charged hadron FFs

An overview

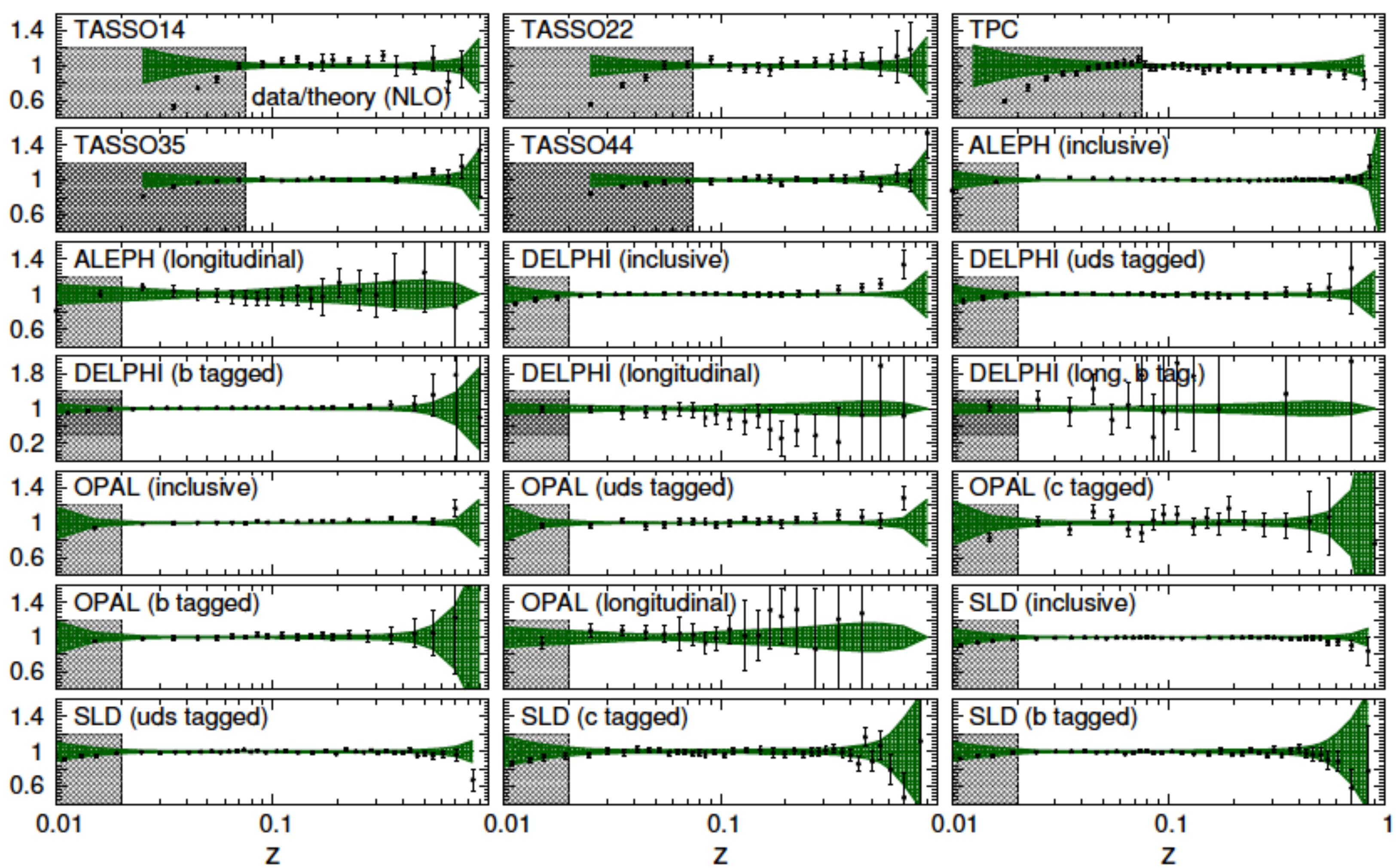
- Many experiments provide data for **charged-hadron** production:
  - this data includes, not only pions, kaons and protons, but **also heavier** (and less abundant) **charged hadrons**.
- Restricting to **SIA experiments**, data is available from:
  - TASSO, TPC, ALEPH, DELPHI, OPAL, SLD.
- Some experiments measure also the **longitudinal** cross section:
  - ALEPH, DELPHI, OPAL.
- Predictions for the longitudinal cross section start at $O(\alpha_s)$:
  - as a consequence it is **not possible to go beyond NLO** (i.e. $O(\alpha_s^2)$) yet.
  - This data provides a strong handle on the **gluon distribution**.

Charged hadron FFs

The NNPDF analysis

E. Nocera [ArXiv:1709.03400]

- General good description of the entire dataset ($\chi^2 / N_{\rm dat} = 0.83$).
- Particularly good the description of the **longitudinal data**.

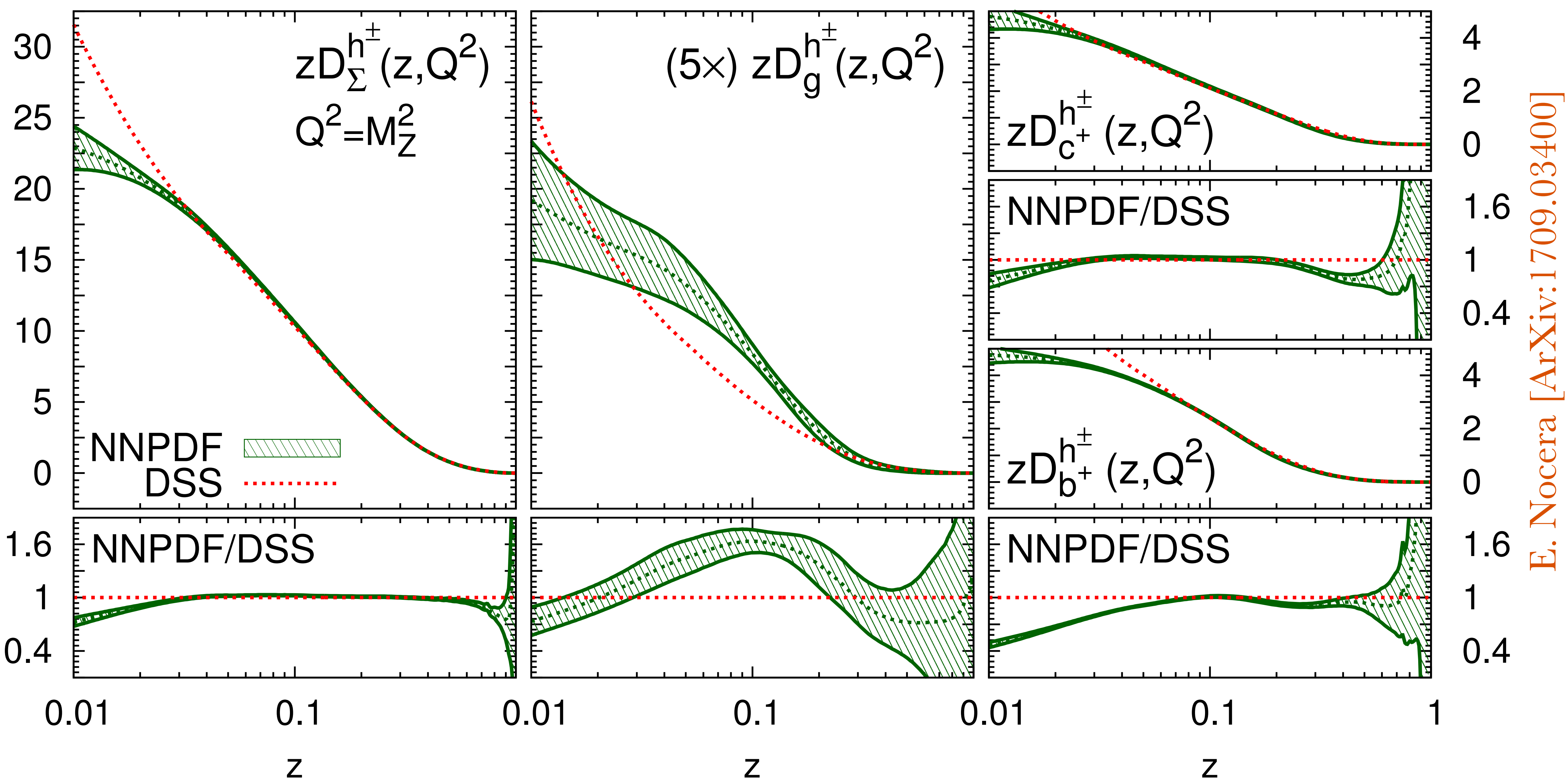


Experiment	Reference	Observable	$\sqrt{s}$ [GeV]	$N_{\rm dat}$	$\chi^2/N_{\rm dat}$
TASSO14	[5]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}$	14.00	15 (20)	1.23
TASSO22	[5]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}$	22.00	15 (20)	0.51
TPC	[6]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}$	29.00	21 (34)	1.65
TASSO35	[5]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}$	35.00	15 (20)	1.14
TASSO44	[5]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}$	44.00	15 (20)	0.68
ALEPH	[7]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}$	91.20	32 (35)	1.04
	[7]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma_L^{h^\pm}}{dz}$	91.20	19 (21)	0.36
DELPHI	[8]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dp_h}$	91.20	21 (27)	0.65
	[8]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dp_h}\Big\vert_{uds}$	91.20	21 (27)	0.17
	[8]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dp_h}\Big\vert_{b}$	91.20	21 (27)	0.82
	[9]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma_L^{h^\pm}}{dz}$	91.20	20 (22)	0.72
	[9]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma_L^{h^\pm}}{dz}\Big\vert_{b}$	91.20	20 (22)	0.44
OPAL	[10]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}$	91.20	20 (22)	2.41
	[10]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{uds}$	91.20	20 (22)	0.90
	[10]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{c}$	91.20	20 (22)	0.61
	[10]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{b}$	91.20	20 (22)	0.21
	[11]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma_L^{h^\pm}}{dz}$	91.20	20 (22)	0.31
SLD	[12]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dp_h}$	91.28	34 (40)	0.75
	[12]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{uds}$	91.28	34 (40)	1.03
	[12]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{c}$	91.28	34 (40)	0.62
	[12]	$\frac{1}{\sigma_{\rm tot}}\frac{d\sigma^{h^\pm}}{dz}\Big\vert_{b}$	91.28	34 (40)	0.97
Total dataset				471 (527)	0.83

Charged hadron FFs

The NNPDF analysis

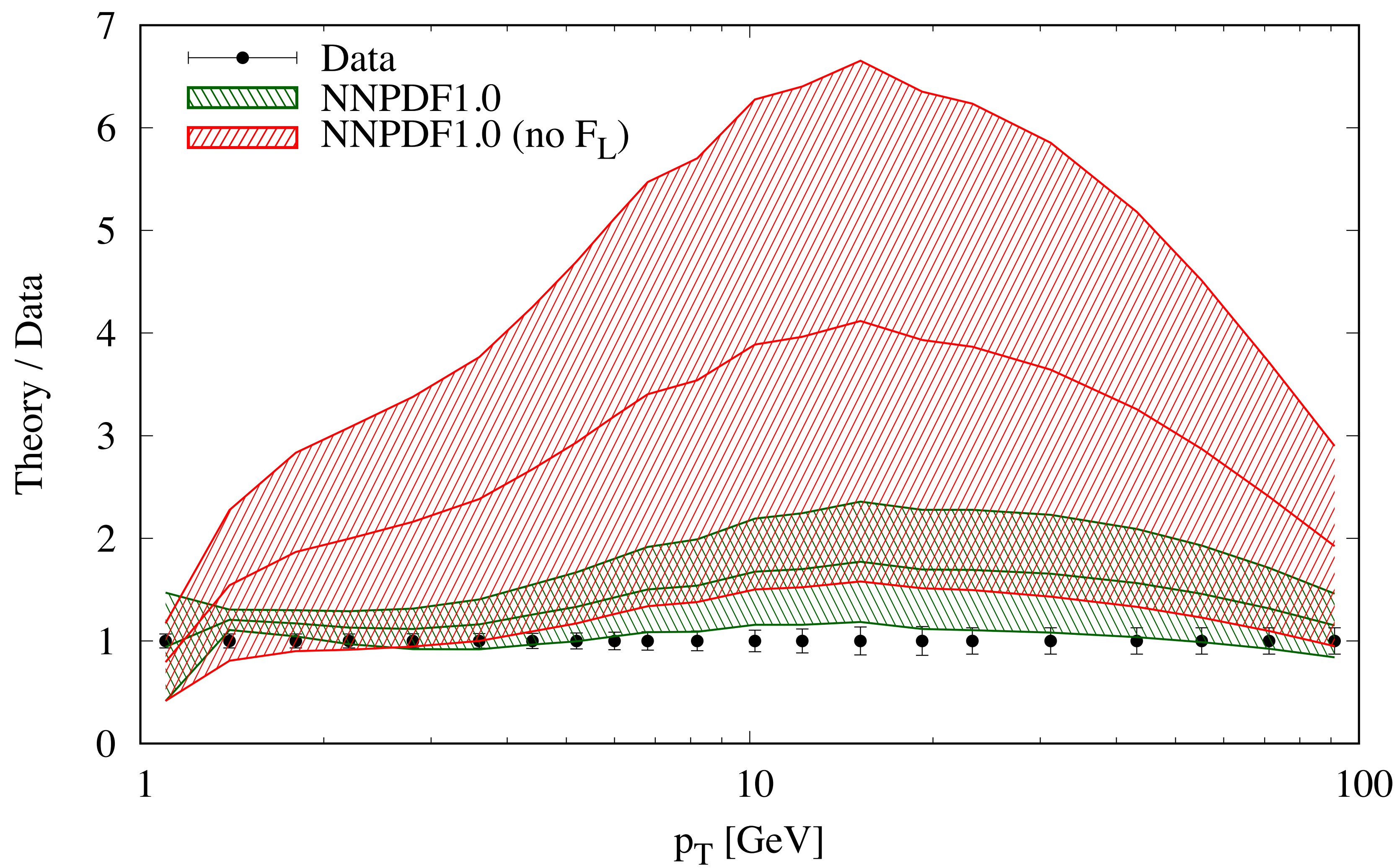
- Charged hadron FFs at the Z-boson mass scale:



- Significant differences w.r.t. DSS, particularly for the gluon.

Charged hadron FFs

Aside: the impact of the longitudinal data for the LHC



- Strong **sensitivity to the gluon distribution**.
- Very significant impact of the **longitudinal data**:
  - dramatic reduction of the **uncertainty**,
  - **better agreement** with CMS data.
- **LHC** and **Tevatron** data expected to have a big impact on FFs.

Charged hadron FFs

Impact hadroproduction in pp collision data

- **CDF** at the *Tevatron*, and **CMS** and **ALICE** at the *LHC* released charged-hadron p_T spectra at different c.o.m. energies:
  - CMS and ALICE at $\sqrt{s}$ = 900, 2760, and 7000 GeV,
  - CDF at $\sqrt{s}$ = 630, 1800, and 1960 GeV.
  - Sensitivity to the **charged-hadron FFs**, particularly to the **gluon**,
- Hard cross sections currently know to **NLO** (i.e. $O(\alpha_s^3)$).
  - **large scale variations** at low p_T. Consider only data with $p_T > 5$ GeV.
  - No CDF data points at 630 GeV survive.
- Include CMS, ALICE, and CDF data in the NNFF1.0 analysis of charged-hadron FFs by means of **Bayesian reweighting**:
  - use **NNPDF3.1** for the PDFs.

Charged hadron FFs

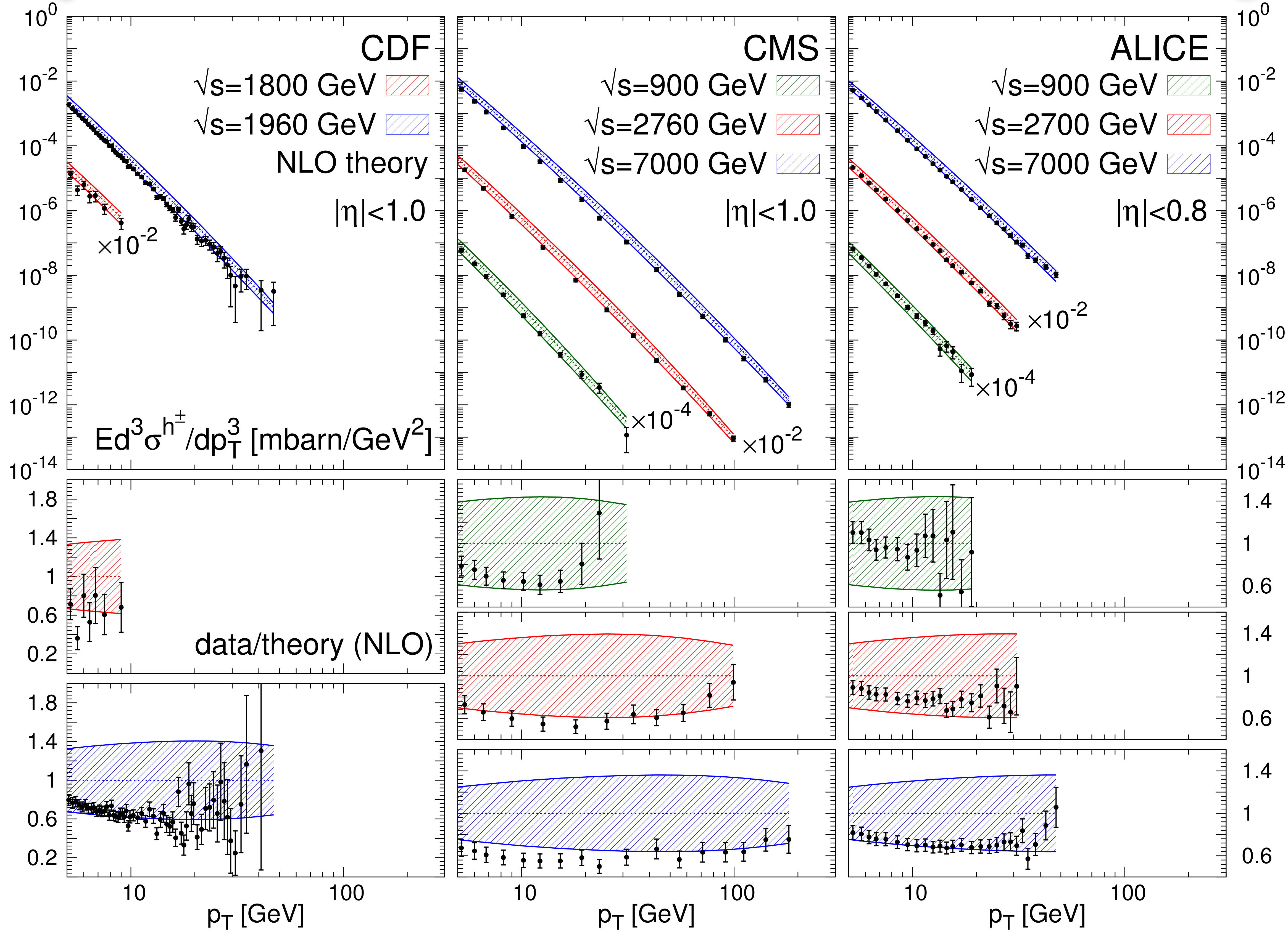
Impact hadroproduction in pp collision data

Experiment	$\sqrt{s}$ [TeV]	$N_{\rm dat}$	$\chi^2_{\rm before}/N_{\rm dat}$	$\chi^2_{\rm after}/N_{\rm dat}$
CDF1800	1.80	7	2.93	**1.36**
CDF1960	1.96	60	3.45	**1.23**
CMS900	0.90	10	3.78	**1.18**
CMS2760	2.76	11	9.31	**1.13**
CMS7000	7.00	17	10.5	**0.98**
ALICE900	0.90	15	4.90	**1.05**
ALICE2760	2.76	21	11.8	**0.96**
ALICE7000	7.00	26	5.21	**0.91**
TOTAL		167	6.01	**1.27**

- Substantial improvement of the single χ^2's as well as of the global one:
  - no tension between the different datasets.

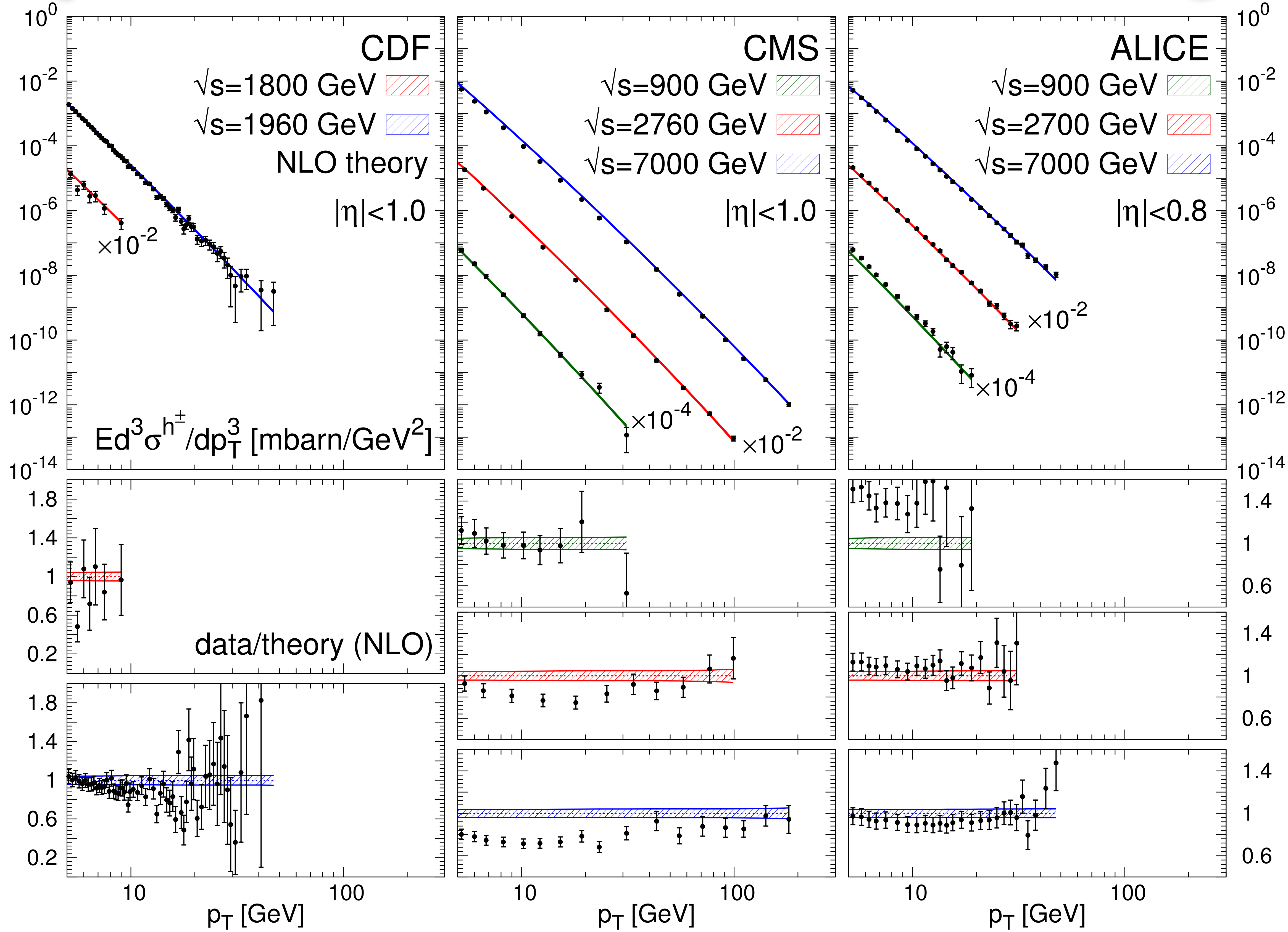
Charged hadron FFs

Comparison to LHC/Tevatron data BEFORE reweighting



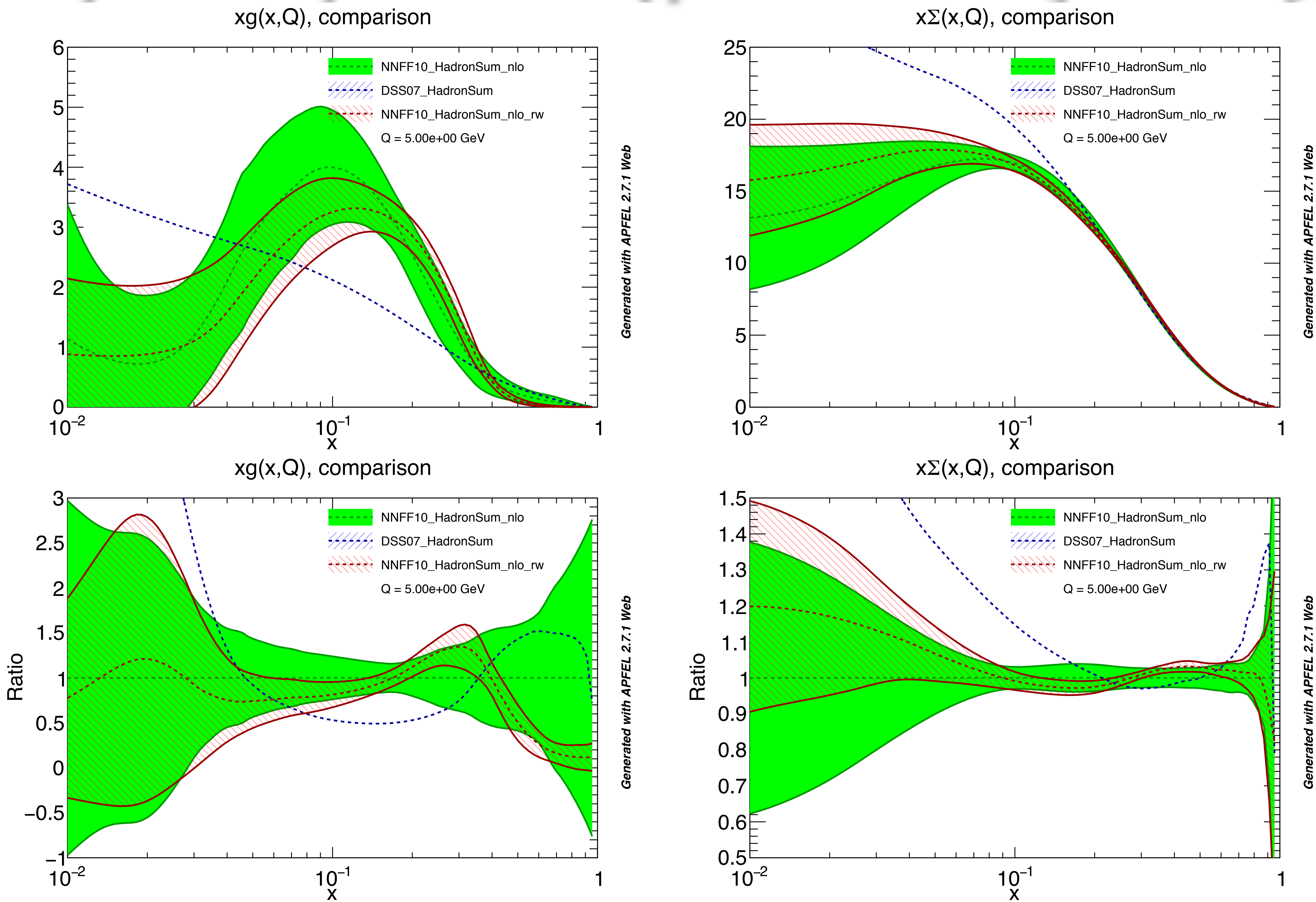
Charged hadron FFs

Comparison to LHC/Tevatron data AFTER reweighting



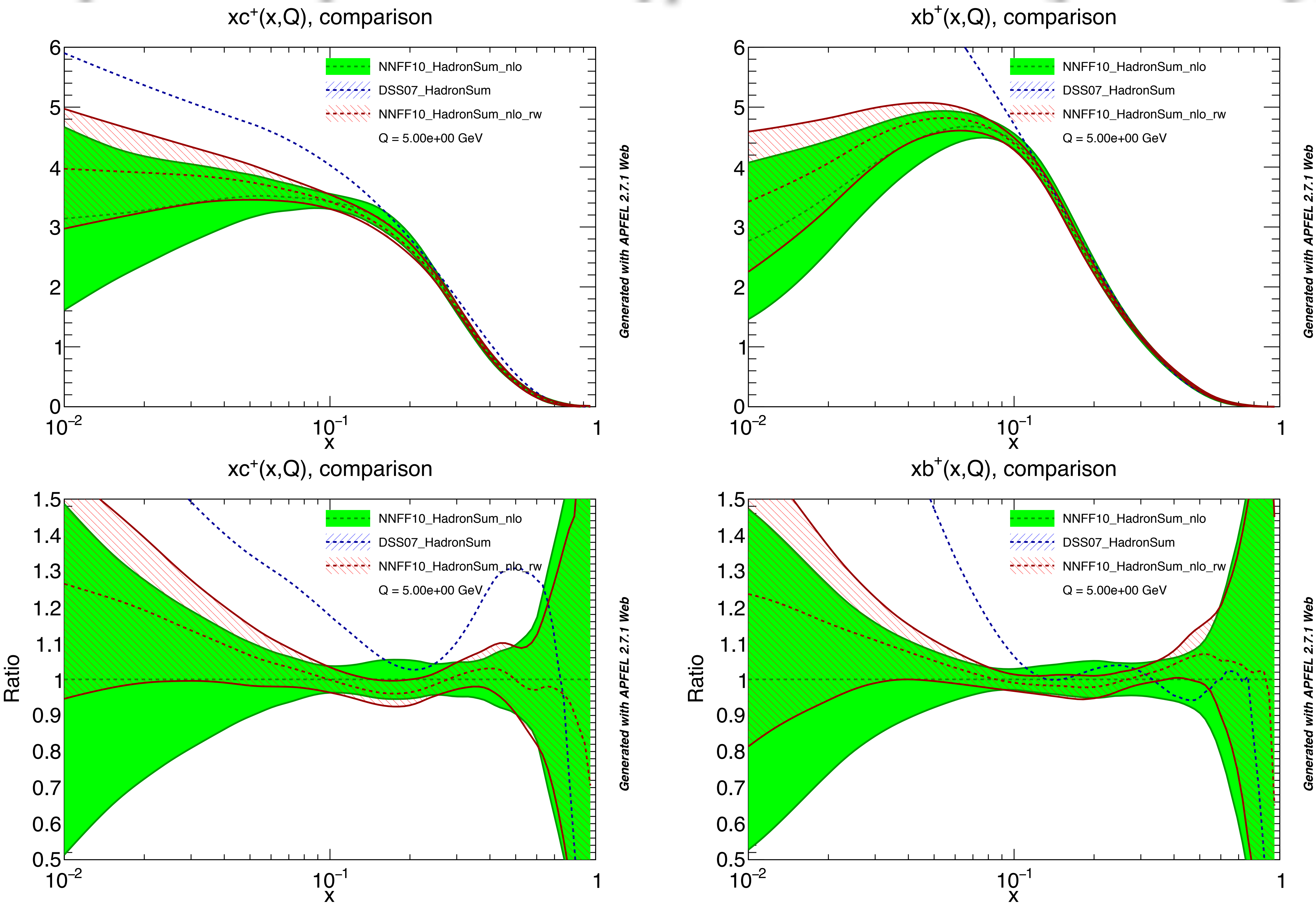
Charged hadron FFs

Impact hadroproduction in pp collision data (preliminary)



Charged hadron FFs

Impact hadroproduction in pp collision data (preliminary)



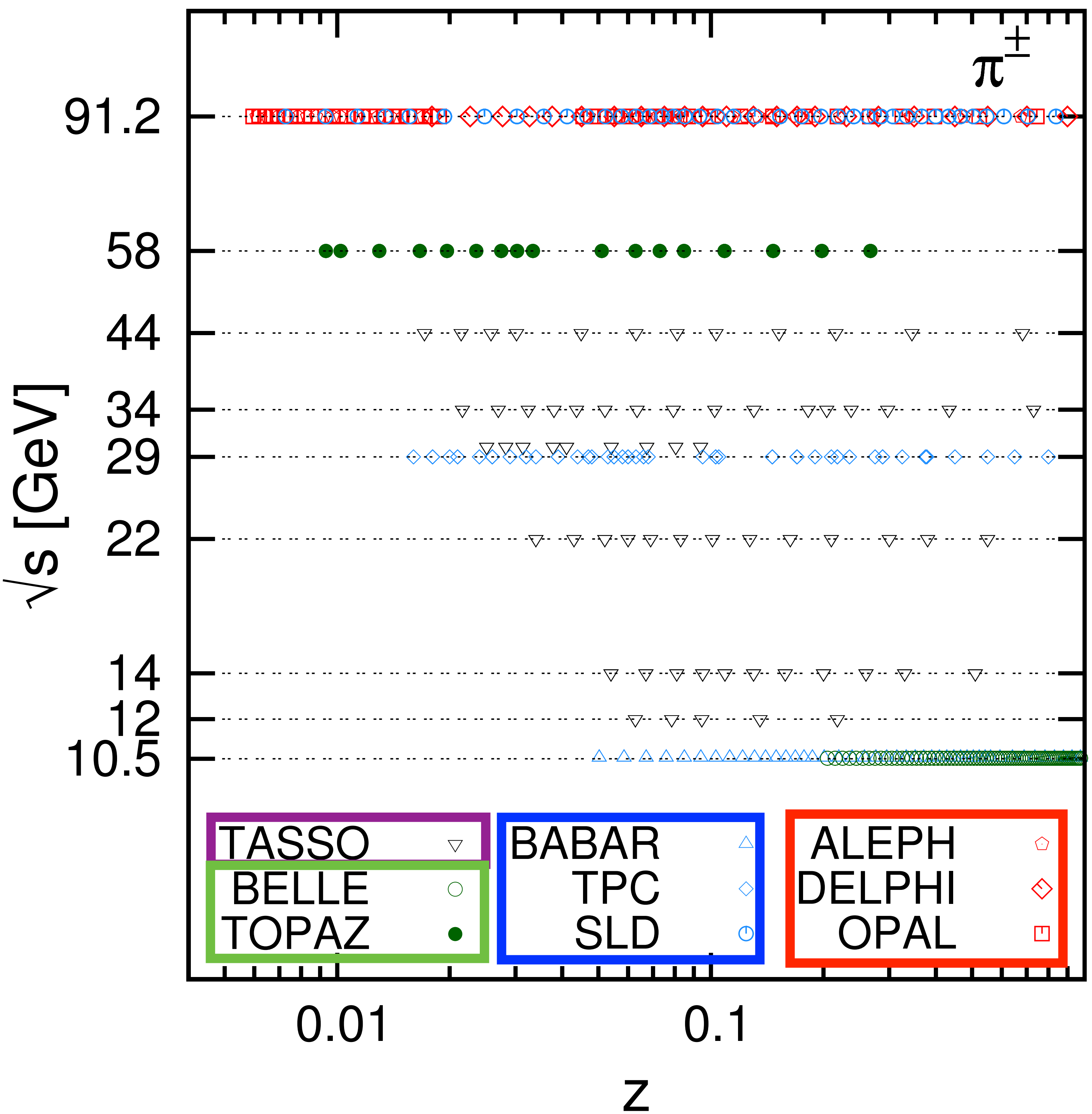
Summary

- **First** determination of charged-hadron FFs including **Tevatron** and **LHC** data:
  - **CMS** and **ALICE** at $\sqrt{s}$ = 900, 2760, and 7000 GeV,
  - **CDF** at $\sqrt{s}$ = 1800, and 1960 GeV.
  - 167 datapoints in total.
- Inclusion of these datasets by **reweighing** of a pre-existing FF set based on **SIA data only**.
- Remarkable **consistency** of these datasets:
  - **simultaneous** inclusion of all of them,
  - very good description of **all single datasets**.
- Strong impact on the **gluon FF**:
  - dramatic reduction of the **uncertainty**, particularly at large z,
  - much **softer** large-z gluon (in agreement with the predictions d'Enterria et. al).

Backup slides

Dataset

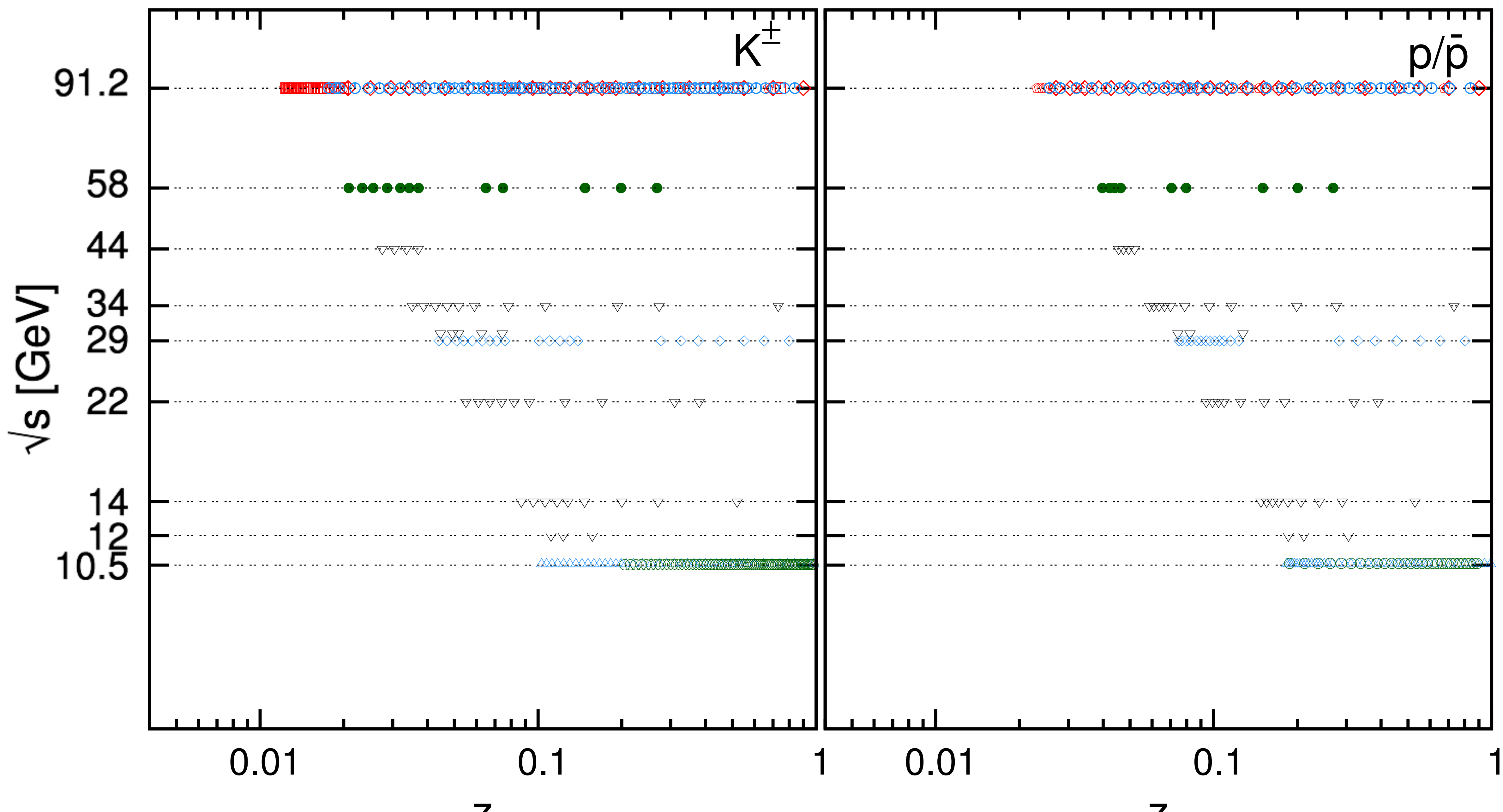
- **Only SIA cross sections** (normalised and absolute) included.



- PETRA (DESY)
  - Phys. Lett. B94 (1980) 444
  - Z.Phys. C17 (1983) 5-15
  - Z.Phys. C42 (1989) 189
- KEK
  - Phys. Rev. Lett. 111 (2013) 062002
  - Phys. Rev. D92 (2015), no. 9 092007
  - Phys. Lett. B345 (1995) 335-342
- SLAC
  - Phys. Rev. D88 (2013) 032011
  - Phys. Rev. Lett. 61 (1988) 1263
  - Phys. Rev. D69 (2004) 072003
  - X.Q. Lu, Ph.D. thesis, Johns Hopkins U., 1986
- LEP (CERN)
  - Z. Phys. C66 (1995) 355-366
  - Eur. Phys. J. C5 (1998) 585-620
  - Z. Phys. C63 (1994) 181-196.
- 428 data points (after cuts)

Dataset

- **Only SIA cross sections** (normalised and absolute) included.
- We have fitted FFs also to $K^{\pm}$ and $p/\overline{p}$ data.



385 data points (after cuts)

360 data points (after cuts)

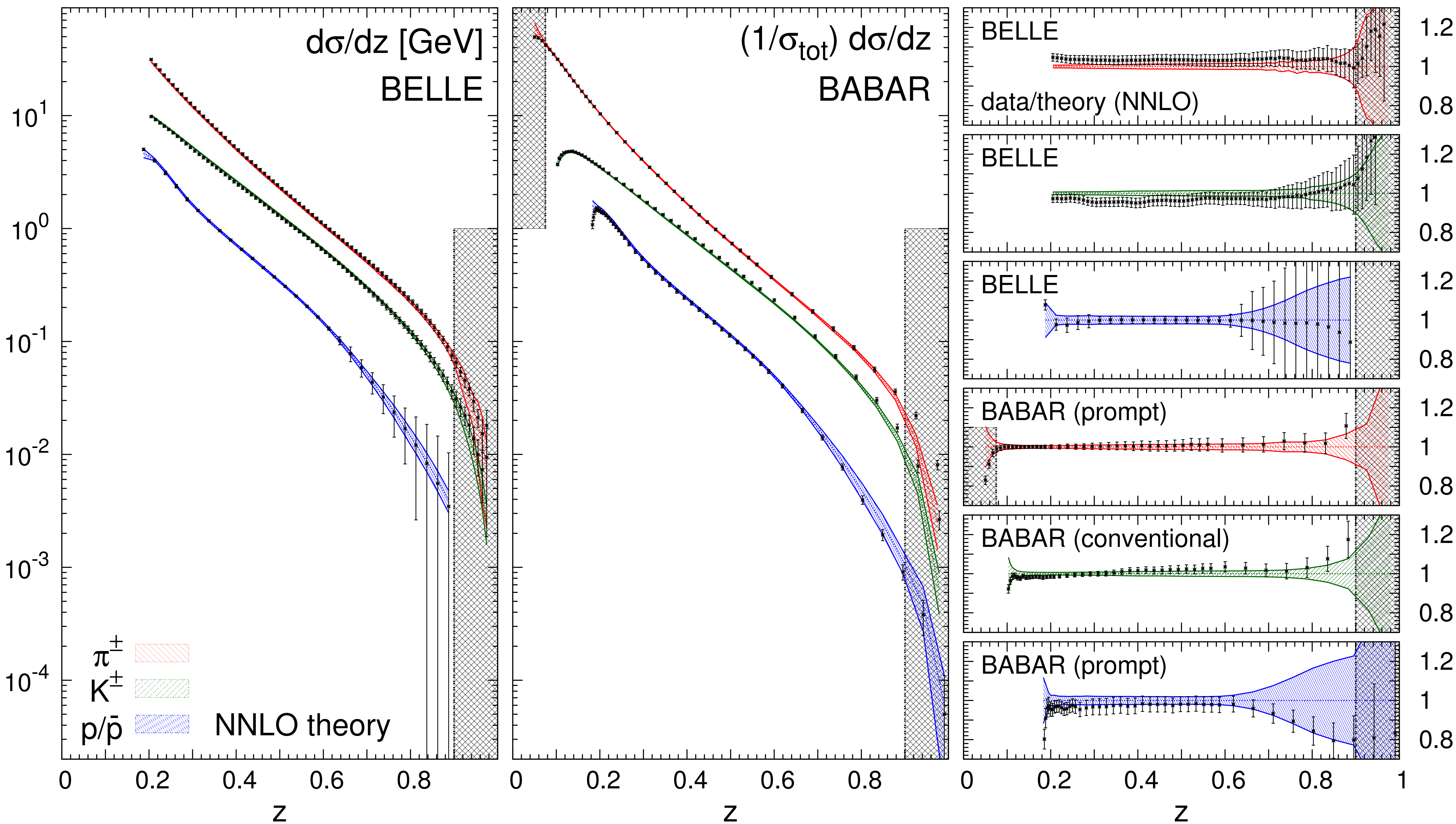
Fit quality

- Fit quality **increasingly better** going from LO to NNLO:
  - substantial from LO to NLO, more moderate from NLO to NNLO.
  - **NNLO** corrections are anyway **beneficial** (particularly for pions).
- **Tension** between BELLE and BABAR for kaons and protons:
  - **opposite trend** upon inclusion of higher-order corrections.
- **Anomalously small** χ^2 for BELLE:
  - possible underestimate of the **uncorrelated systematic** uncertainty.
- **Possible tension** also between DELPHI inclusive and the other experiments at M_Z:
  - opposite trend upon inclusion of higher-order corrections.

	$\chi^2/N_{\rm dat}$ $(h=\pi^\pm)$			$\chi^2/N_{\rm dat}$ $(h=K^\pm)$			$\chi^2/N_{\rm dat}$ $(h=p/\bar{p})$		
Exp.	LO	NLO	NNLO	LO	NLO	NNLO	LO	NLO	NNLO
BELLE	0.60	0.11	0.09	0.21	0.32	0.33	0.10	0.31	0.50
BABAR	1.91	1.77	0.78	2.86	1.11	0.95	4.74	3.75	3.25
TASSO12	0.70	0.85	0.87	1.10	1.03	1.02	0.69	0.70	0.72
TASSO14	1.55	1.67	1.70	2.17	2.13	2.07	1.32	1.25	1.22
TASSO22	1.64	1.91	1.91	2.14	2.77	2.62	0.98	0.92	0.93
TPC (incl.)	0.46	0.65	0.85	0.94	1.09	1.01	1.04	1.10	1.08
TPC (*uds* tag)	0.78	0.55	0.49	—	—	—	—	—	—
TPC (*c* tag)	0.55	0.53	0.52	—	—	—	—	—	—
TPC (*b* tag)	1.44	1.43	1.43	—	—	—	—	—	—
TASSO30	—	—	—	—	—	—	0.25	0.19	0.18
TASSO34	1.16	0.98	1.00	0.27	0.44	0.36	0.82	0.81	0.78
TASSO44	2.01	2.24	2.34	—	—	—	—	—	—
TOPAZ	1.04	0.82	0.80	0.61	1.19	0.99	0.79	1.21	1.19
ALEPH	1.68	0.90	0.78	0.47	0.55	0.56	1.36	1.43	1.28
DELPHI (incl.)	1.44	1.79	1.86	0.28	0.33	0.34	0.48	0.49	0.49
DELPHI (*uds* tag)	1.30	1.48	1.54	1.38	1.49	1.32	0.47	0.46	0.45
DELPHI (*b* tag)	1.21	0.99	0.95	0.58	0.49	0.52	0.89	0.89	0.91
OPAL	2.29	1.88	1.84	1.67	1.57	1.66	—	—	—
SLD (incl.)	2.33	1.14	0.83	0.86	0.62	0.57	0.66	0.65	0.64
SLD (*uds* tag)	0.95	0.65	0.52	1.31	1.02	0.93	0.77	0.76	0.78
SLD (*c* tag)	3.33	1.33	1.06	0.92	0.47	0.38	1.22	1.22	1.21
SLD (*b* tag)	0.45	0.38	0.36	0.59	0.67	0.62	1.12	1.29	1.33
Total dataset	**1.44**	**1.02**	**0.87**	**1.02**	**0.78**	**0.73**	**1.31**	**1.23**	**1.17**

Description of the data

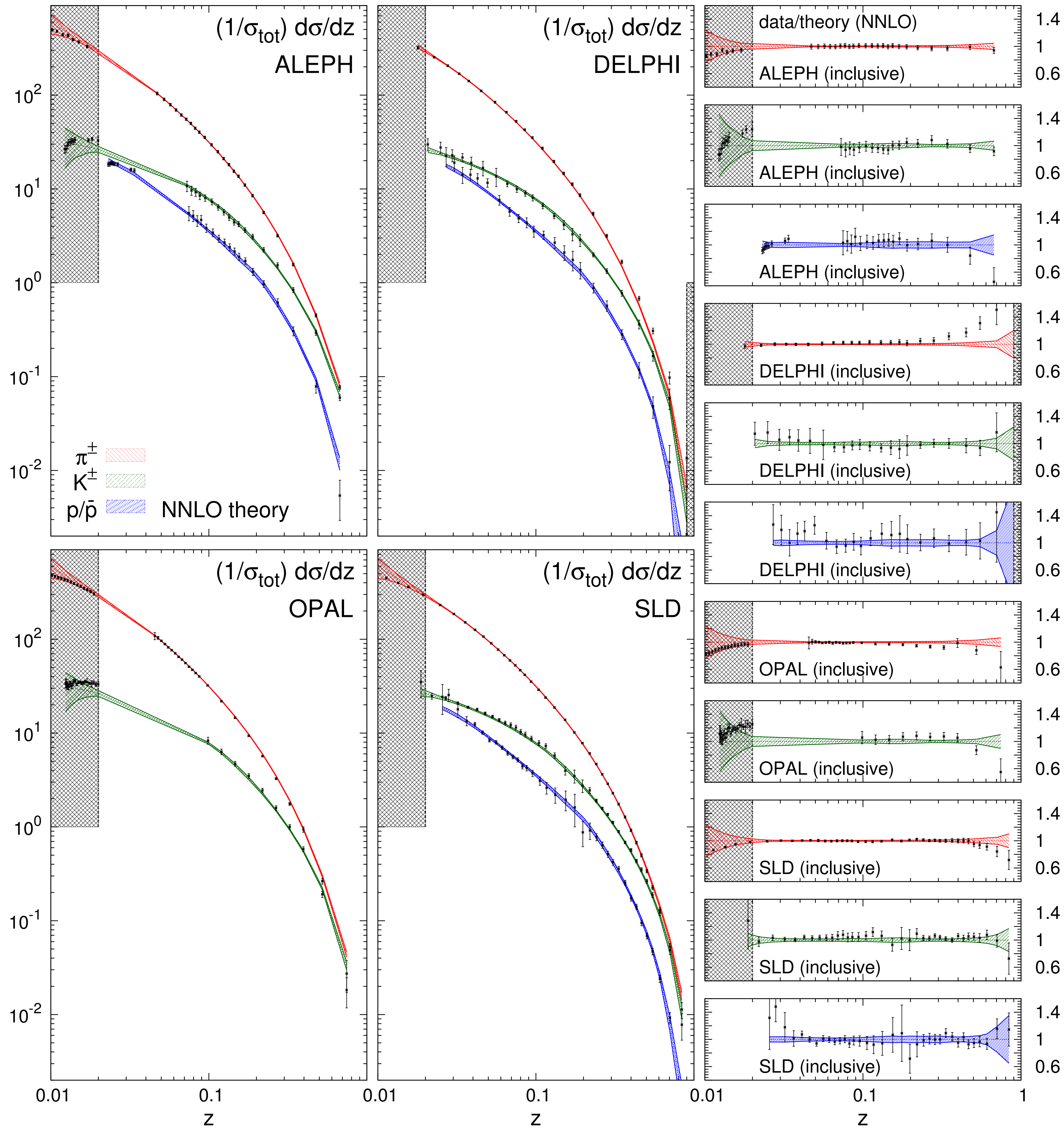
- Data/Theory comparison for **BELLE** and **BABAR** using NNFF1.0 at NNLO:
  - the bands indicate the 1-σ uncertainty.



- Very good description in the region not excluded by the kinematic cuts (shaded areas).
- Different **trend** of the data at **low *z*** for **kaons** and particularly for **protons**:
  - possible reason of the worsening of the χ^2.

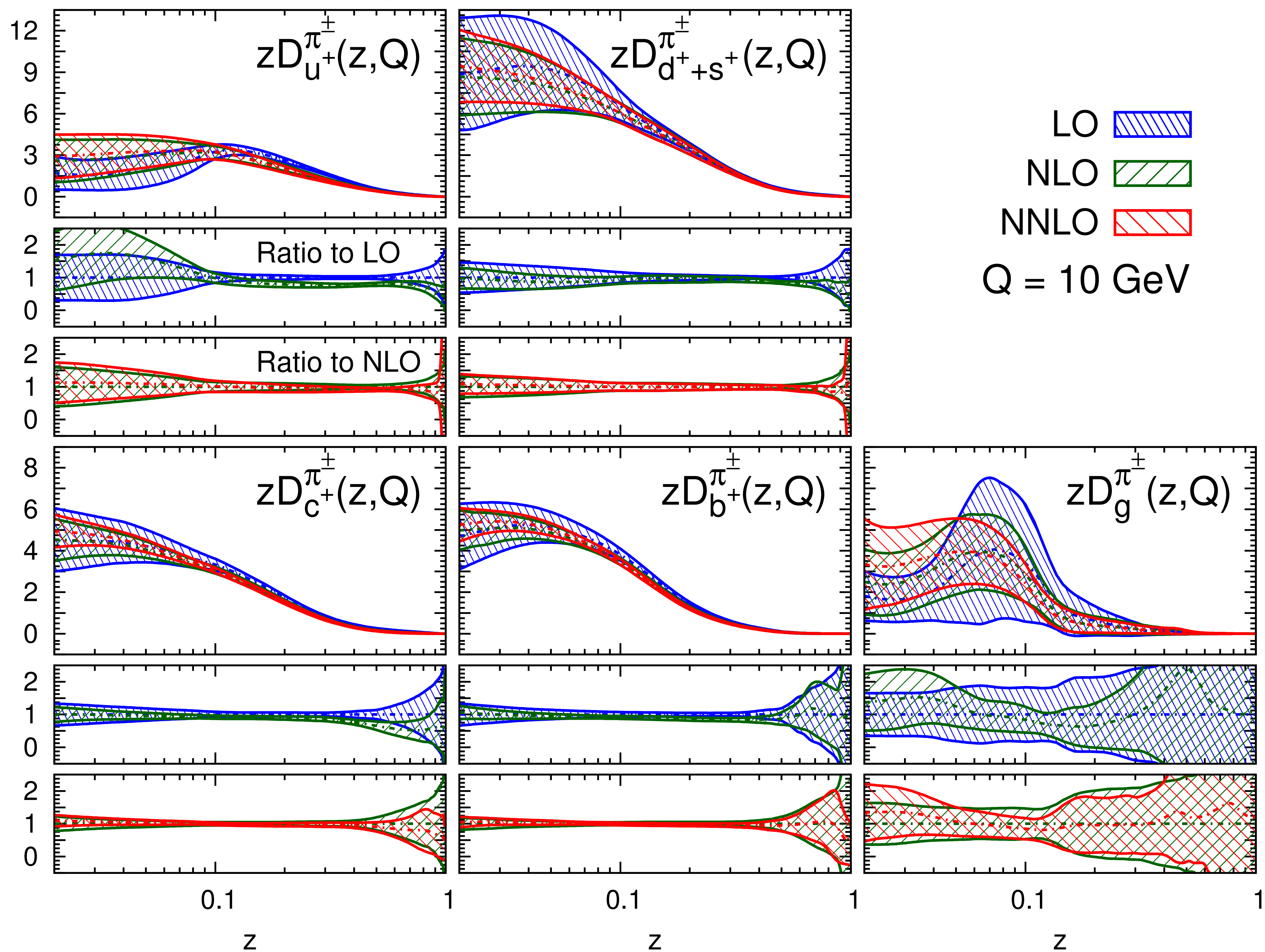
Description of the data

- Data/Theory comparison for the experiments at M_Z using NNFF1.0 at NNLO.
- Very good description in the region allowed by the kinematic cuts.
- Often also the data excluded by the cuts are well described.
- The predictions for pions for **DELPHI** overshoot the data:
  - origin of the worse χ^2 as compared to the other experiments at M_Z.



Fragmentation functions

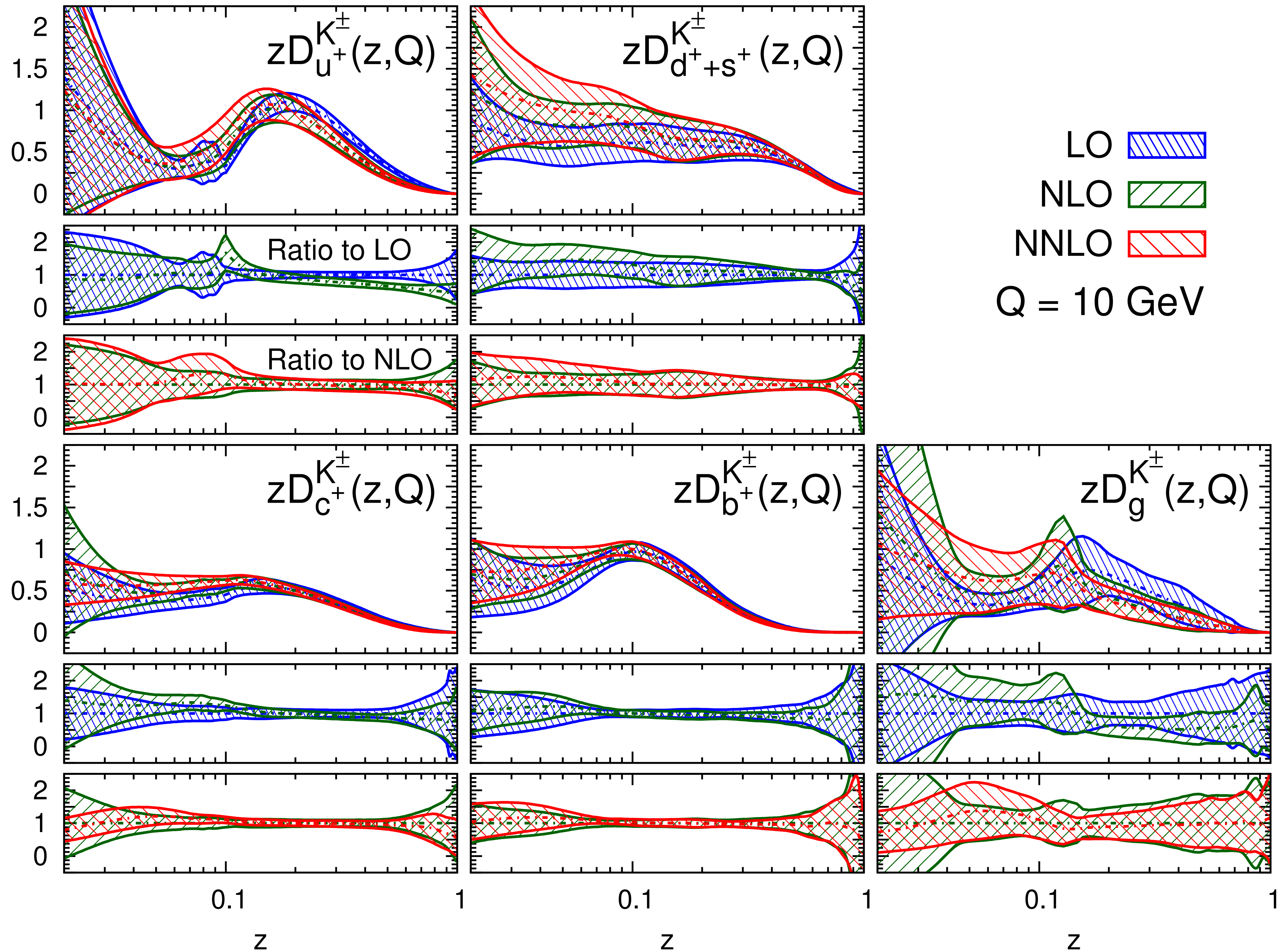
Perturbative stability (Pions)



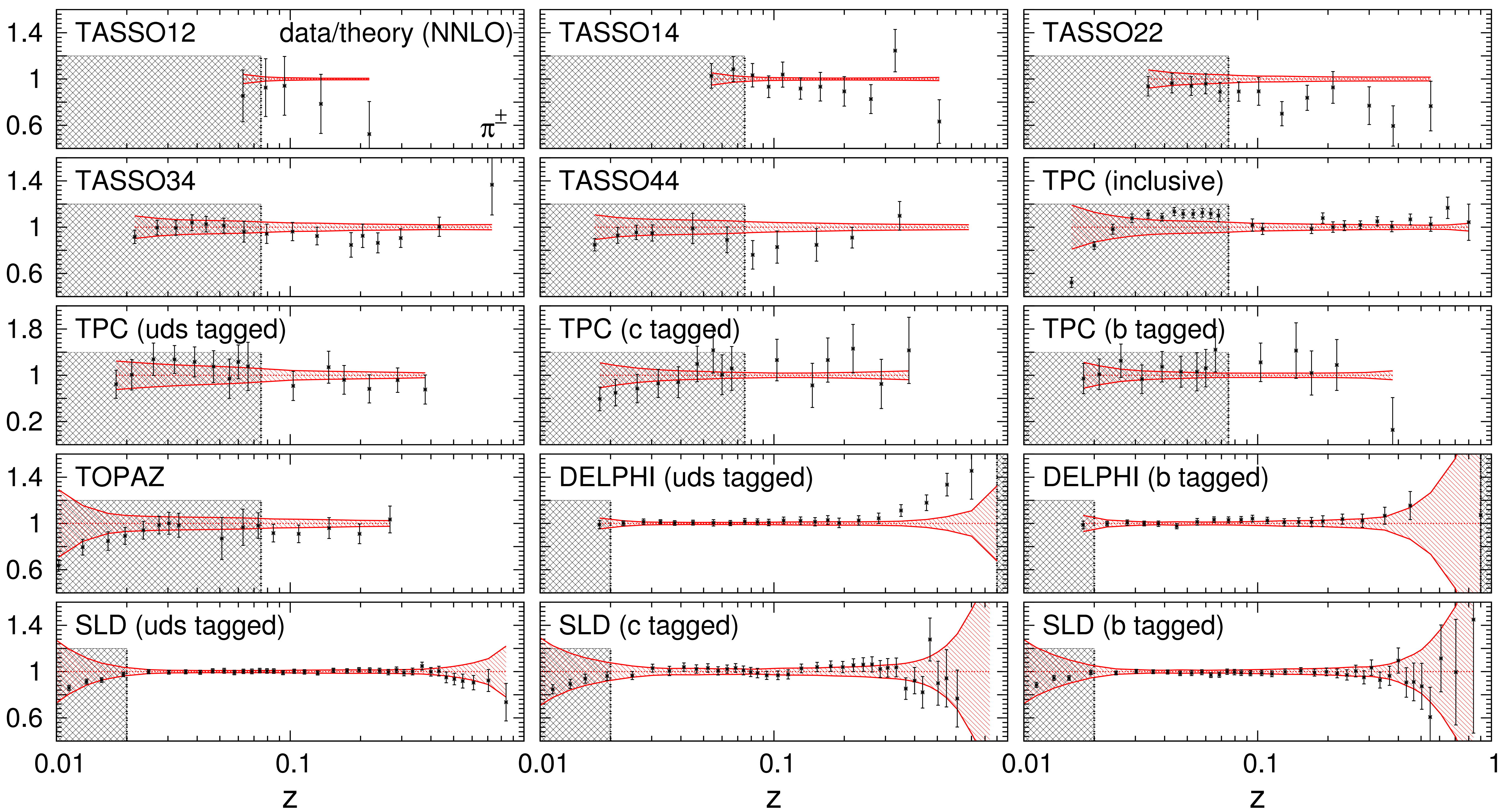
- **Stabilisation** going from LO to NNLO,
- LO uncertainties slightly larger: poorer theoretical description.

Fragmentation functions

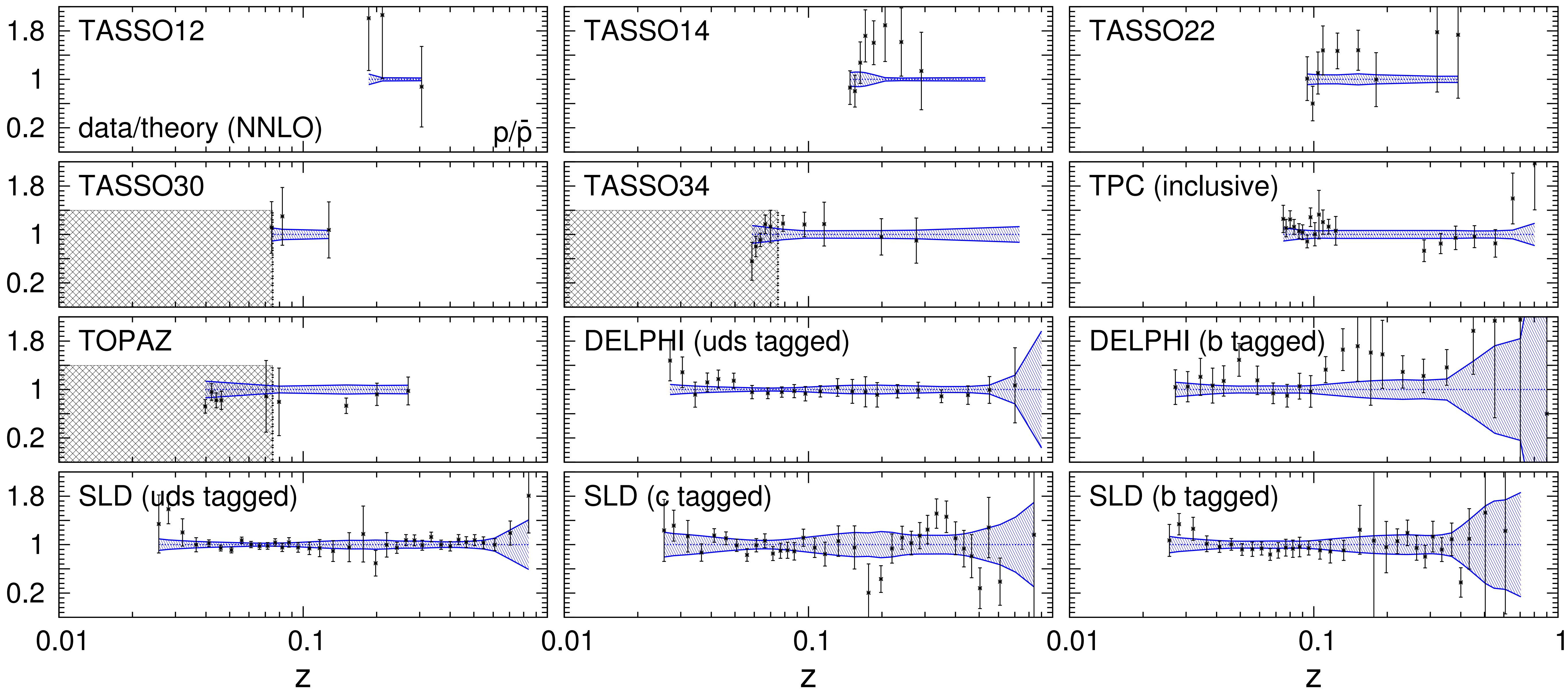
Perturbative stability (Kaons)



- Same for kaons…



TASSO12
data/theory (NNLO)
$\pi^{\pm}$
TASSO14
TASSO22
TASSO34
TASSO44
TPC (inclusive)
TPC (uds tagged)
TPC (c tagged)
TPC (b tagged)
TOPAZ
DELPHI (uds tagged)
DELPHI (b tagged)
SLD (uds tagged)
SLD (c tagged)
SLD (b tagged)
1.4
1
0.6
1.8
0.2
0.01
0.1
1
z


TASSO12
data/theory (NNLO)
p/p̄
TASSO14
TASSO22
TASSO30
TASSO34
TPC (inclusive)
TOPAZ
DELPHI (uds tagged)
DELPHI (b tagged)
SLD (uds tagged)
SLD (c tagged)
SLD (b tagged)
1.8
1
0.2
0.01
0.1
1
z