Progress on aMCfast

Valerio Bertone

University of Oxford

ERC miniworkshop

July 4 - 5, 2016, CERN

Recap on aMCfast

- The inclusion of hadron-collider data into fits of **parton distribution functions** (PDFs) is a crucial requirements for precision physics at the LHC.

- Presently, hadronic NLO(+PS) calculations are too **time-consuming** to be directly employed in a PDF fit.

- Need for **fast interfaces** to such calculations (*e.g.* APPLgrid).

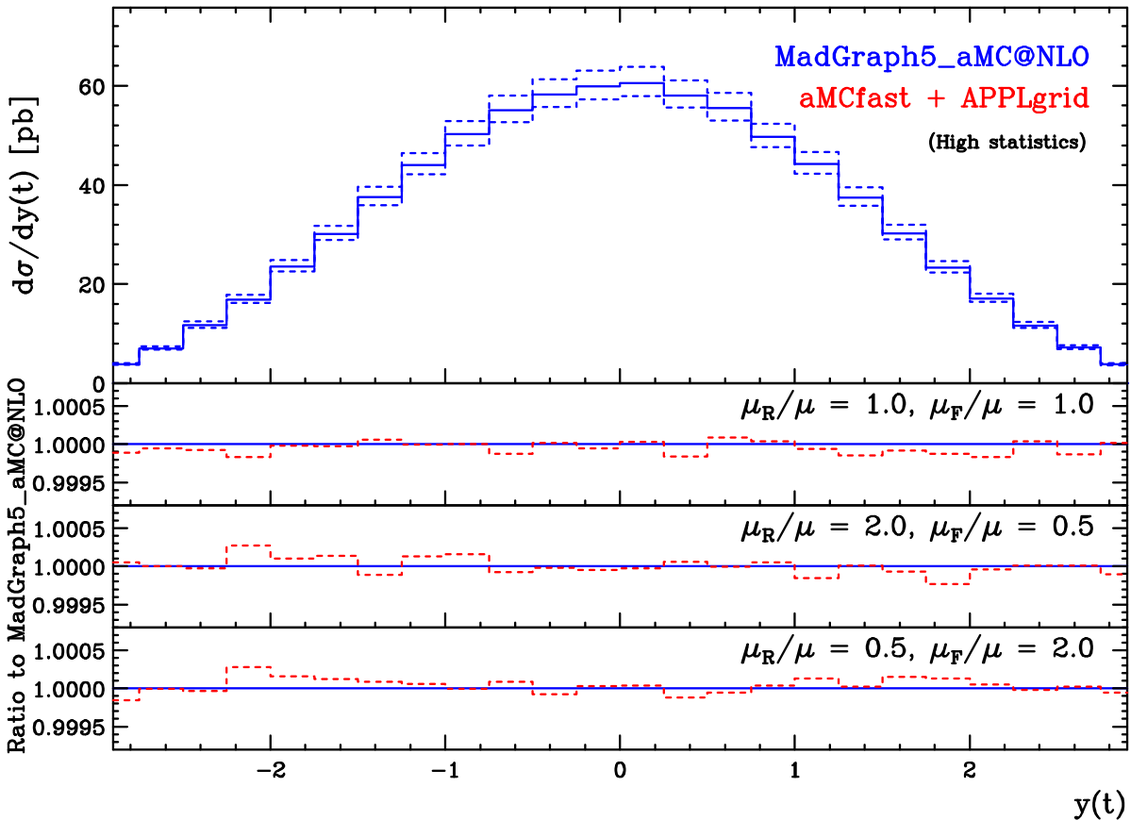
- **aMCfast** provides an interface that bridges:
  - **MadGraph5_aMC@NLO** [arXiv:1405.0301]
    - an **automated** cross section calculator of LO and NLO cross sections, with and without matching to parton showers.
  - **APPLgrid** [arXiv:0911.2985]
    - a framework for the fast computation of cross sections ⇒ APPLgrid interpolation grids.

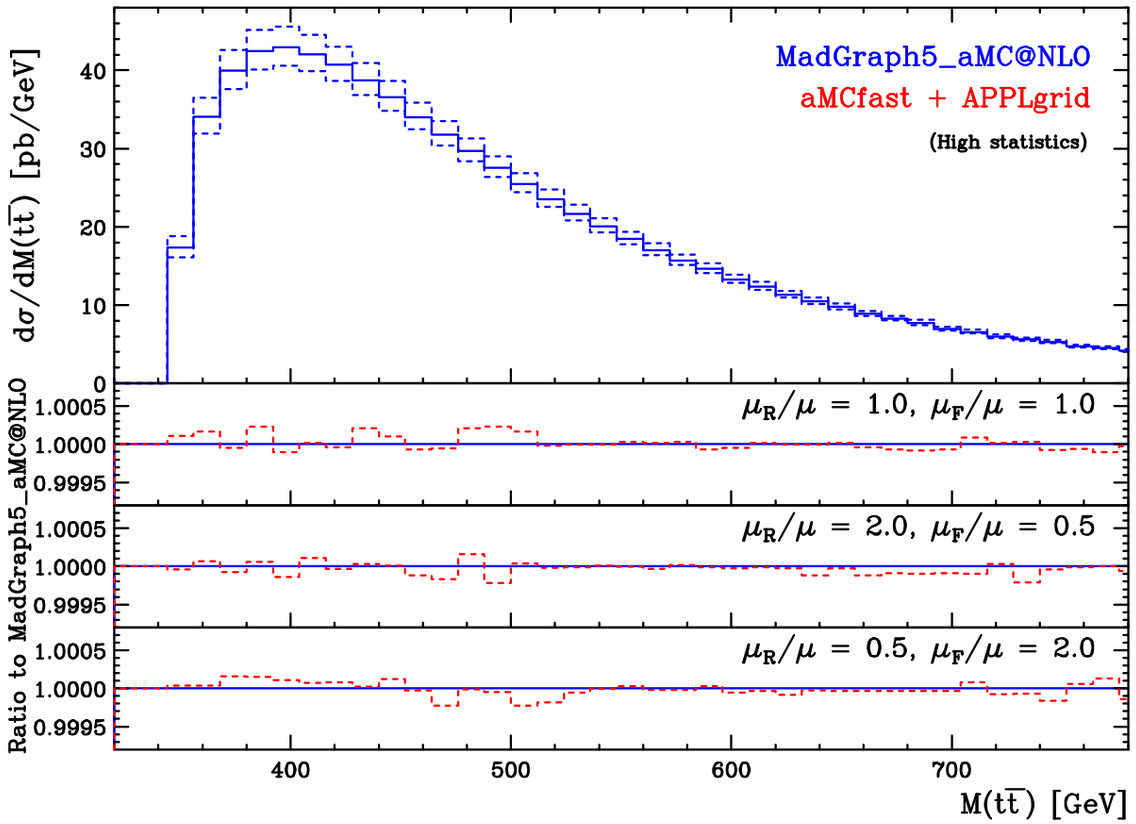
- The **full automation** of MG5_aMC@NLO allows one to produce interpolation grids for any process without any additional work (like *e.g.* for MCFM).

The aMCfast Interface

The NLO case

- At fixed order the the fast interpolation technique is **well-established** and presents no conceptual problems:
  - exact **independence** of the interpolation grids from PDFs and α_s,
  - commonly used in most of the modern PDF fits.
- The aMCfast interface for NLO computation provides a high accuracy:





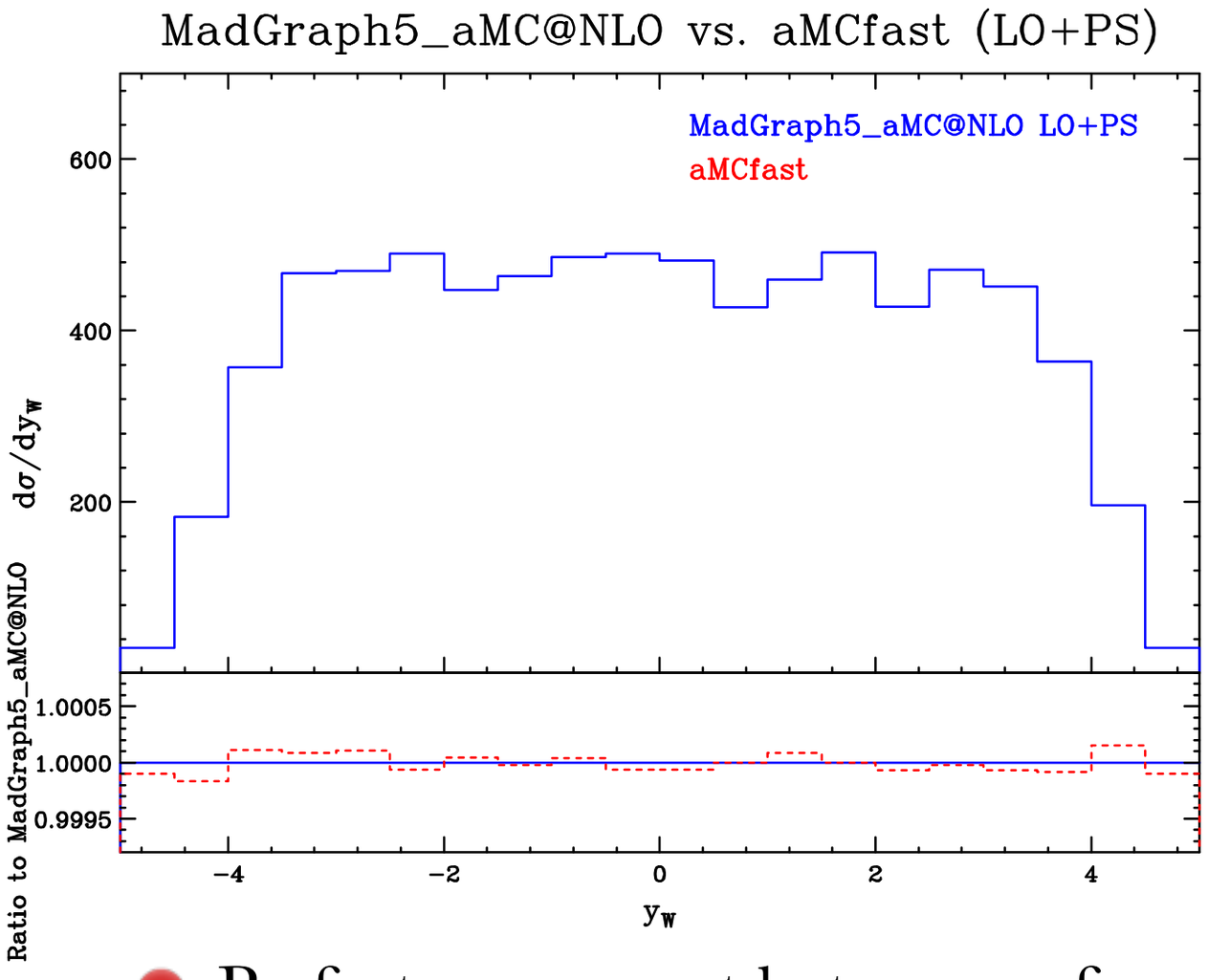
- and has been already employed in e few public studies.

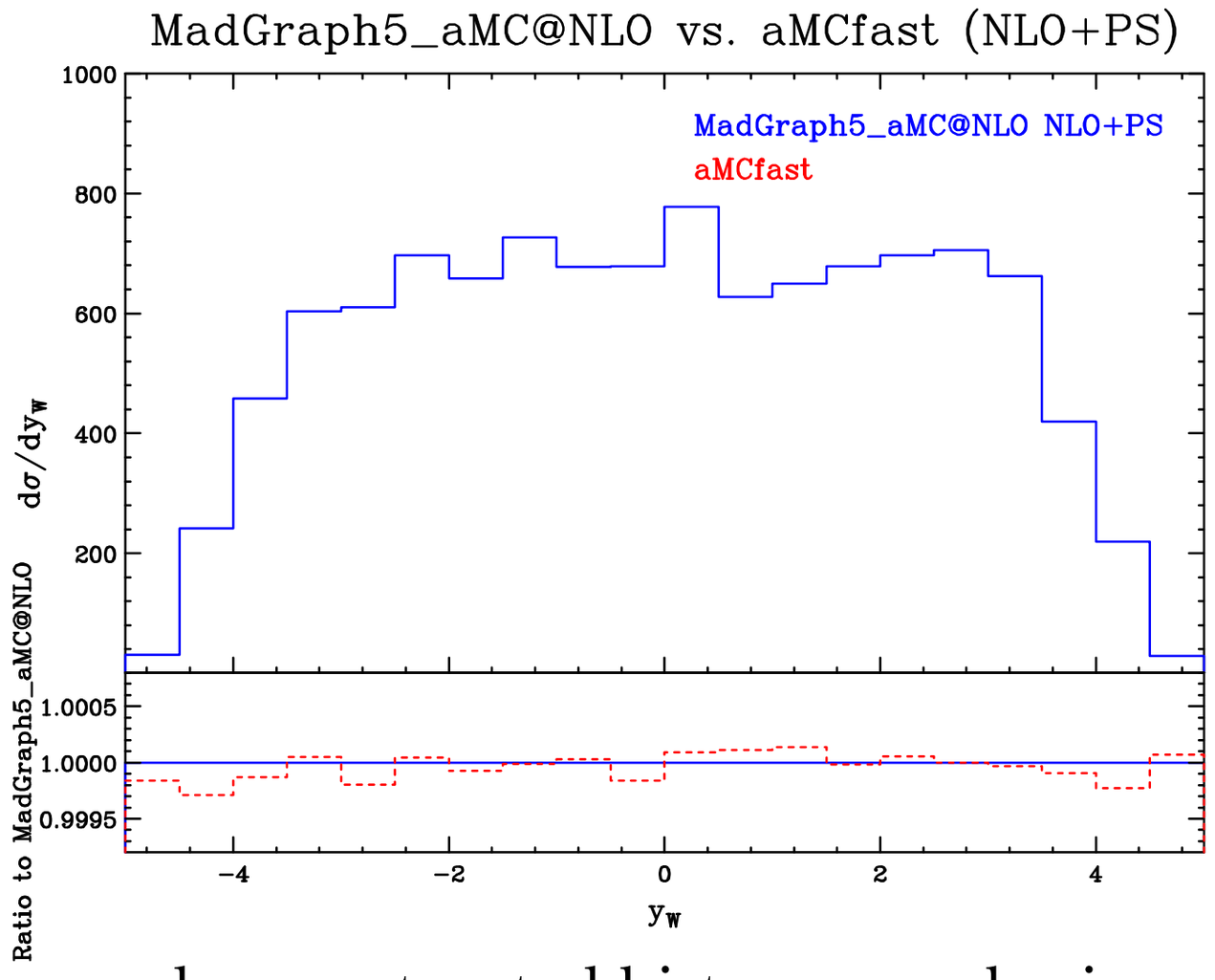
The aMCfast Interface

The NLO + PS Case

- Technically speaking, the aMCfast interface at NLO+PS mode of has already been implemented and tested for **all** PSs interfaced to MG5_aMC@NLO.

- W rapidity for $e^+\nu$ production at LO and NLO + Herwig6:





- Perfect agreement between reference and reconstructed histograms also in the low statistics regime, as in the fixed-order case.

The aMCfast Interface

The NLO + PS Case

- The production of interpolation grids in the presence of PS poses more **conceptual questions** as compared to the fixed-order case.

- There are **two main issues**:

  1) Dependence on PDFs of the **PS evolution** as a results of different kinematic configurations at the **matrix element** (ME) level when the latter is computed with different PDF sets cannot be removed.

  2) Dependence on PDFs of the **initial-state PS** (IS-PS) cannot be disentangled:

     - expected to be small as it appears as a ratio of PDFs at the same x but different Q^2.

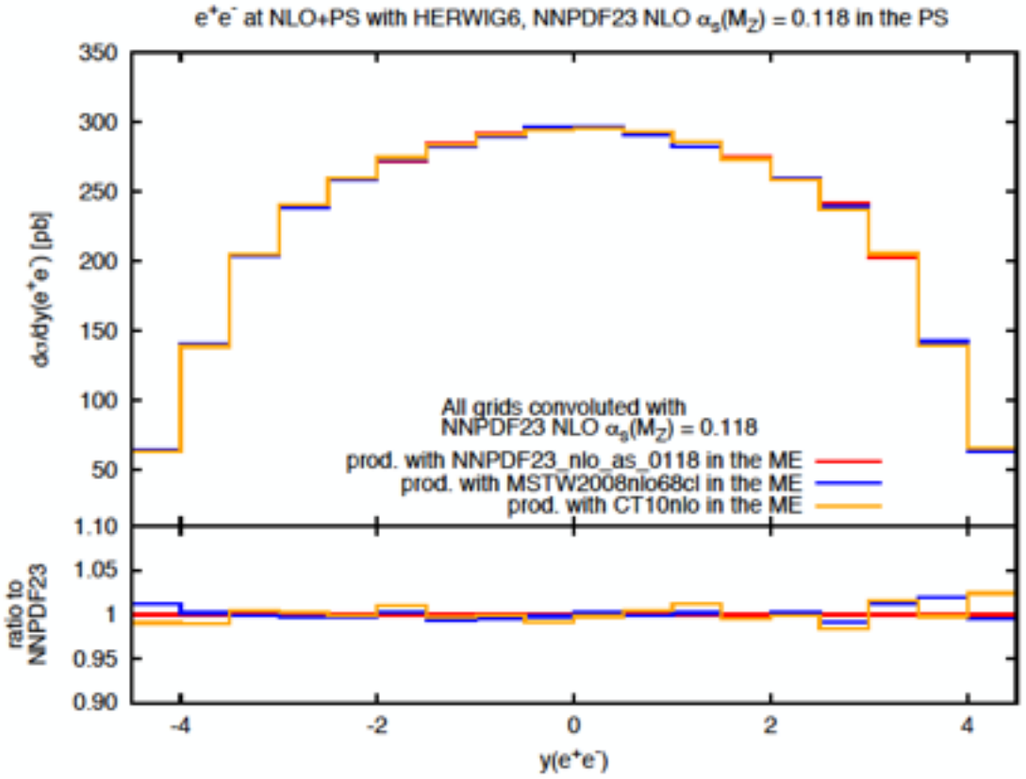
- In order for such grids to be safely used in a PDF fit one needs to check **explicitly** that the dependence on the PDFs used for the production is mild:

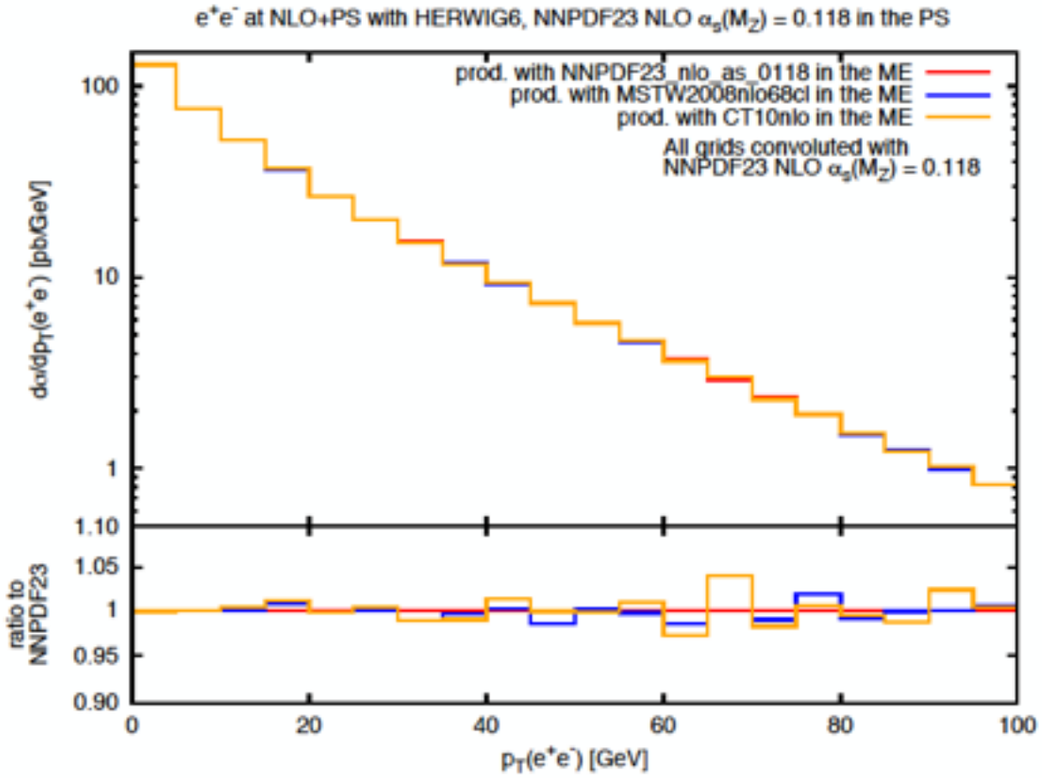
  - process/observable dependent,

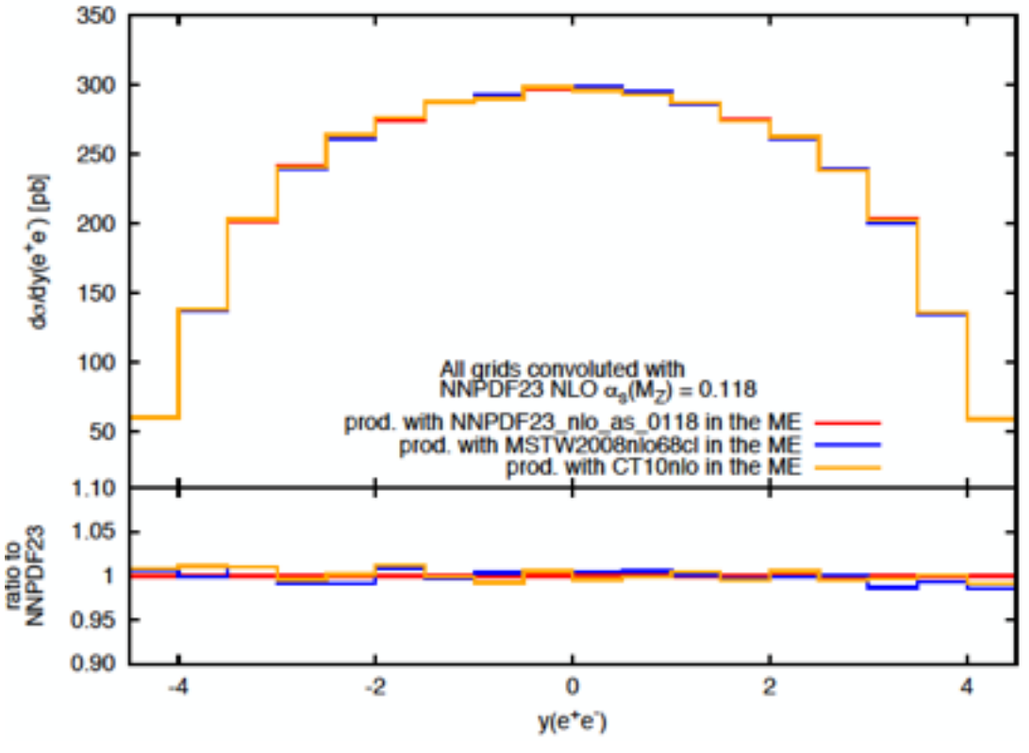
  - parton shower dependent.

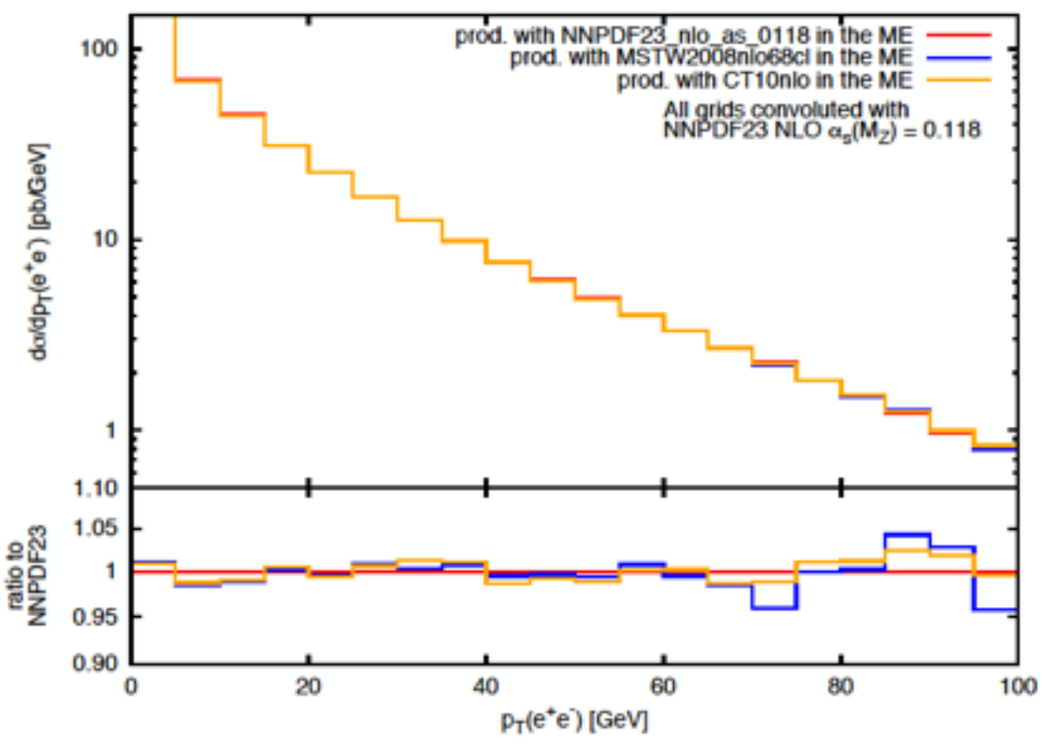
The aMCfast Interface

The NLO + PS Case: Dependence on the ME



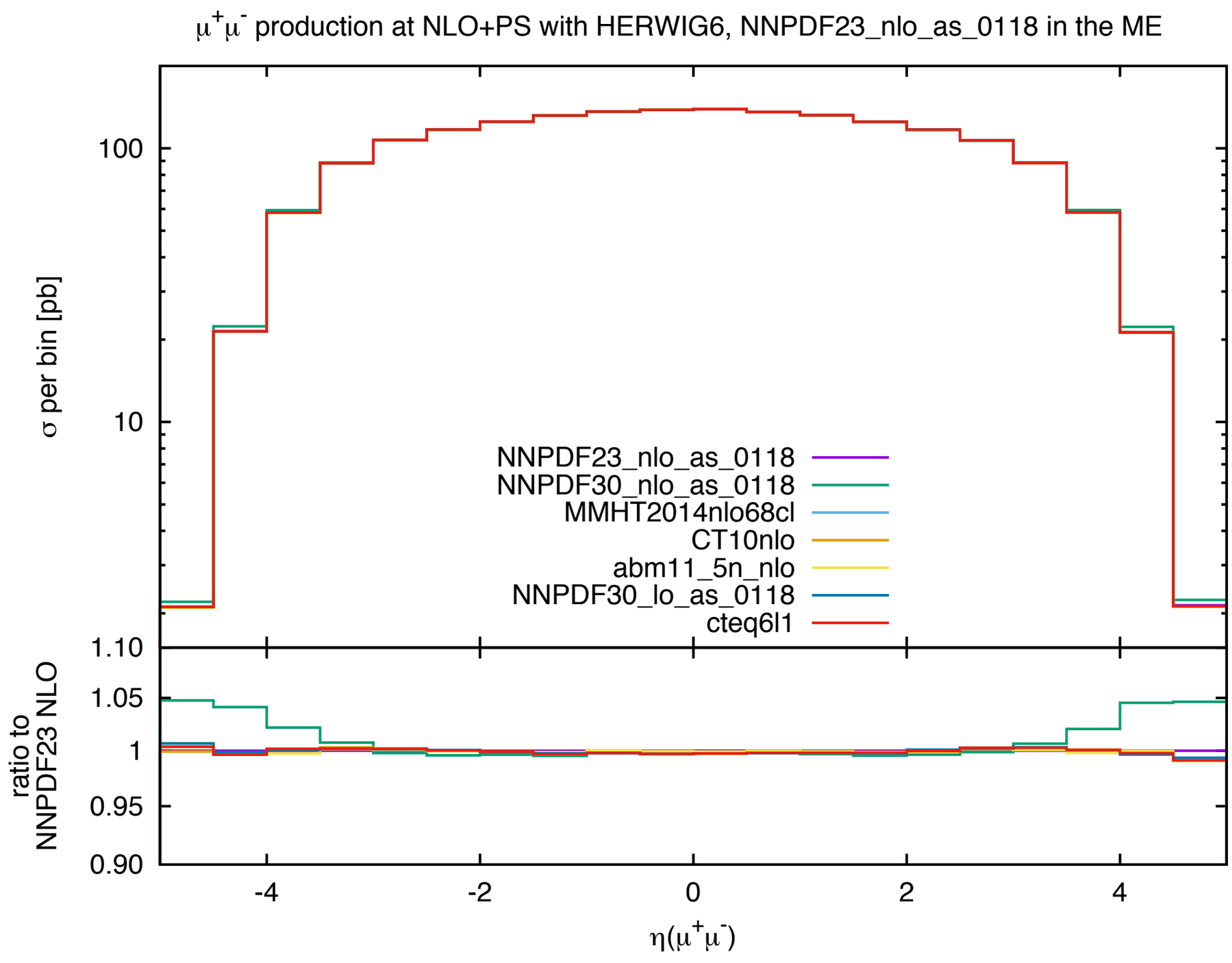






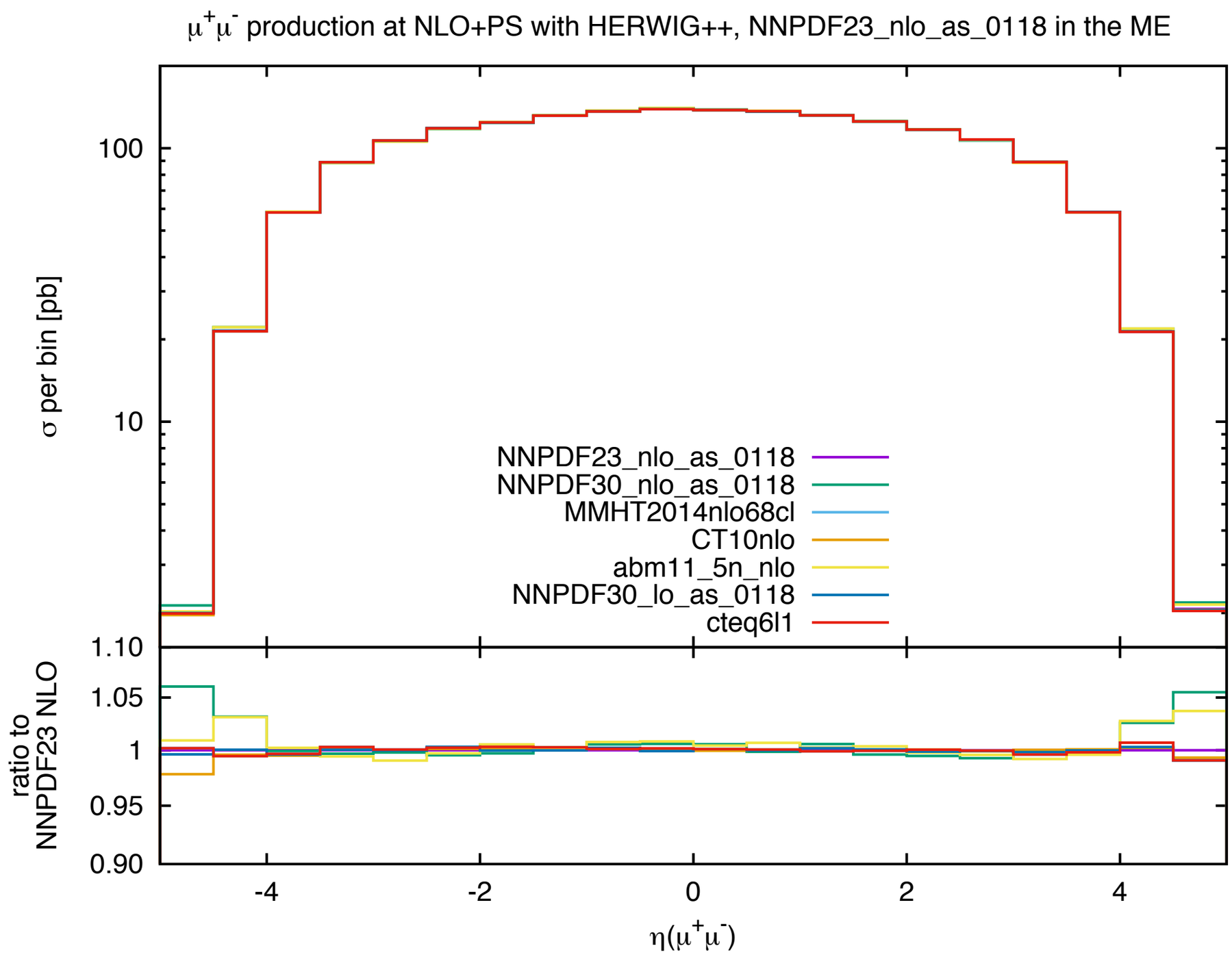
The aMCfast Interface

The NLO + PS Case: Dependence on the IS-PS



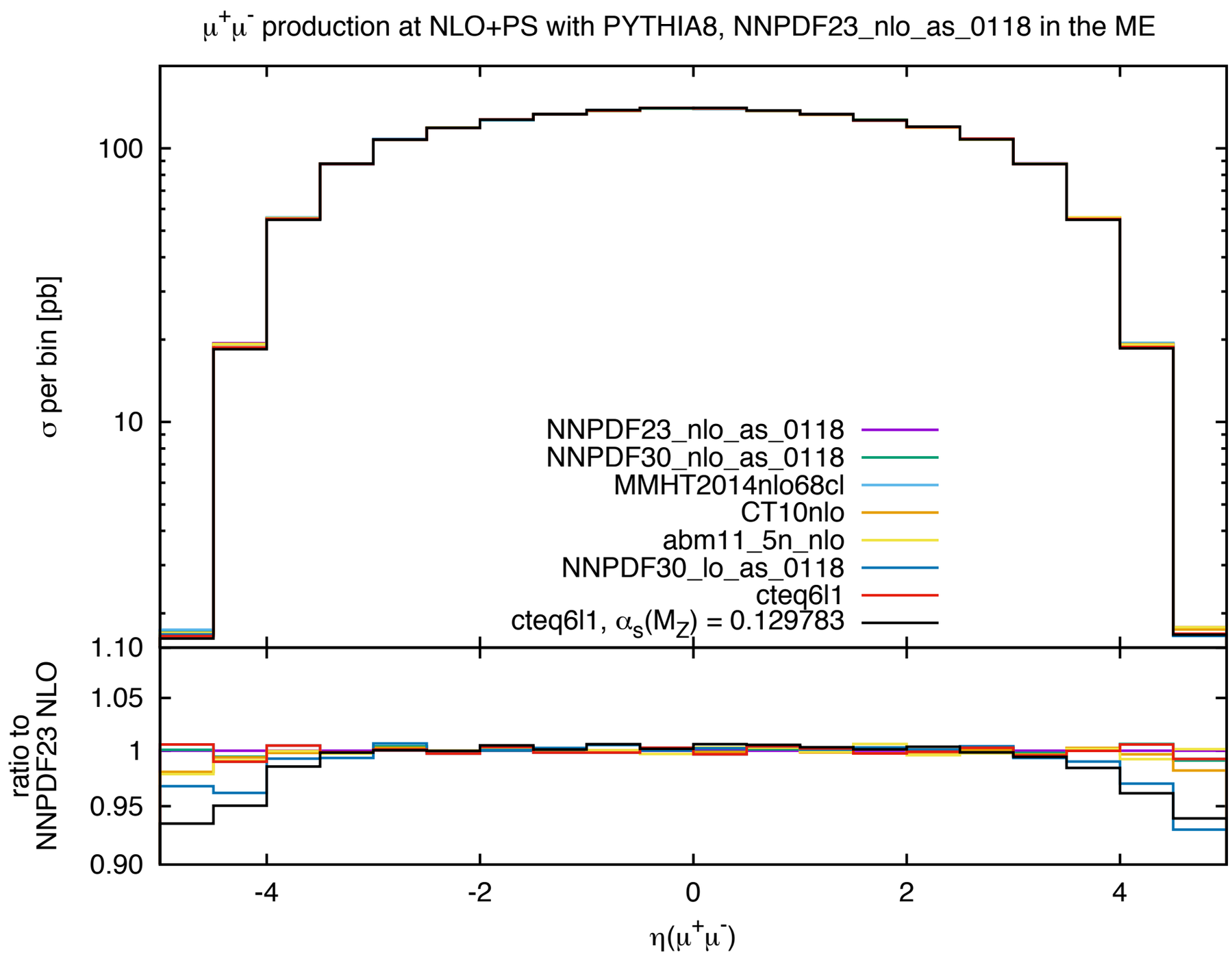
The aMCfast Interface

The NLO + PS Case: Dependence on the IS-PS



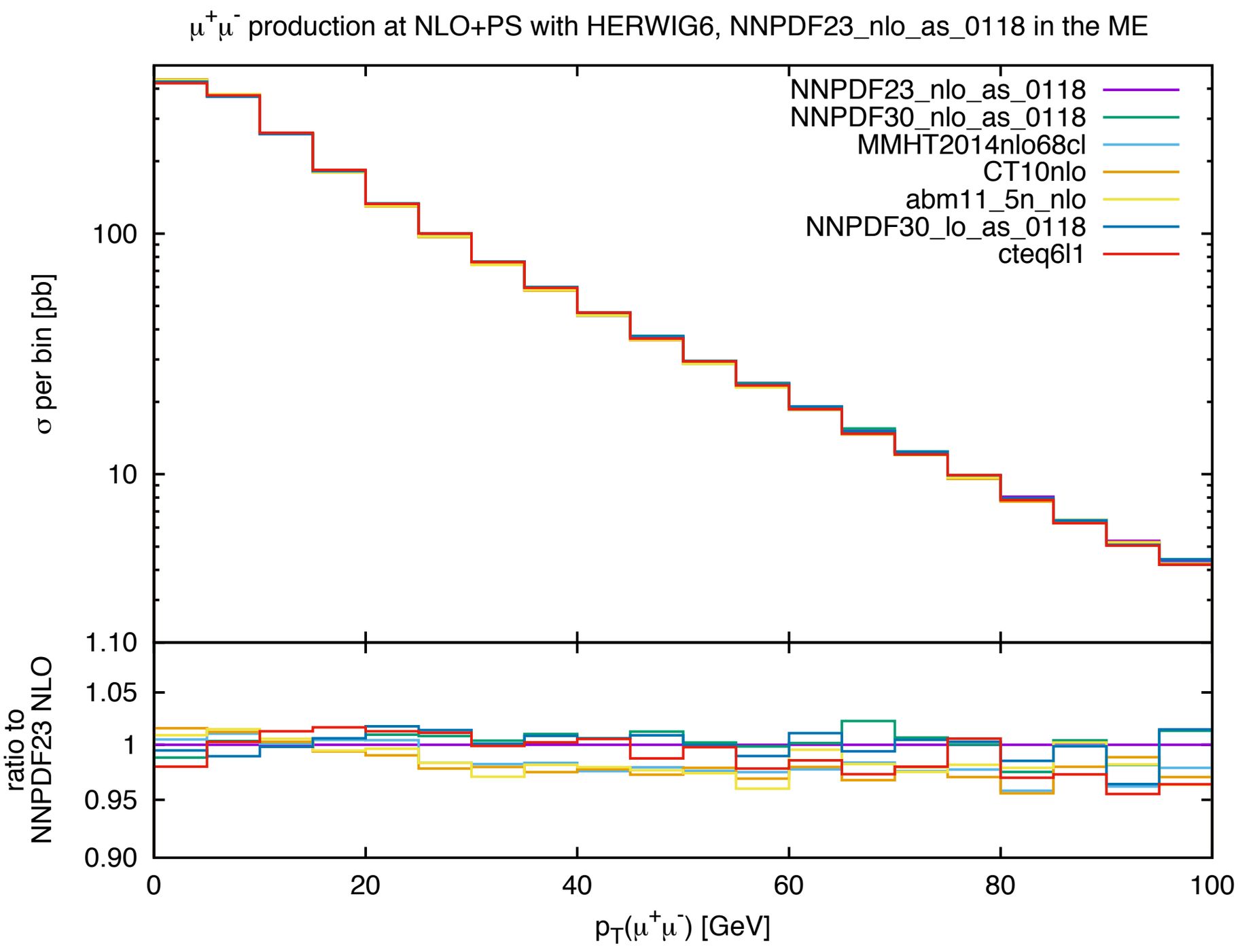
The aMCfast Interface

The NLO + PS Case: Dependence on the IS-PS



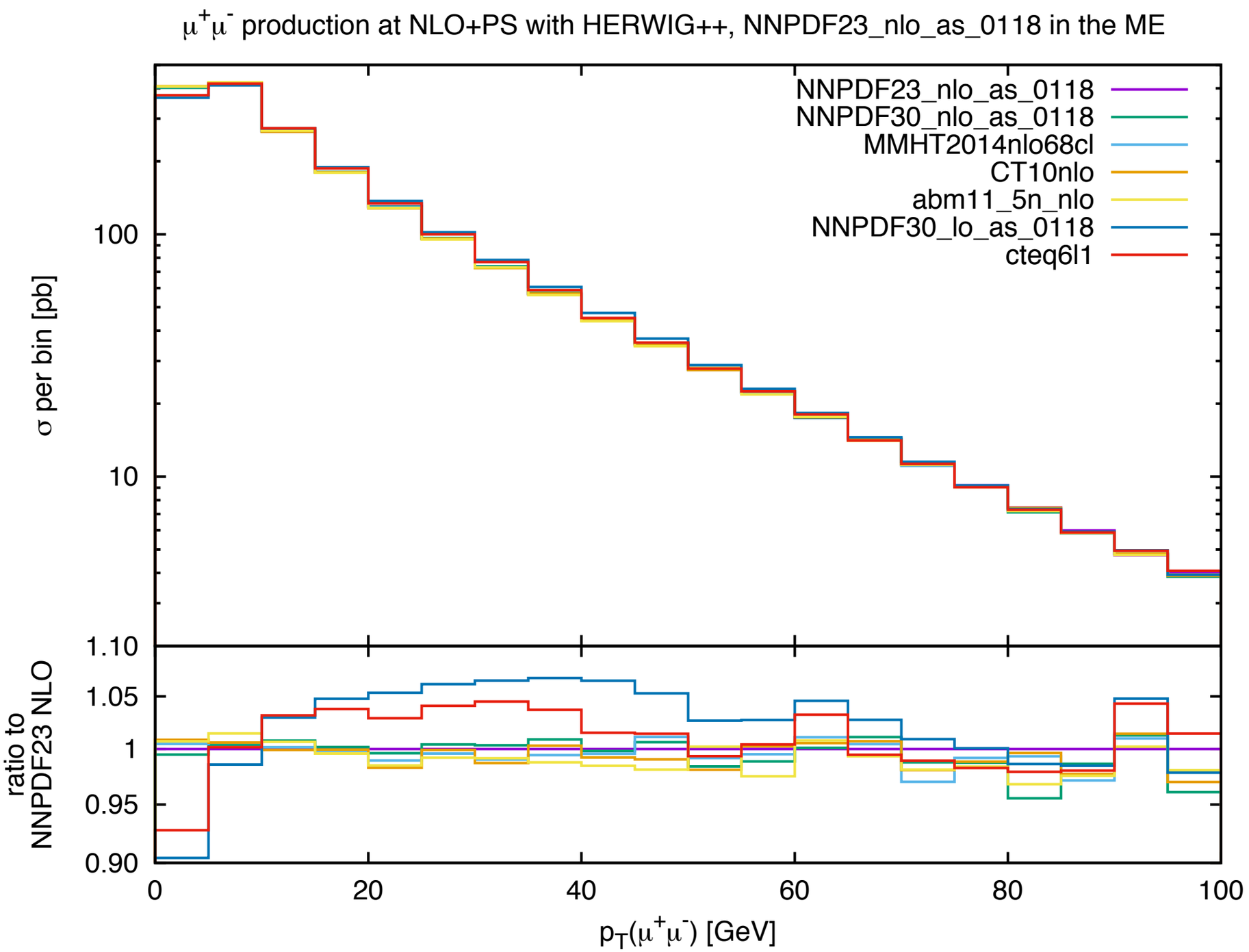
The aMCfast Interface

The NLO + PS Case: Dependence on the IS-PS



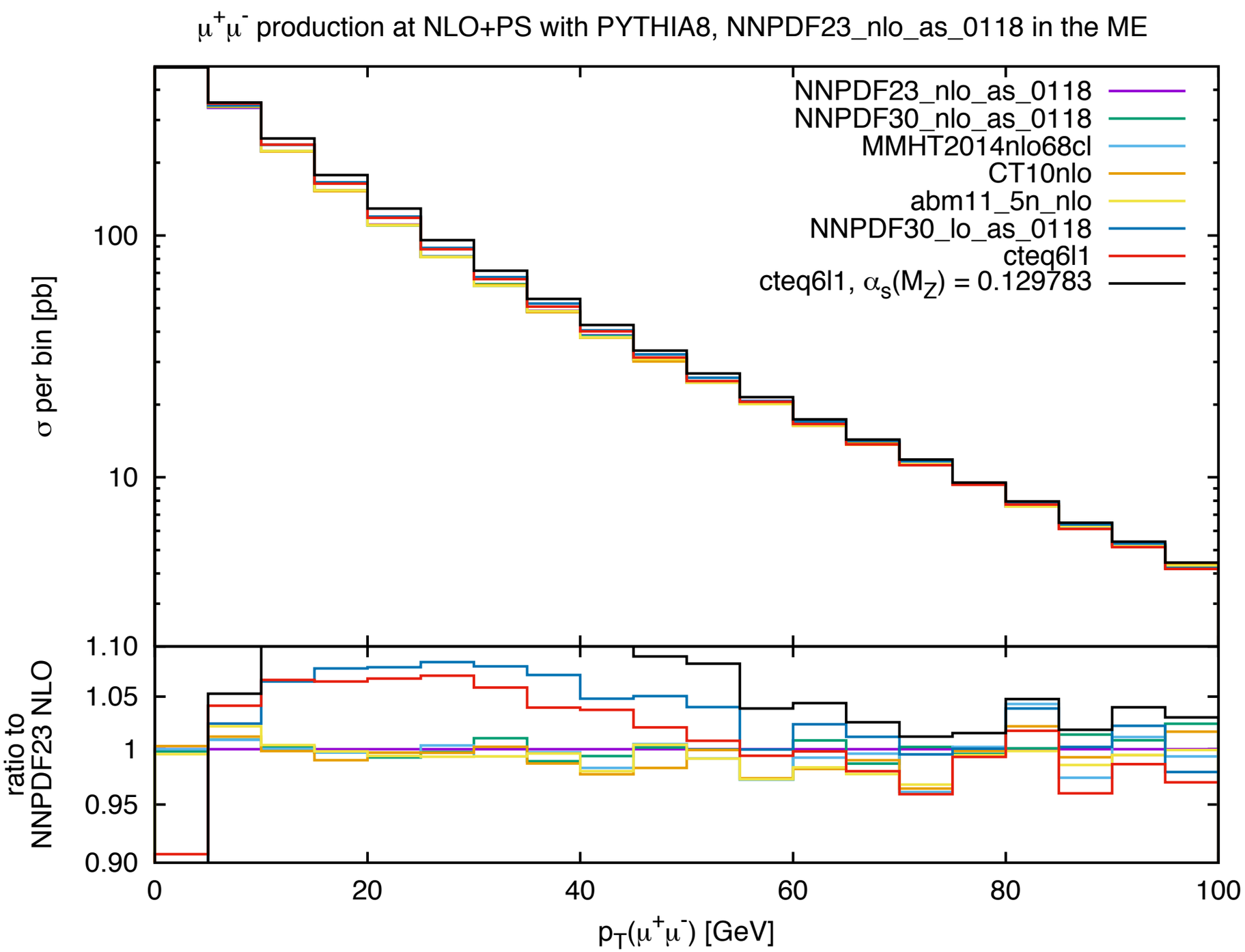
The aMCfast Interface

The NLO + PS Case: Dependence on the IS-PS



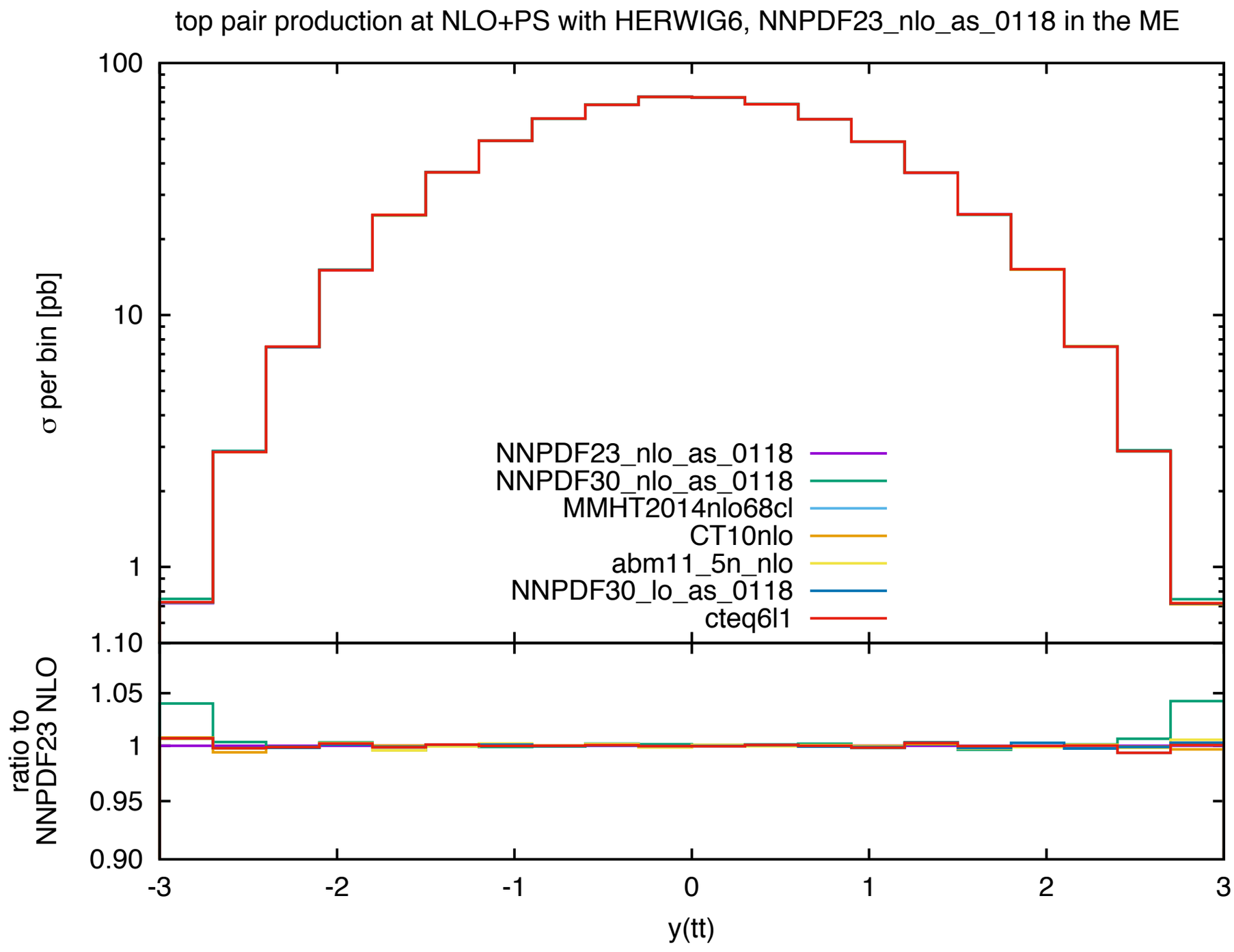
The aMCfast Interface

The NLO + PS Case: Dependence on the IS-PS



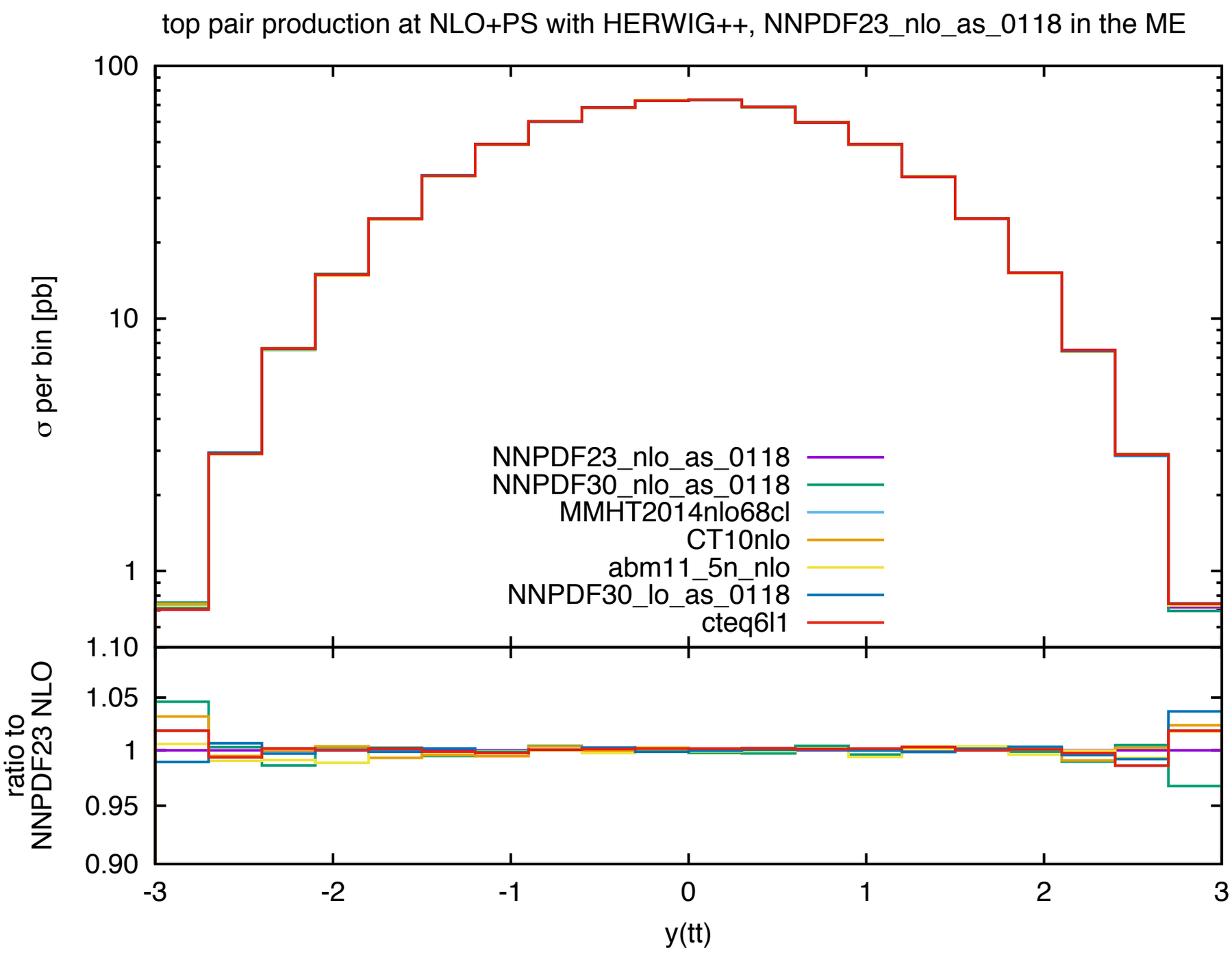
The aMCfast Interface

The NLO + PS Case: Dependence on the IS-PS



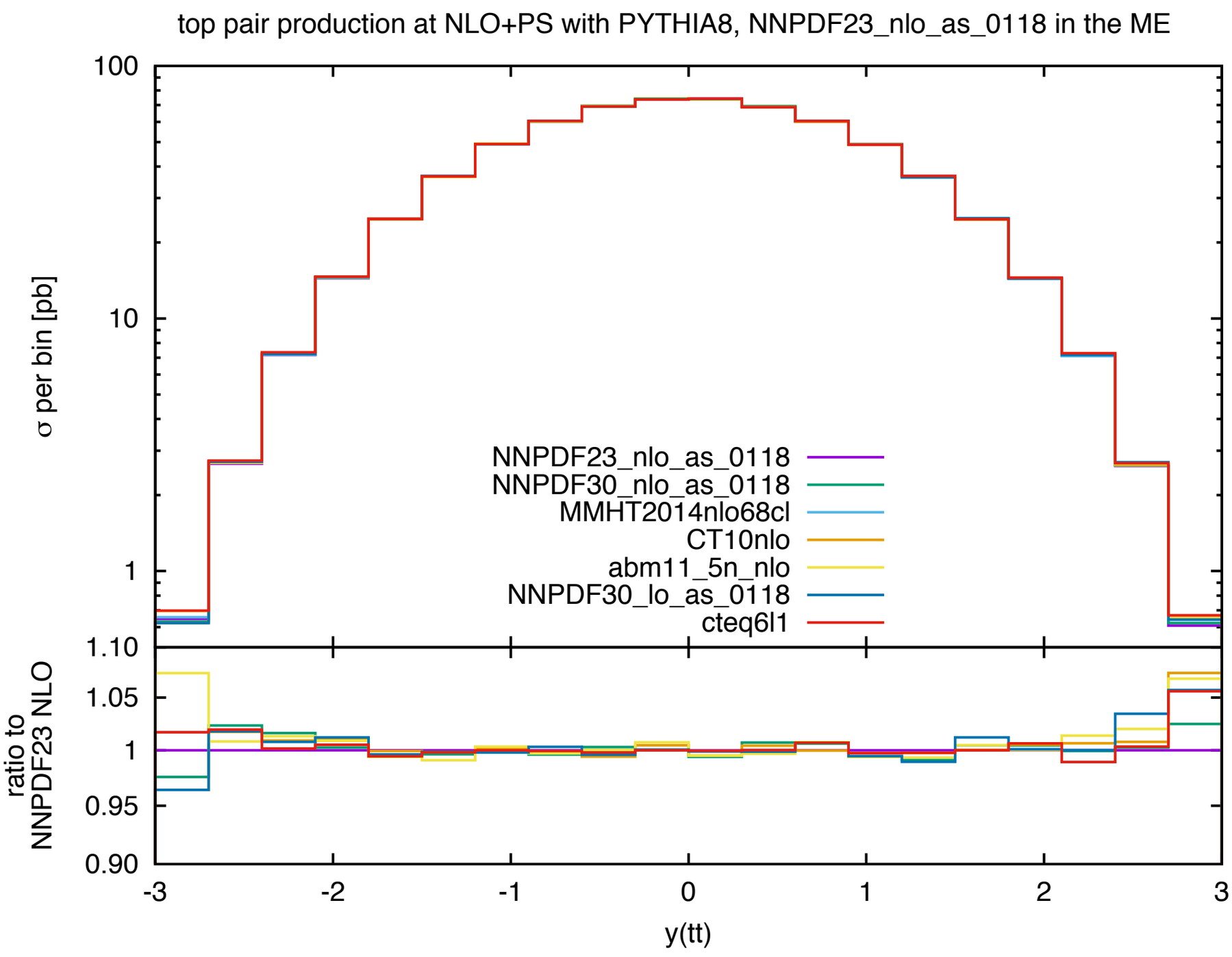
The aMCfast Interface

The NLO + PS Case: Dependence on the IS-PS



The aMCfast Interface

The NLO + PS Case: Dependence on the IS-PS



The aMCfast Interface

The NLO + PS Case

- In general, the **dependence** on the PDF set used in the PS is **contained**:
  - provided that "reasonable" PDFs are used,
  - using LO or NLO PDFs can make a substantial difference.
- However, there are corners (*e.g.* **large rapidity regions**) where such dependence is more marked.
- In addition, in those corners different PSs might behave differently.
- Also, less inclusive observables present a stronger dependence (merging should help).
- In conclusion, even though deviations are typically contained, there is **no guarantee** that the procedure is generally unbiased and thus formally suitable for PDF fits. However, how relevant is the bias?
  - To answer this question we need to estimate the **potential impact** of the PS corrections on PDFs.

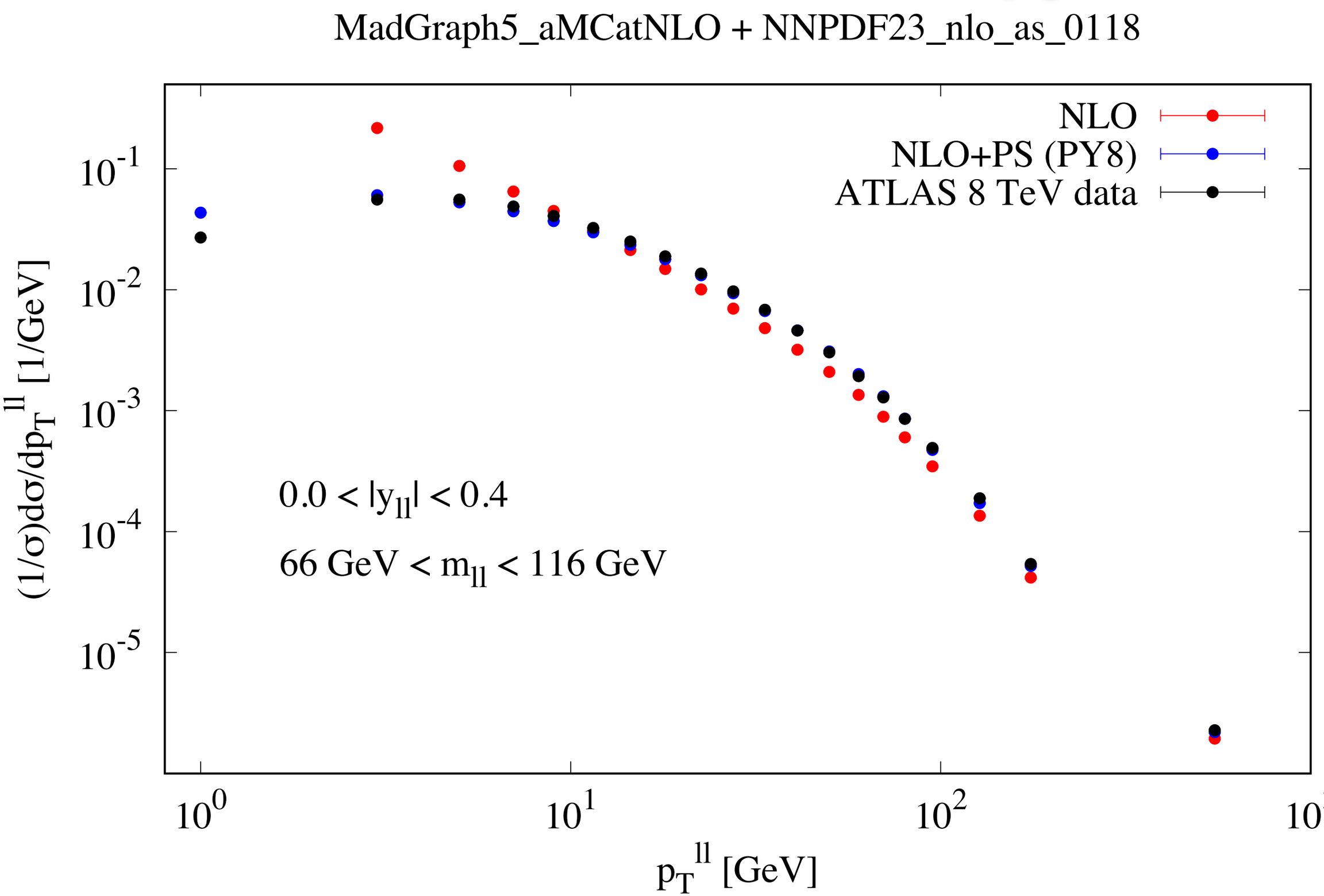
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data

- As first step we considered the ATLAS **Z p_T distribution** data at 8 TeV in the lepton-pair invariant mass bin 66 GeV < m_{ll} < 116 GeV:
  - 6 rapidity bins in the range 0 < $|y_{ll}|$ < 2.4 with 20 data points each.
  - Extremely **small uncertainties**.
- Data points in each rapidity bin extend between 0 and 900 GeV in p_T:
  - when fitting PDFs at fixed order low p_T data must be excluded, typically **p_T > 25 GeV**
  - in addition, since the p_T at LO is identically zero, one needs to supplement the NLO predictions with NNLO k-factors (or alternatively consider Z+j).
  - PS is expected to help in describing low p_T data and thus the cut can be lowered accordingly.

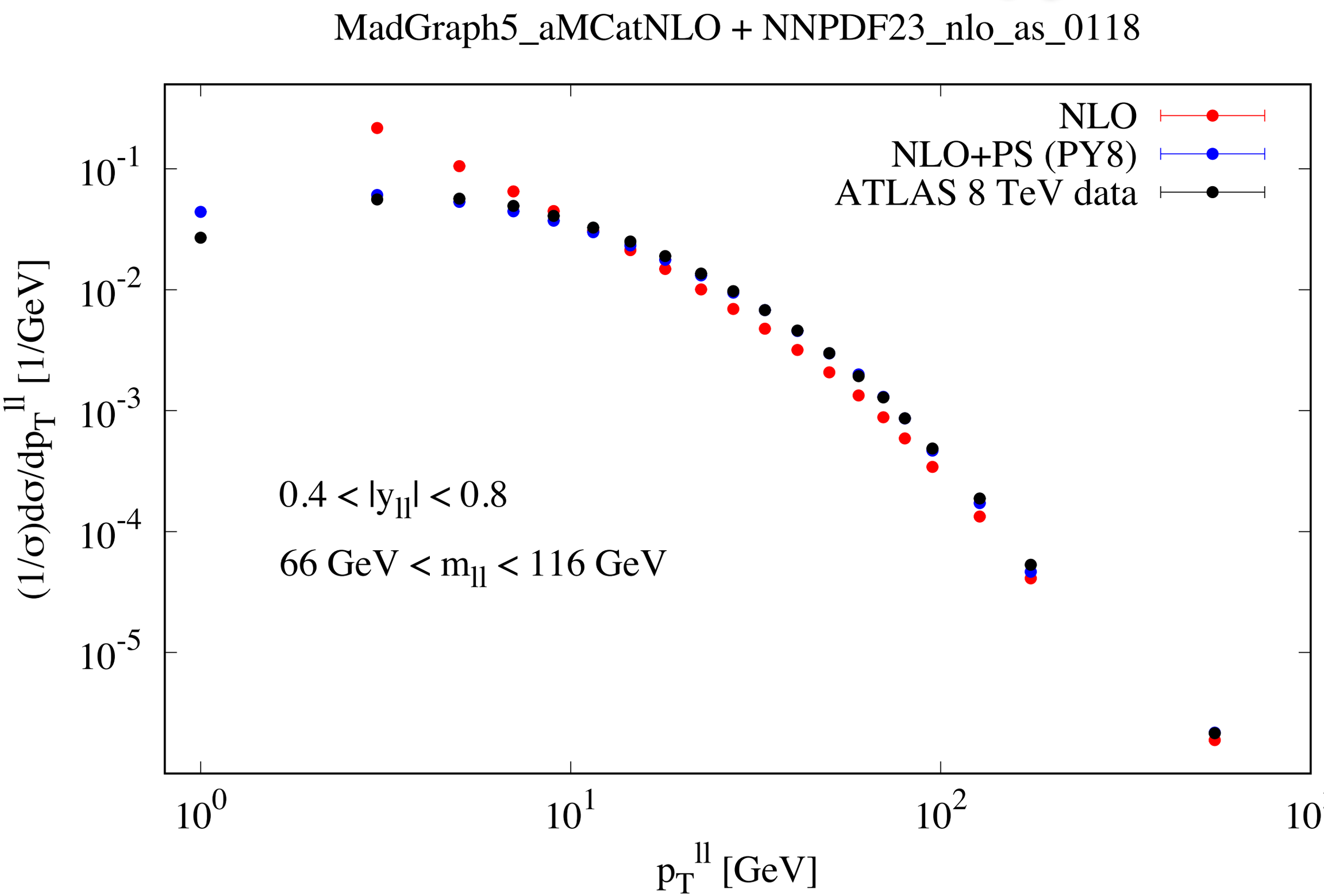
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



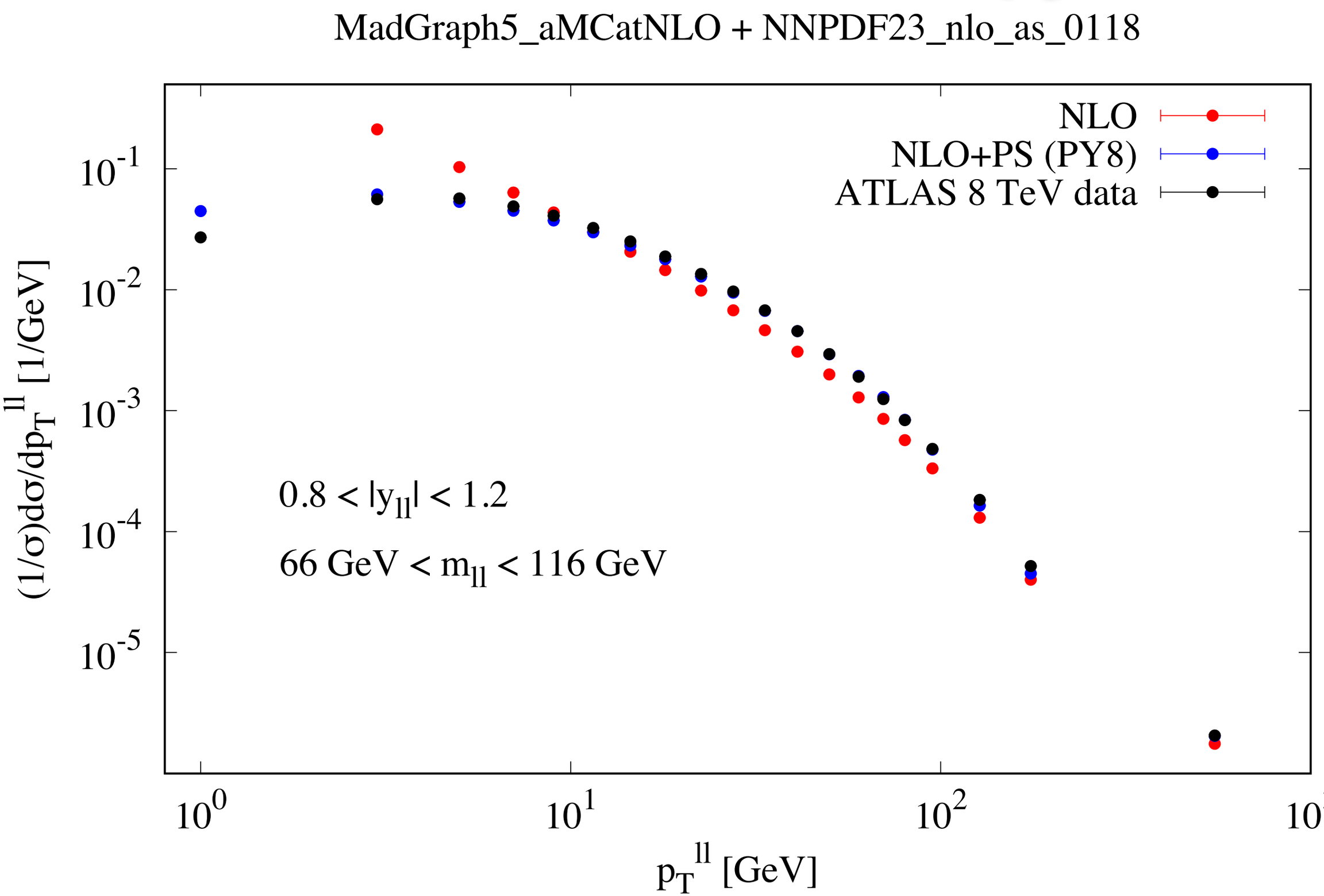
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



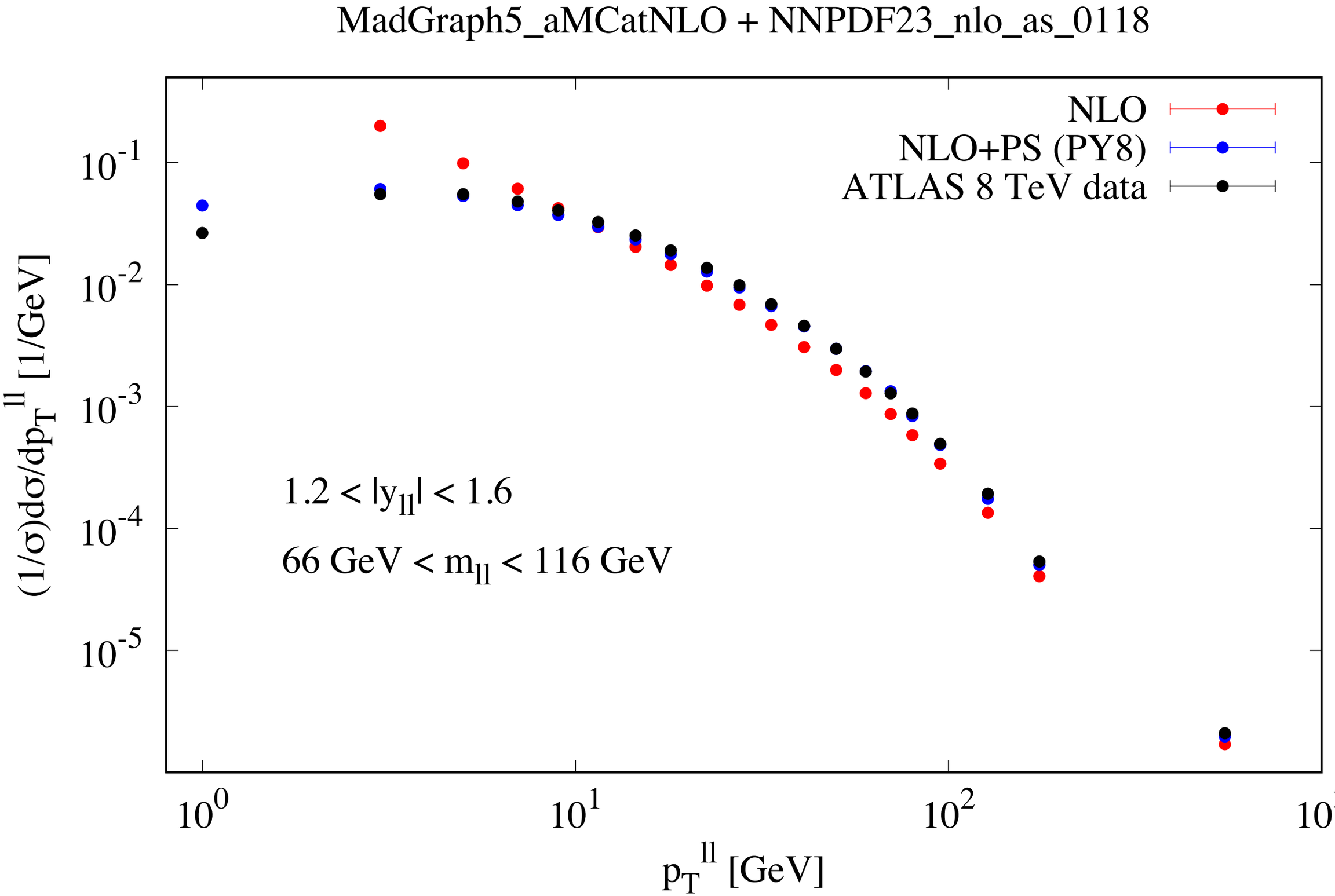
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



The aMCfast Interface

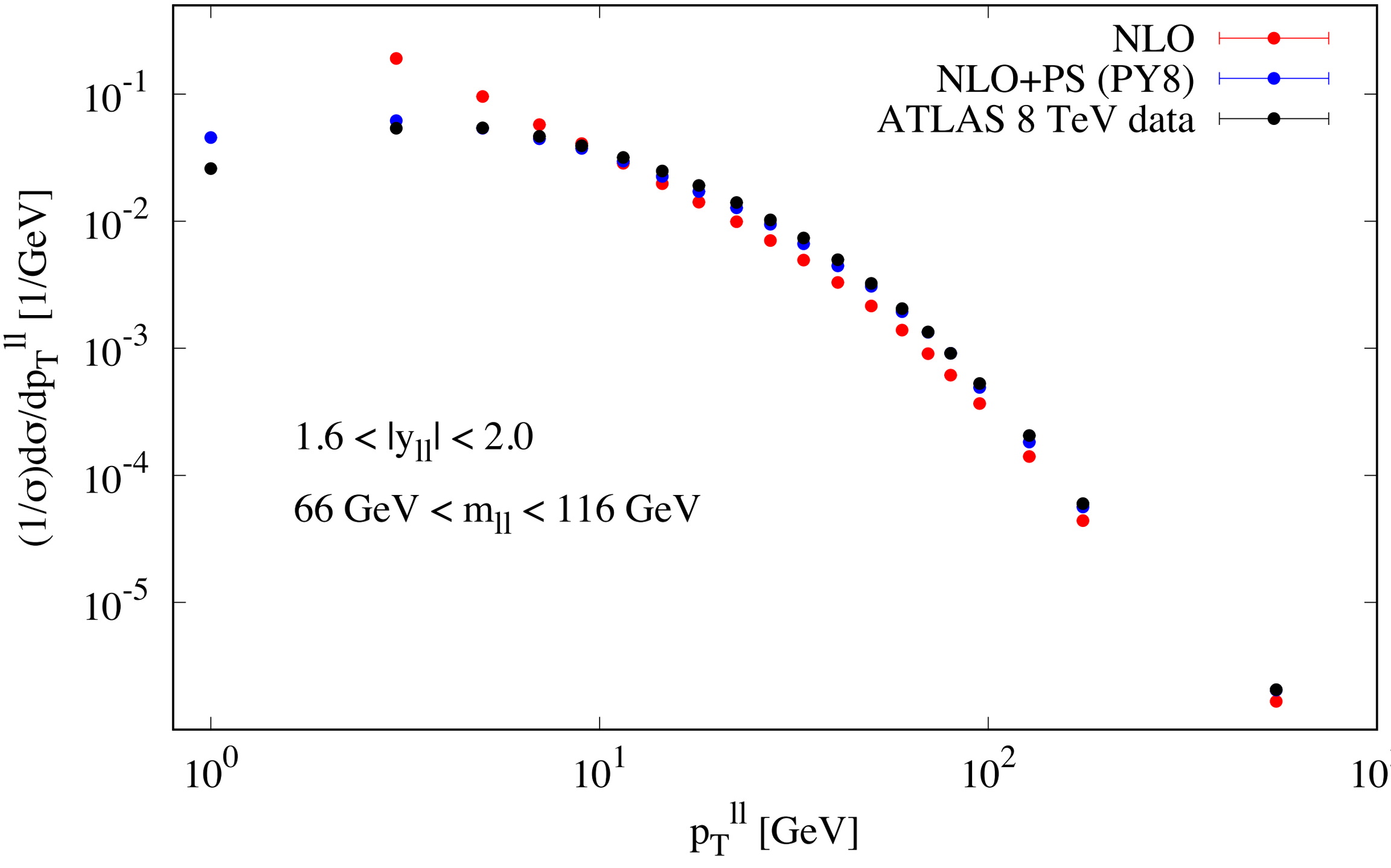
The NLO + PS Case: The ATLAS Z p_T data



The aMCfast Interface

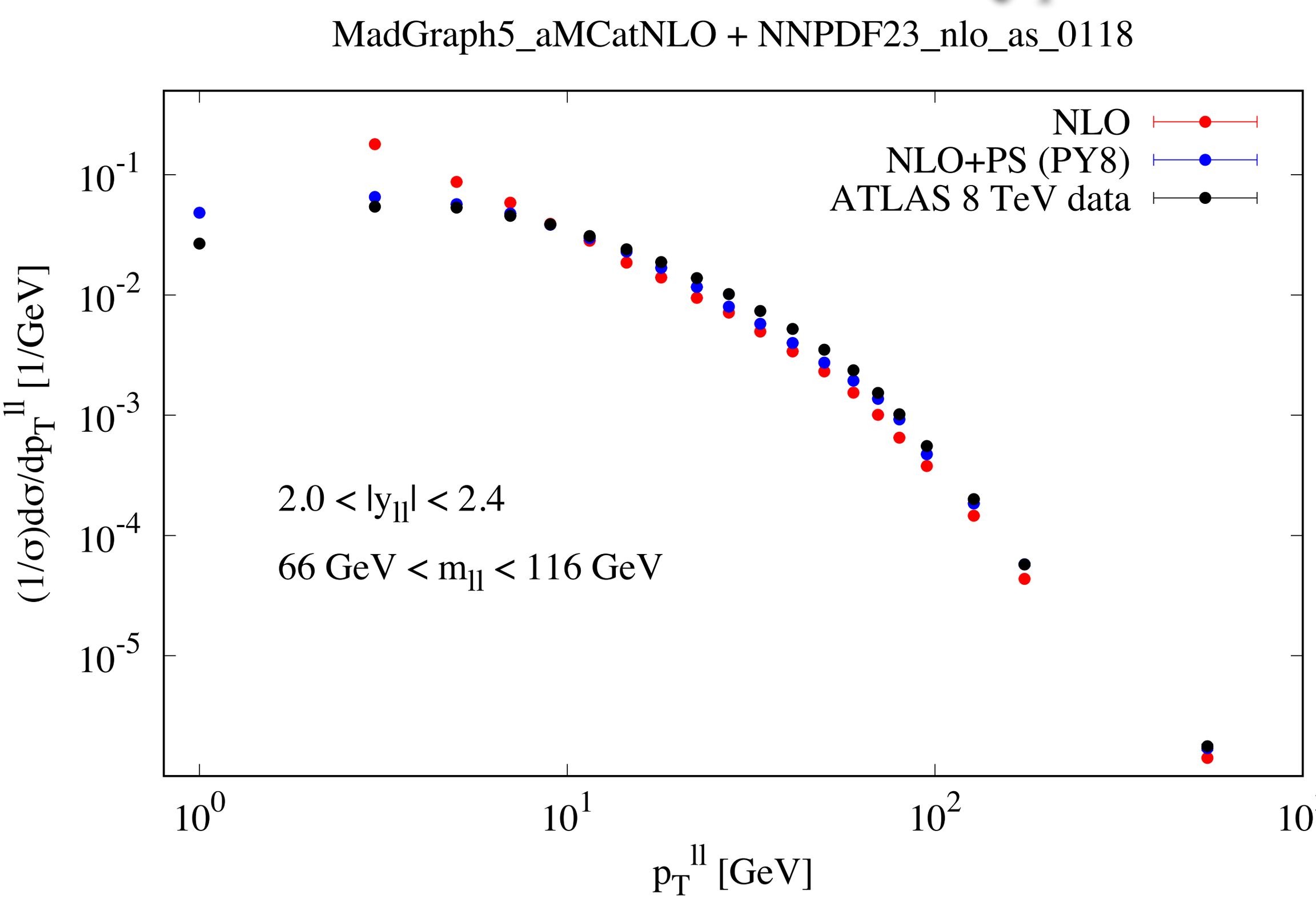
The NLO + PS Case: The ATLAS Z p_T data

MadGraph5_aMCatNLO + NNPDF23_nlo_as_0118



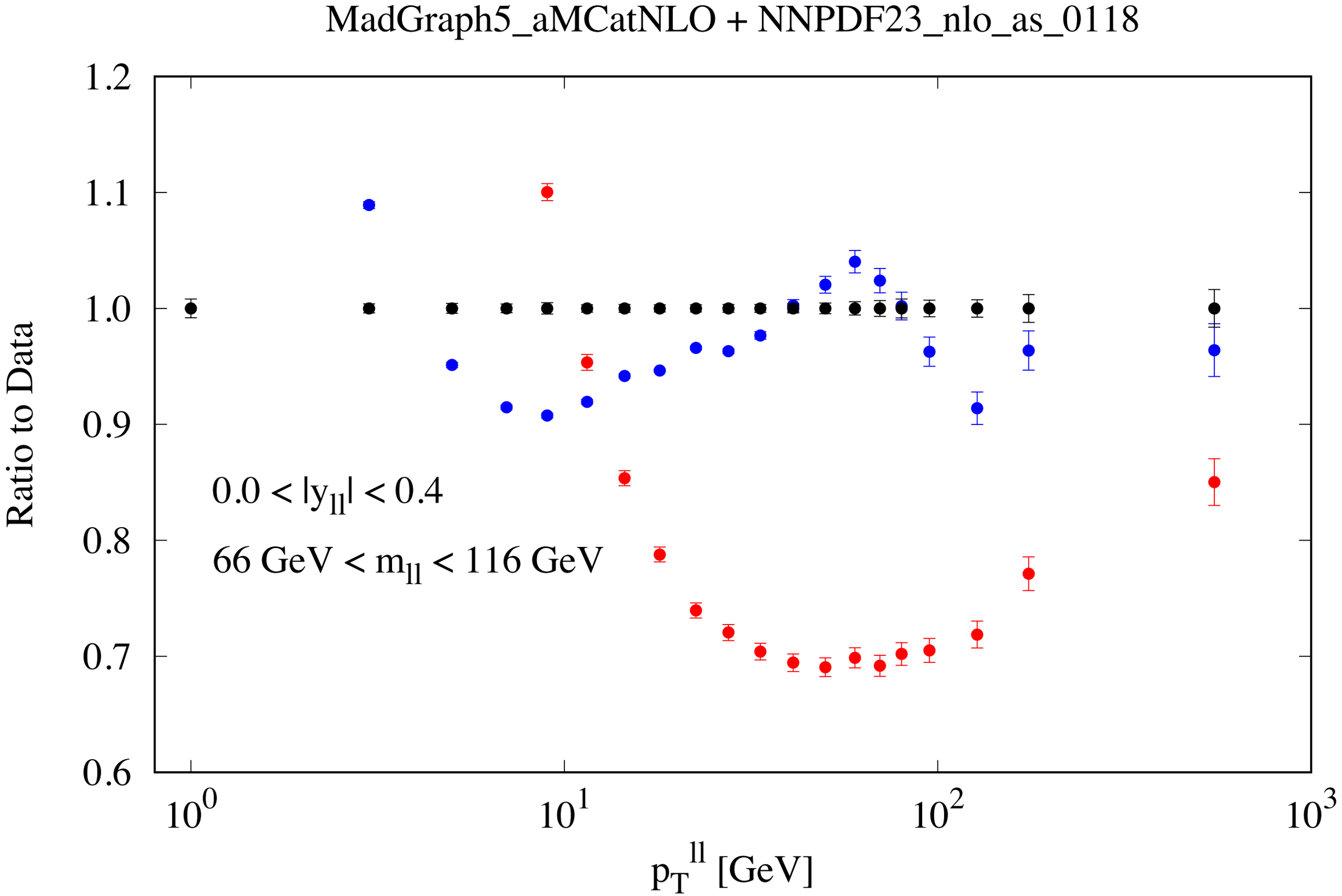
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



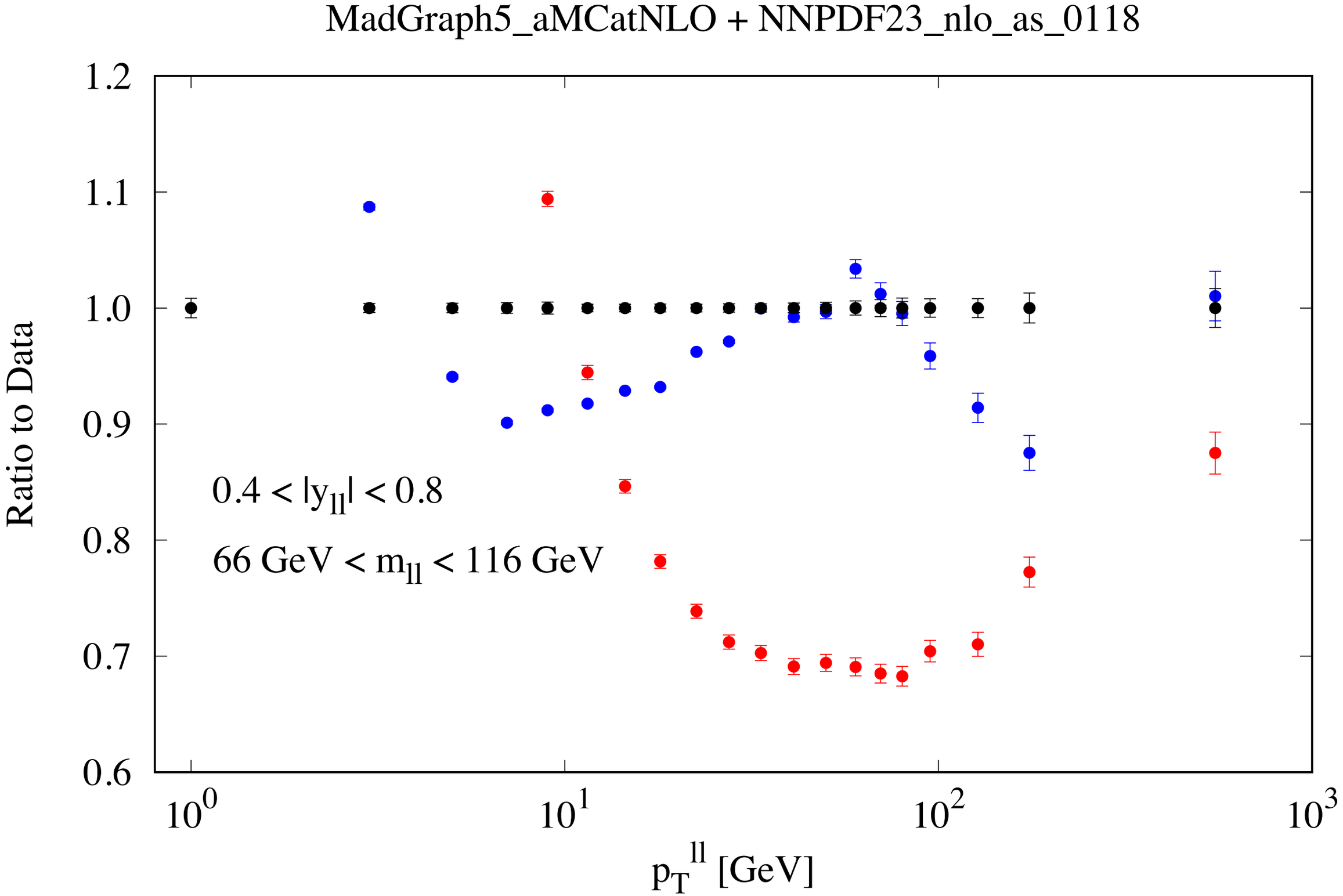
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



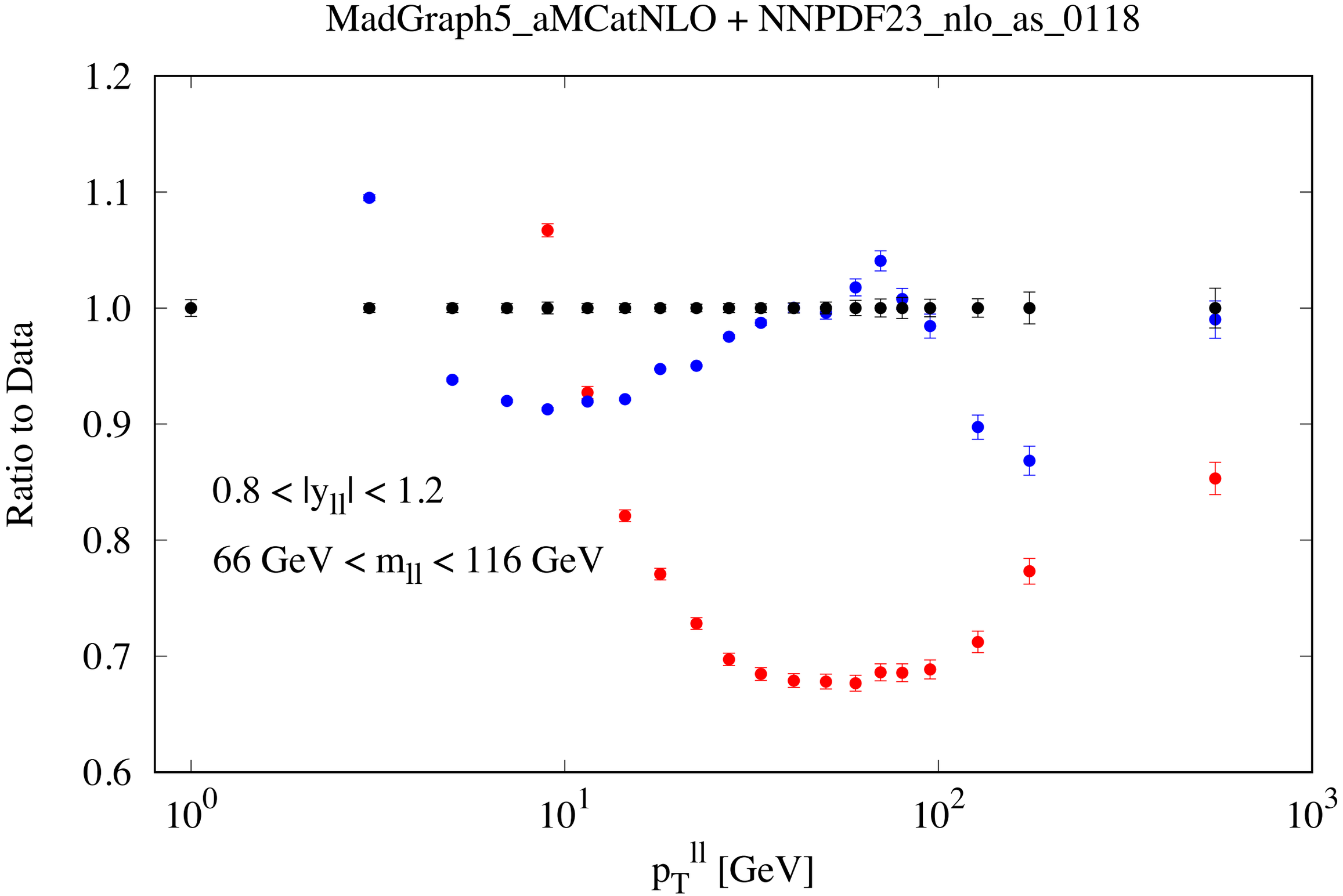
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



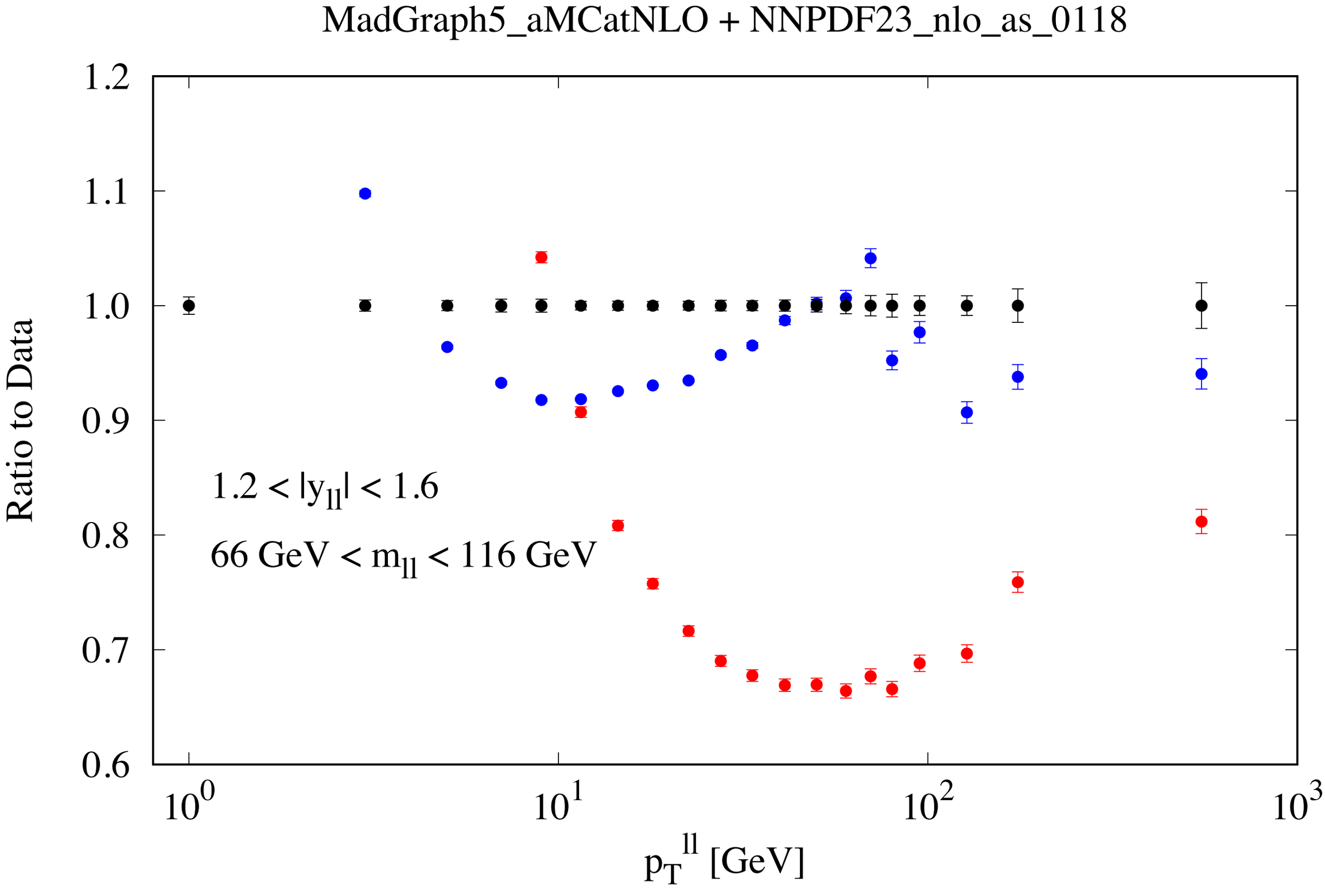
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



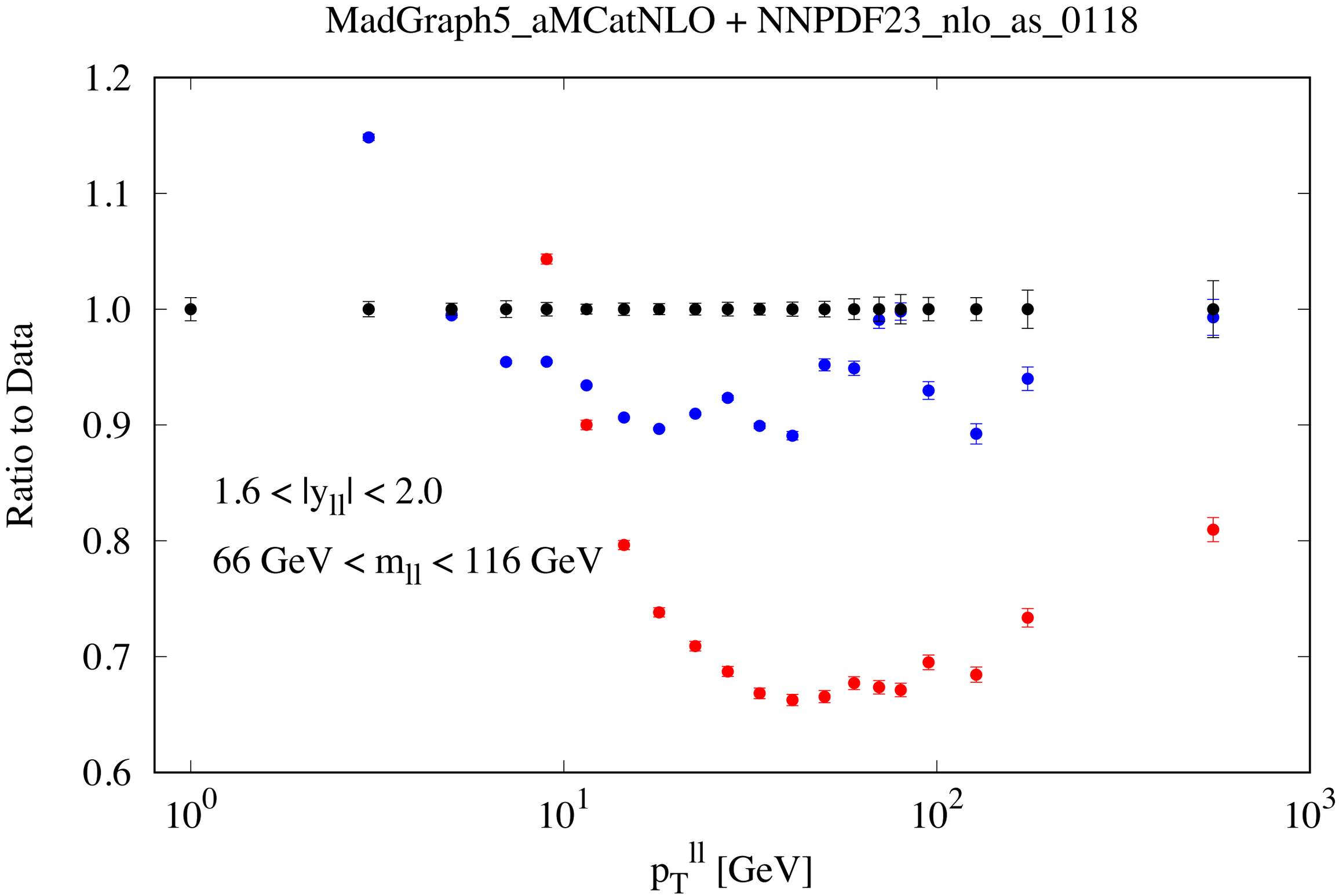
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



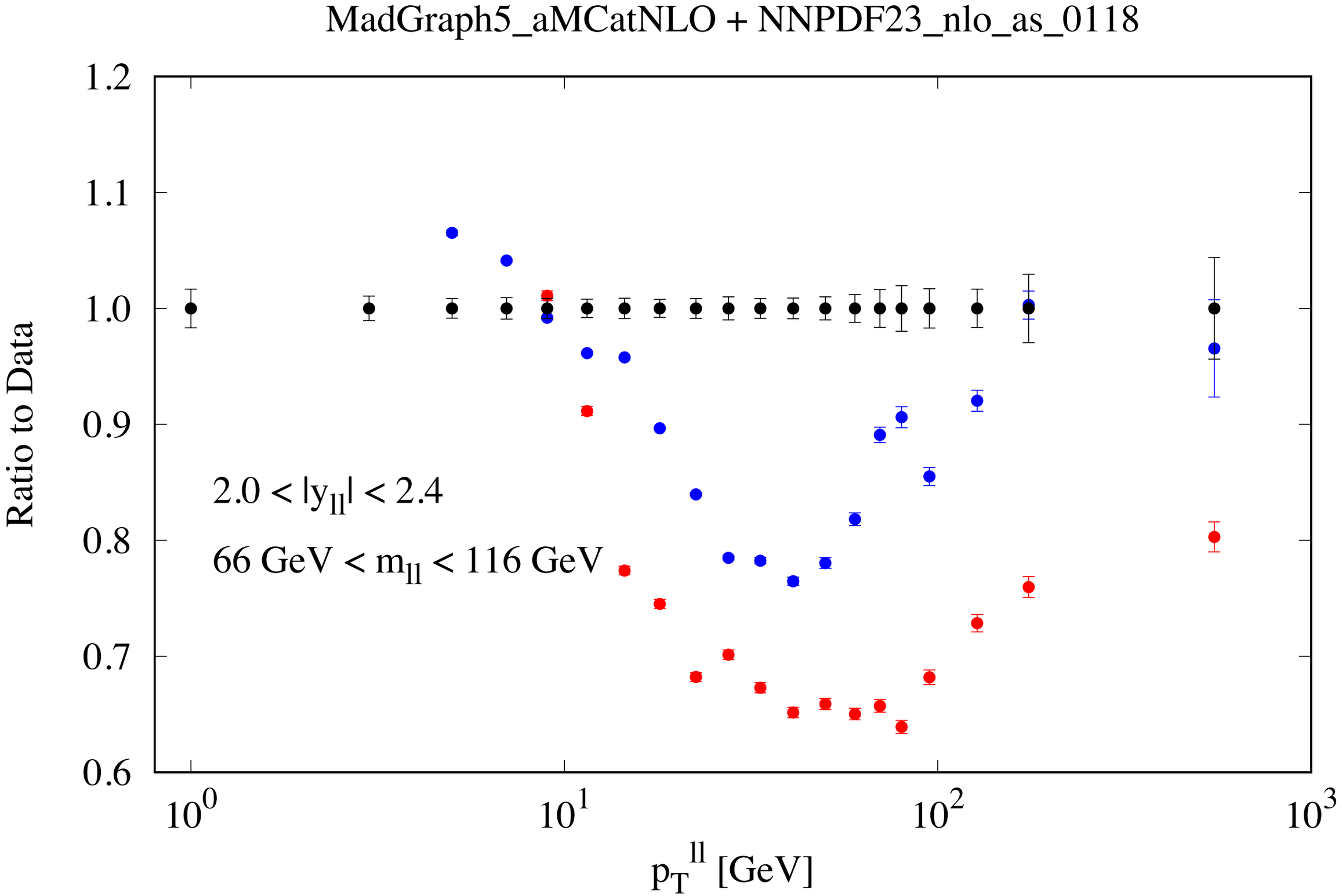
The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



The aMCfast Interface

The NLO + PS Case: The ATLAS Z p_T data



The aMCfast Interface

The NLO + PS Case: Attempting a First PDF Fit

- Actually the PS does help but the experimental uncertainties are so small that a good **description** is **hardly achievable**.

- Try to fit a set of PDFs to this data:

  - use HERA1+2 combined inclusive DIS data for the baseline,

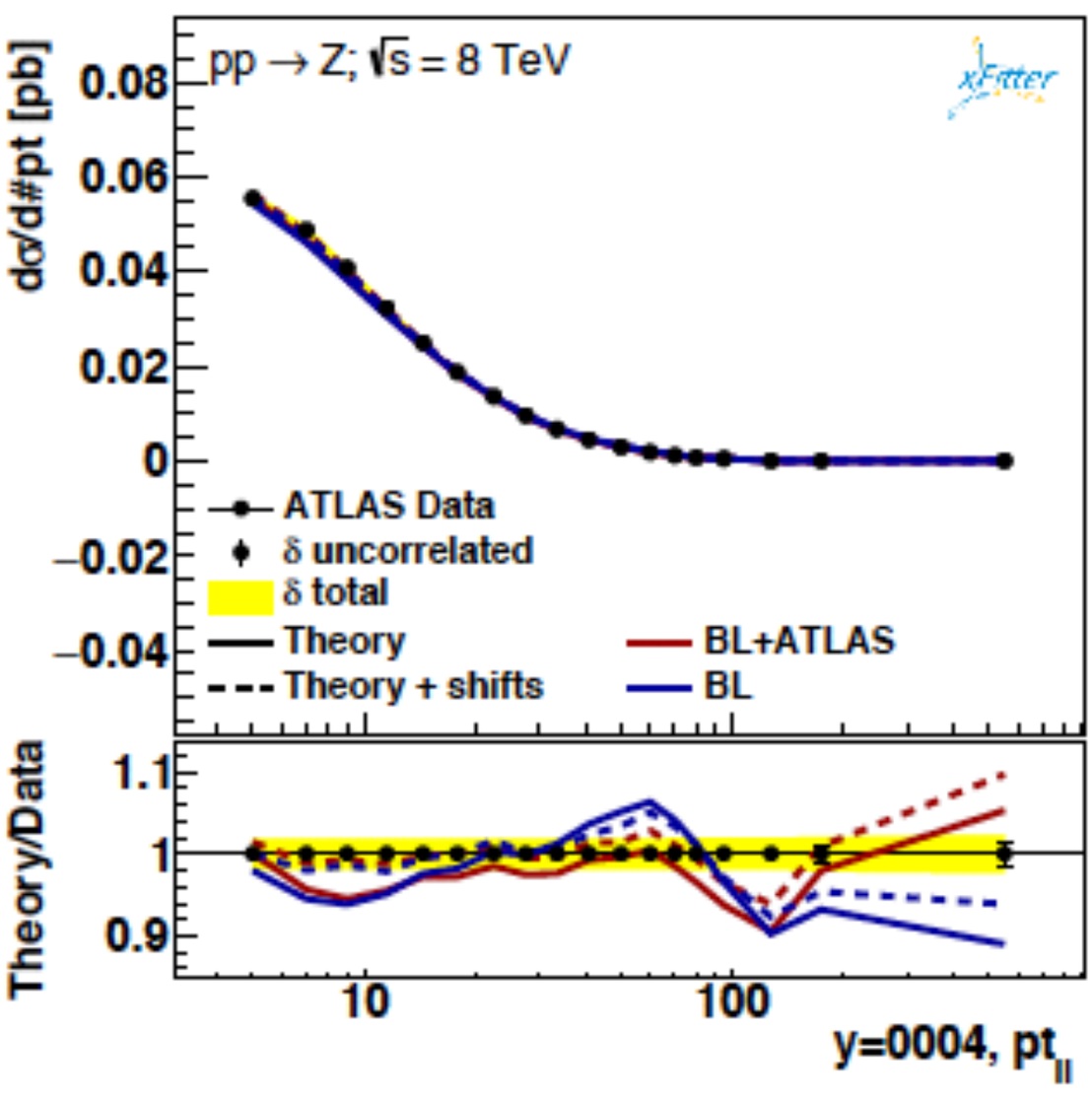
  - include ATLAS Z p_T data on top of them and check the impact on PDF.

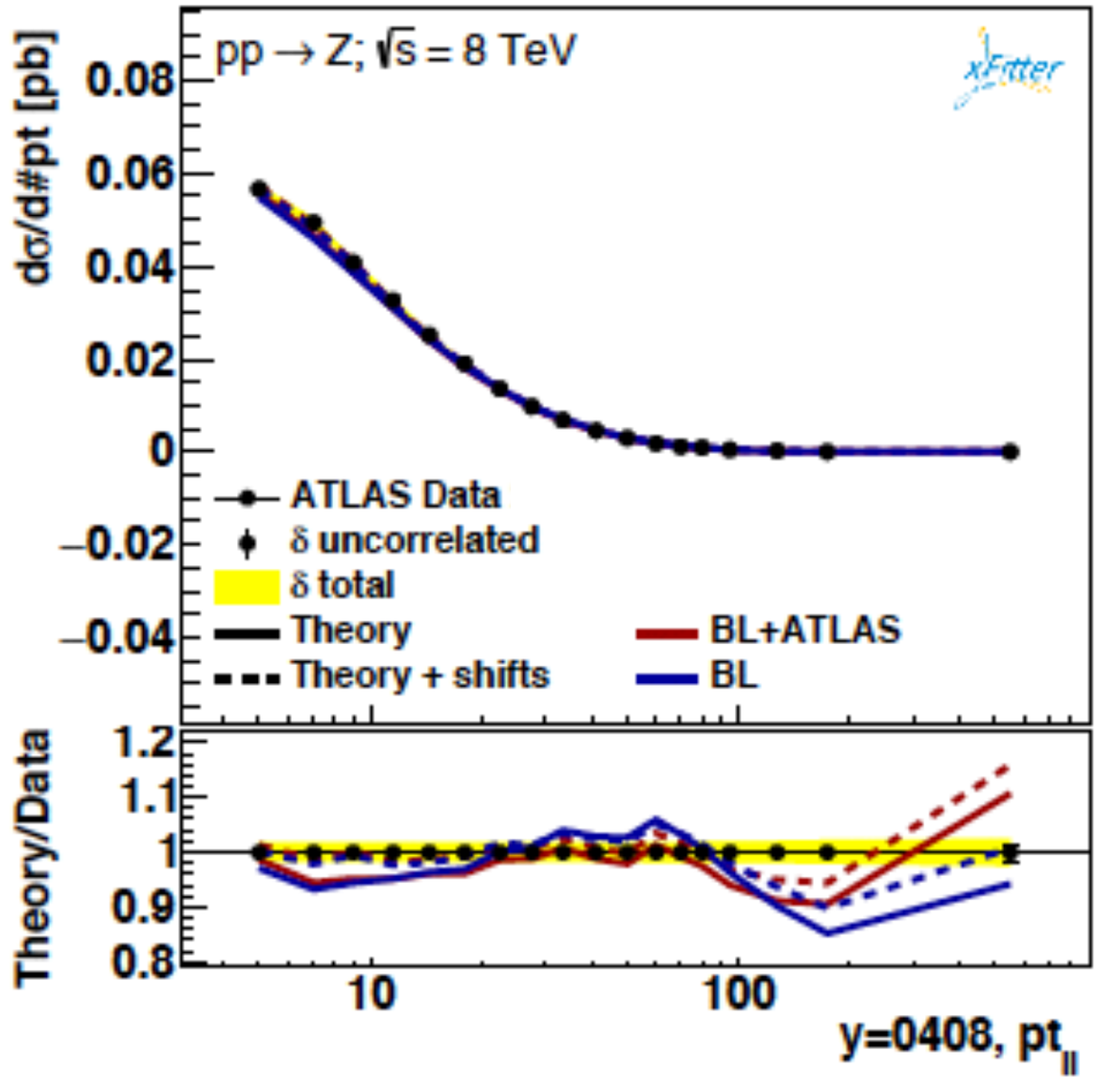
  - Exclude the first two p_T bins because they are very off.

- To do so we have used the the open-source fitting code **xFitter** (former HERAFitter).

The aMCfast Interface

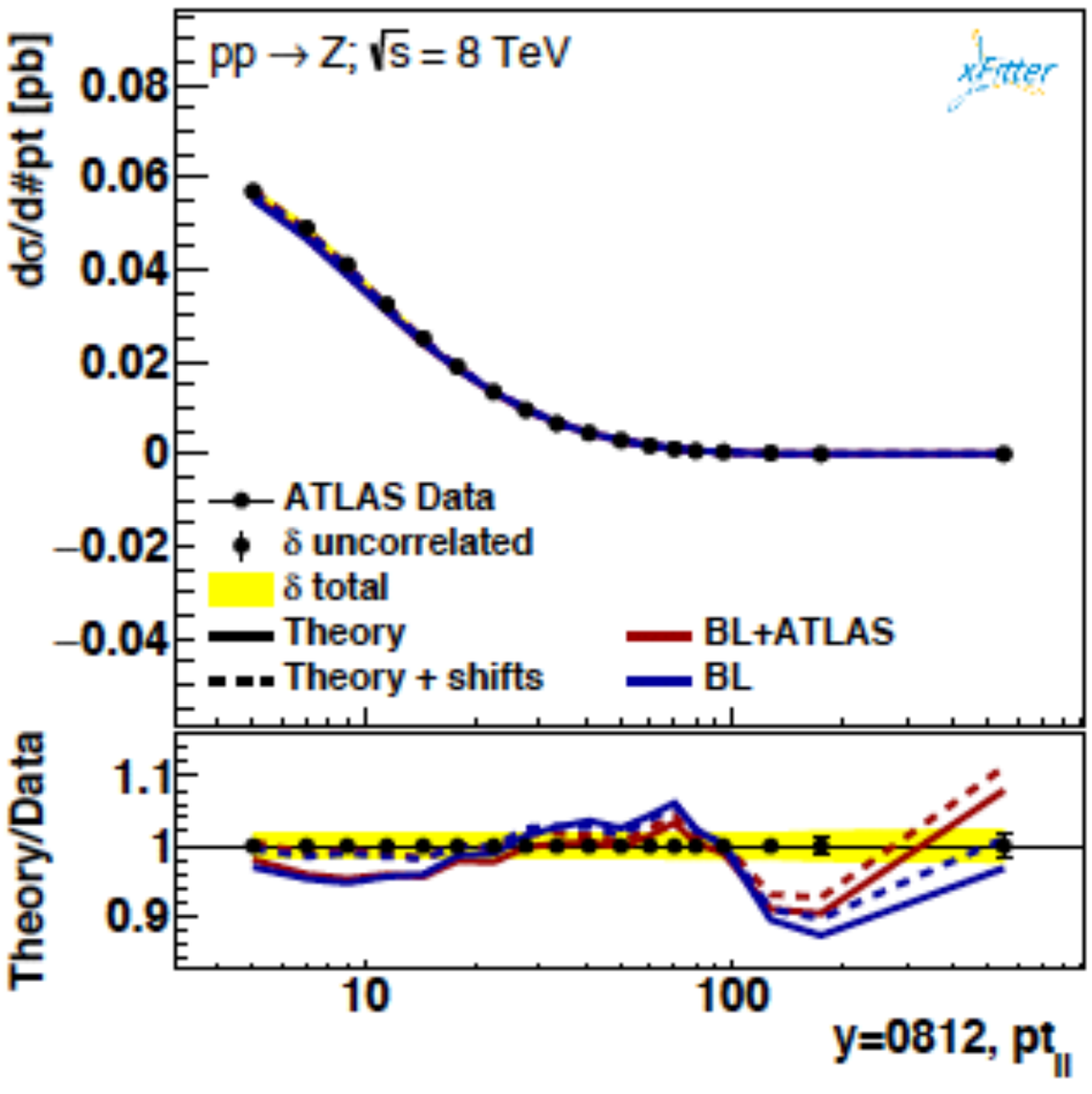
The NLO + PS Case: Attempting a First PDF Fit

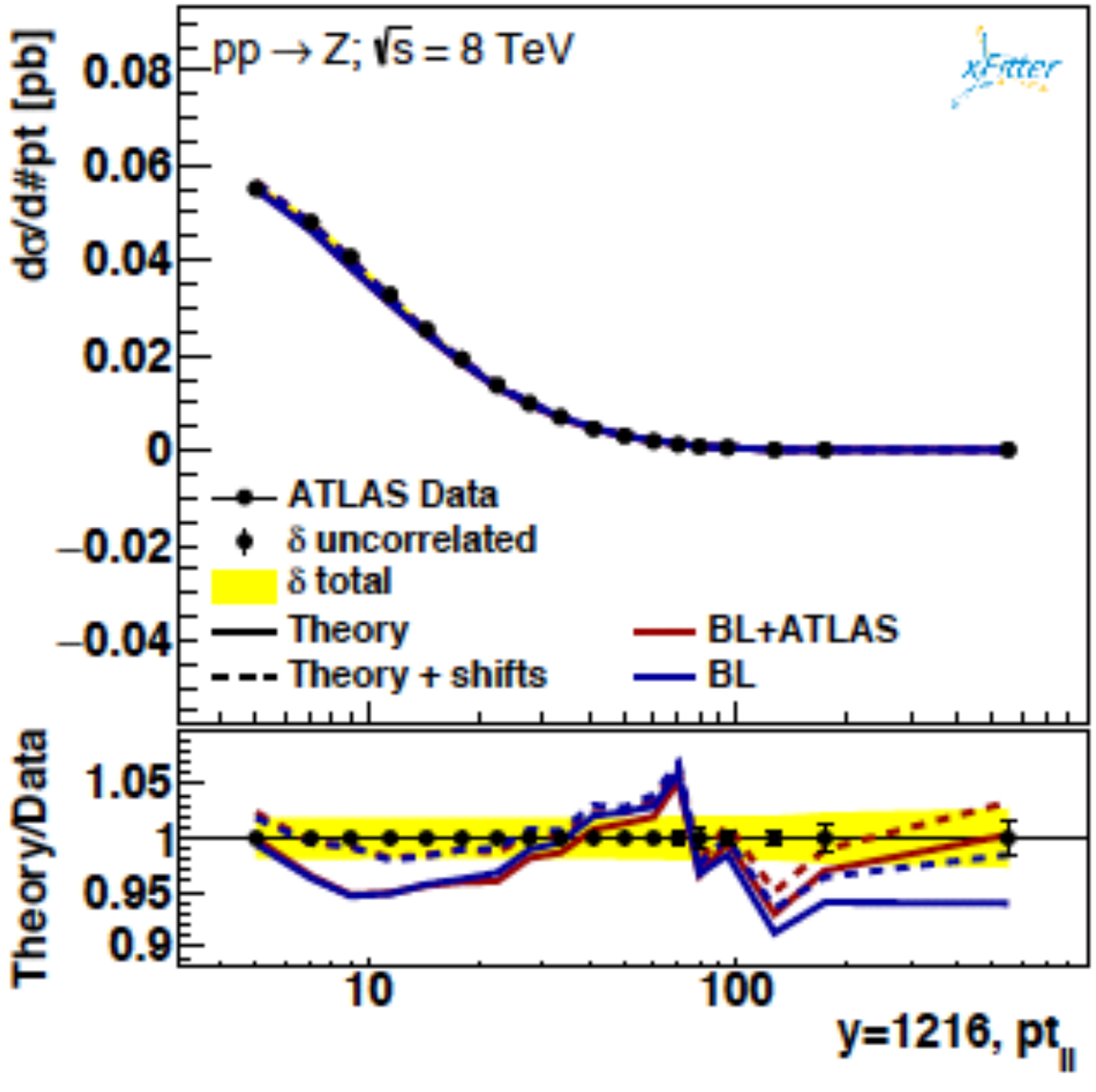




The aMCfast Interface

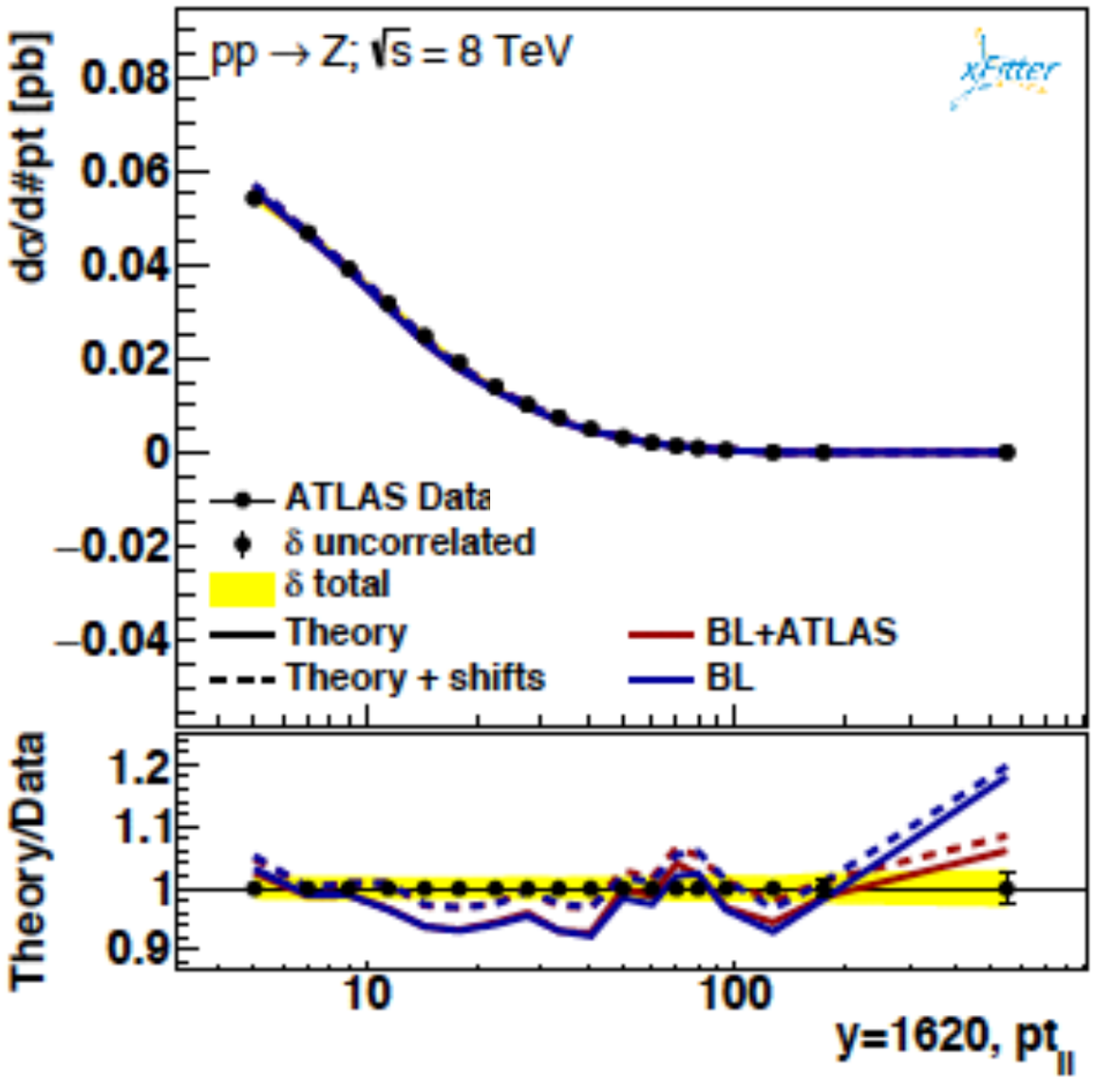
The NLO + PS Case: Attempting a First PDF Fit

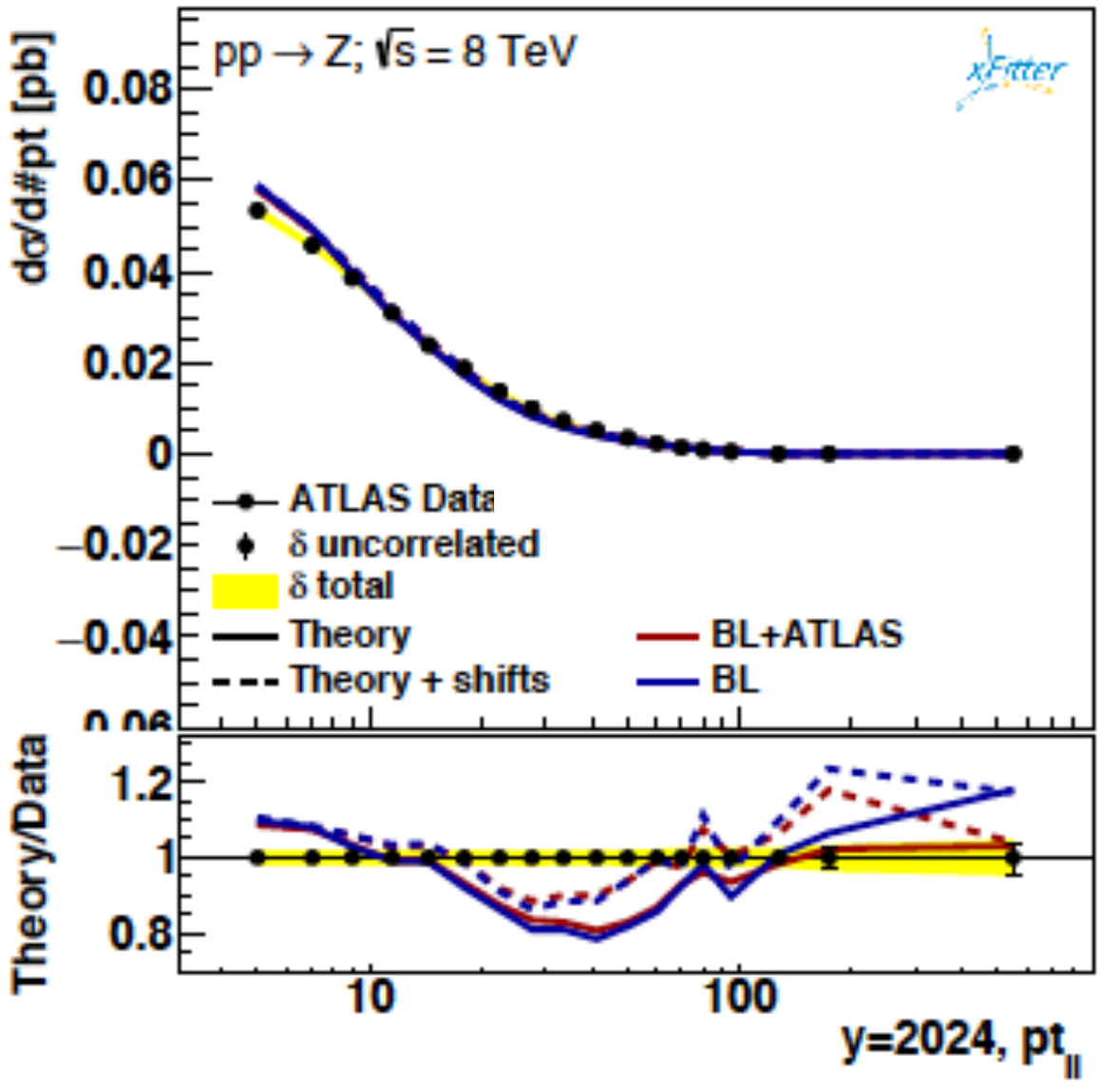




The aMCfast Interface

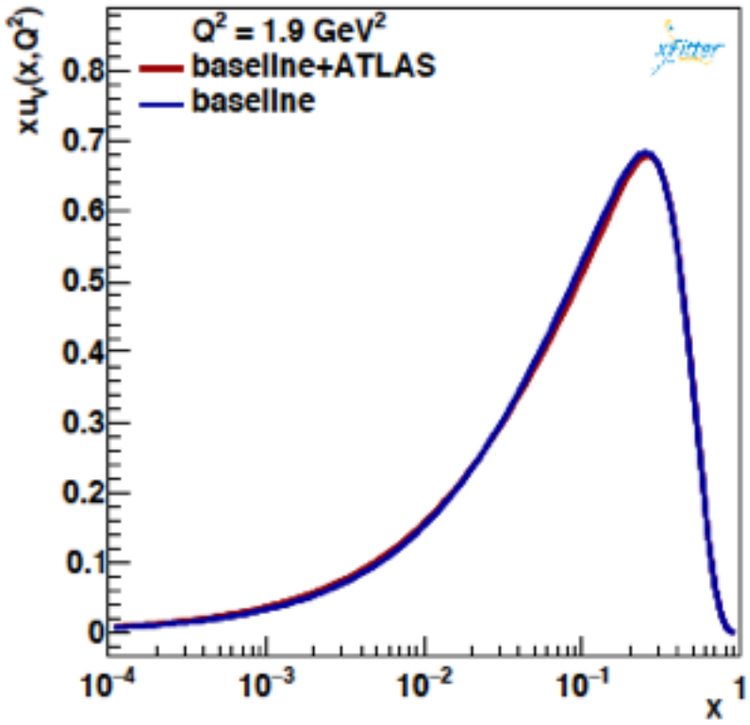
The NLO + PS Case: Attempting a First PDF Fit

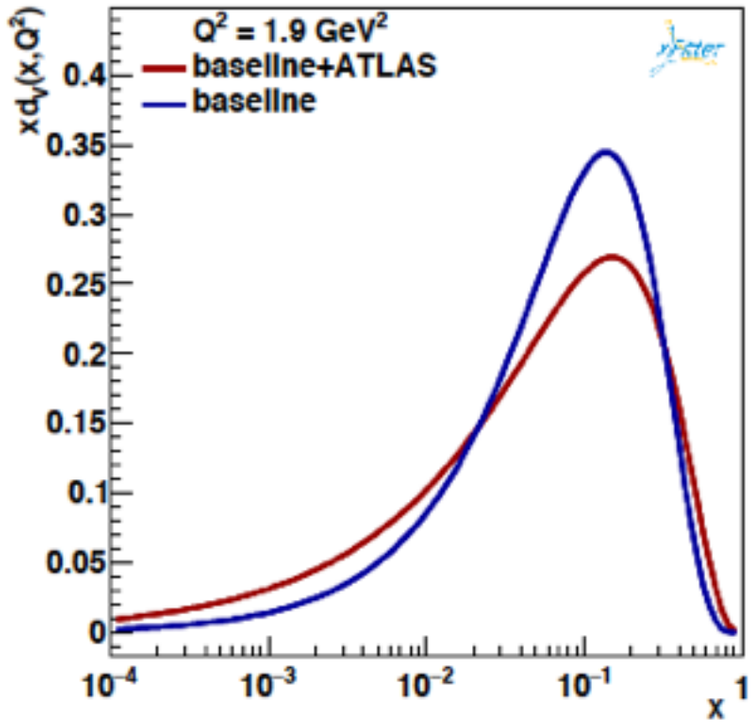


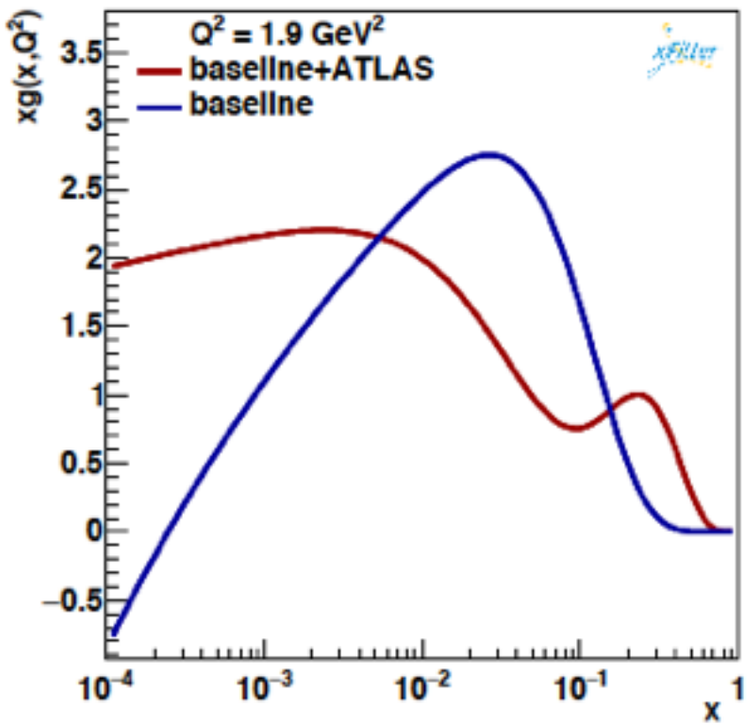


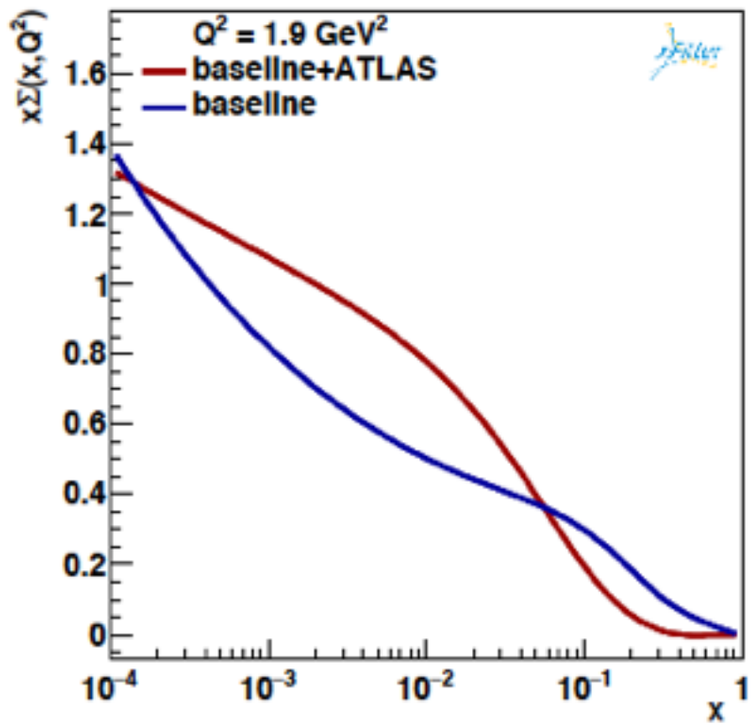
The aMCfast Interface

The NLO + PS Case: Attempting a First PDF Fit









The aMCfast Interface

The NLO + PS Case: Attempting a First PDF Fit

- Compare χ^2 before and after the fit:

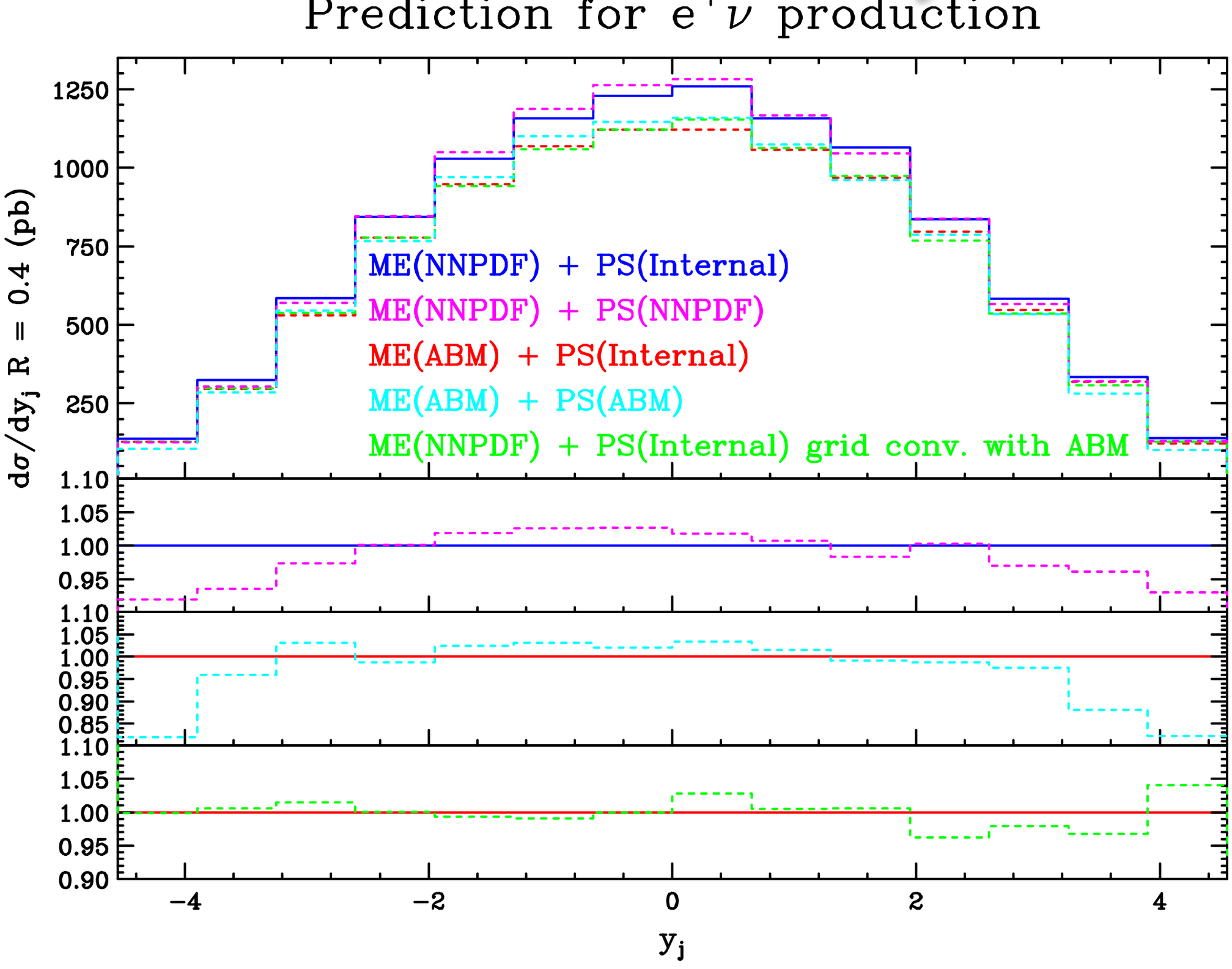
Dataset	BL+ATLAS	BL
Average y1216 101	372 / 18	449 / 18
Average y1620 101	516 / 18	583 / 18
Average y0408 101	520 / 18	798 / 18
Average y0812 101	429 / 18	731 / 18
Average y2024 101	1299 / 18	1750 / 18
Average y0004 101	356 / 18	693 / 18

- The fit does help reduce the χ^2 but the description is far from being acceptable.
- Very substantial effect on PDFs.
- Very preliminary results… this is not the last word.

Backup Slides

The aMCfast Interface

The NLO + PS Case: Preliminary Results



Fast NLO Computations

The Interpolating Grids

- The basic idea is that of a Lagrange-polynomial expansion:

$$F(z) = \sum_i F(z_i) I_i^{(s)}(z)$$

Grid nodes

Interpolation functions

Fast NLO Computations

The Interpolating Grids

- The basic idea is that of a Lagrange-polynomial expansion:

$$F(z) = \sum_i F(z_i) I_i^{(s)}(z)$$

- Suppose you want to compute numerically the following integral, *e.g.* by Monte Carlo methods:

$$J = \int_a^b dz F(z) S(z) = \sum_{k=1}^{M} \Phi_k F(z_k) S(z_k)$$

Normalization factor

Random points in the interval [a,b]

Fast NLO Computations

The Interpolating Grids

- The basic idea is that of a Lagrange-polynomial expansion:

$$F(z) = \sum_i F(z_i) I_i^{(s)}(z)$$

- Suppose you want to compute numerically the following integral, *e.g.* by Monte Carlo methods:

$$J = \int_a^b dz F(z) S(z) = \sum_{k=1}^{M} \Phi_k F(z_k) S(z_k)$$

Fast function (PDFs)

Slow function (Partonic cross sections)

Fast NLO Computations

The Interpolating Grids

- The basic idea is that of a Lagrange-polynomial expansion:

$$F(z) = \sum_i F(z_i) I_i^{(s)}(z)$$

- Suppose you want to compute numerically the following integral, *e.g.* by Monte Carlo methods:

$$J = \int_a^b dz F(z) S(z) = \sum_{k=1}^{M} \Phi_k F(z_k) S(z_k)$$

- Using the interpolation formula:

$$J = \sum_i F(z_i) G_i \quad \text{with} \quad G_i = \sum_{k=1}^{M} \Phi_k S(z_k) I_i^{(s)}(z_k)$$

(1-dimentional) interpolation grid independent of $F(z)$: precomputed and stored

Fast NLO Computations

The Interpolating Grids

- The basic idea is that of a Lagrange-polynomial expansion:

$$F(z) = \sum_i F(z_i) I_i^{(s)}(z)$$

- Suppose you want to compute numerically the following integral, *e.g.* by Monte Carlo methods:

$$J = \int_a^b dz F(z) S(z) = \sum_{k=1}^{M} \Phi_k F(z_k) S(z_k)$$

- Using the interpolation formula:

$$J = \sum_i F(z_i) G_i \quad \text{with} \quad G_i = \sum_{k=1}^{M} \Phi_k S(z_k) I_i^{(s)}(z_k)$$

- Once G_i has been precomputed, the *a posteriori* computation of J with any function $F(z)$ will be extremely fast.

Fast NLO Computations

The Hard Cross Sections in aMC@NLO at NLO

- The generalization of this procedure to the realistic case of a hard NLO cross section is straightforward, considering that:

$$d\sigma^{(\mathrm{NLO})} \quad \longleftrightarrow \quad \left\{ d\sigma^{(\mathrm{NLO},\alpha)} \right\}_{\alpha = E,S,C,SC}$$

Event & Counterevents

Fast NLO Computations

The Hard Cross Sections in aMC@NLO at NLO

- The generalization of this procedure to the realistic case of a hard NLO cross section is straightforward, considering that:

$$d\sigma^{(\mathrm{NLO})} \quad \longleftrightarrow \quad \left\{ d\sigma^{(\mathrm{NLO},\alpha)} \right\}_{\alpha=E,S,C,SC}$$

$$d\sigma^{(\mathrm{NLO},\alpha)} \quad = \quad f_1(x_1^{(\alpha)},\mu_F^{(\alpha)}) f_2(x_2^{(\alpha)},\mu_F^{(\alpha)}) W^{(\alpha)} d\chi_{Bj} d\chi_{n+1}$$

PDFs

Partonic cross section

Fast NLO Computations

The Hard Cross Sections in aMC@NLO at NLO

- The generalization of this procedure to the realistic case of a hard NLO cross section is straightforward, considering that:

$$d\sigma^{(\mathrm{NLO})} \quad \longleftrightarrow \quad \left\{ d\sigma^{(\mathrm{NLO},\alpha)} \right\}_{\alpha=E,S,C,SC}$$

$$d\sigma^{(\mathrm{NLO},\alpha)} \quad = \quad f_1(x_1^{(\alpha)},\mu_F^{(\alpha)}) f_2(x_2^{(\alpha)},\mu_F^{(\alpha)}) W^{(\alpha)} d\chi_{Bj} d\chi_{n+1}$$

$$W^{(\alpha)} \quad = \quad \alpha_s^{b+1}(\mu_R^{(\alpha)}) \left[W_0^{(\alpha)} + W_F^{(\alpha)} \ln\left(\frac{\mu_F^{(\alpha)}}{Q}\right) + W_R^{(\alpha)} \ln\left(\frac{\mu_R^{(\alpha)}}{Q}\right) \right]$$

$$+ \quad \alpha_s^{b}(\mu_R^{(\alpha)}) W_B \delta_{\alpha S}$$

NLO term

Born term

Fast NLO Computations

The Hard Cross Sections in aMC@NLO at NLO

- The generalization of this procedure to the realistic case of a hard NLO cross section is straightforward, considering that:

$$d\sigma^{(\mathrm{NLO})} \quad \longleftrightarrow \quad \left\{ d\sigma^{(\mathrm{NLO},\alpha)} \right\}_{\alpha=E,S,C,SC}$$

Slow functions

Fast functions

$$d\sigma^{(\mathrm{NLO},\alpha)} = f_1(x_1^{(\alpha)},\mu_F^{(\alpha)}) f_2(x_2^{(\alpha)},\mu_F^{(\alpha)}) W^{(\alpha)} d\chi_{Bj} d\chi_{n+1}$$

$$W^{(\alpha)} = \alpha_s^{b+1}(\mu_R^{(\alpha)}) \left[W_0^{(\alpha)} + W_F^{(\alpha)} \ln\left(\frac{\mu_F^{(\alpha)}}{Q}\right) + W_R^{(\alpha)} \ln\left(\frac{\mu_R^{(\alpha)}}{Q}\right) \right]$$
$$+ \alpha_s^{b}(\mu_R^{(\alpha)}) W_B \delta_{\alpha S}$$

Fast NLO Computations

The Hard Cross Sections in aMC@NLO at NLO

- The generalization of this procedure to the realistic case of a hard NLO cross section is straightforward, considering that:

$$d\sigma^{(\mathrm{NLO})} \quad \longleftrightarrow \quad \left\{ d\sigma^{(\mathrm{NLO},\alpha)} \right\}_{\alpha=E,S,C,SC}$$

Slow functions

Fast functions

$$d\sigma^{(\mathrm{NLO},\alpha)} \quad = \quad f_1(x_1^{(\alpha)},\mu_F^{(\alpha)}) f_2(x_2^{(\alpha)},\mu_F^{(\alpha)}) W^{(\alpha)} d\chi_{Bj} d\chi_{n+1}$$

$$W^{(\alpha)} \quad = \quad \alpha_s^{b+1}(\mu_R^{(\alpha)}) \left[W_0^{(\alpha)} + W_F^{(\alpha)} \ln\left(\frac{\mu_F^{(\alpha)}}{Q}\right) + W_R^{(\alpha)} \ln\left(\frac{\mu_R^{(\alpha)}}{Q}\right) \right]$$

$$+ \quad \alpha_s^{b}(\mu_R^{(\alpha)}) W_B \delta_{\alpha S}$$

- 4 **slow functions** $\Rightarrow$ **4 interpolation grids**.

- The **fast functions** are functions of 4 independent variables (x_1,x_2,μ_F,μ_R) $\Rightarrow$ 4-dimentional interpolation grids needed.

Fast NLO Computations

The Hard Cross Sections in aMC@NLO at NLO

- The generalization of this procedure to the realistic case of a hard NLO cross section is straightforward, considering that:

$$d\sigma^{(\mathrm{NLO})} \quad \longleftrightarrow \quad \left\{ d\sigma^{(\mathrm{NLO},\alpha)} \right\}_{\alpha=E,S,C,SC}$$

Slow functions

Fast functions

$$d\sigma^{(\mathrm{NLO},\alpha)} \quad = \quad f_1(x_1^{(\alpha)},\mu_F^{(\alpha)}) f_2(x_2^{(\alpha)},\mu_F^{(\alpha)}) W^{(\alpha)} d\chi_{Bj} d\chi_{n+1}$$

$$W^{(\alpha)} \quad = \quad \alpha_s^{b+1}(\mu_R^{(\alpha)}) \left[W_0^{(\alpha)} + W_F^{(\alpha)} \ln\left(\frac{\mu_F^{(\alpha)}}{Q}\right) + W_R^{(\alpha)} \ln\left(\frac{\mu_R^{(\alpha)}}{Q}\right) \right]$$

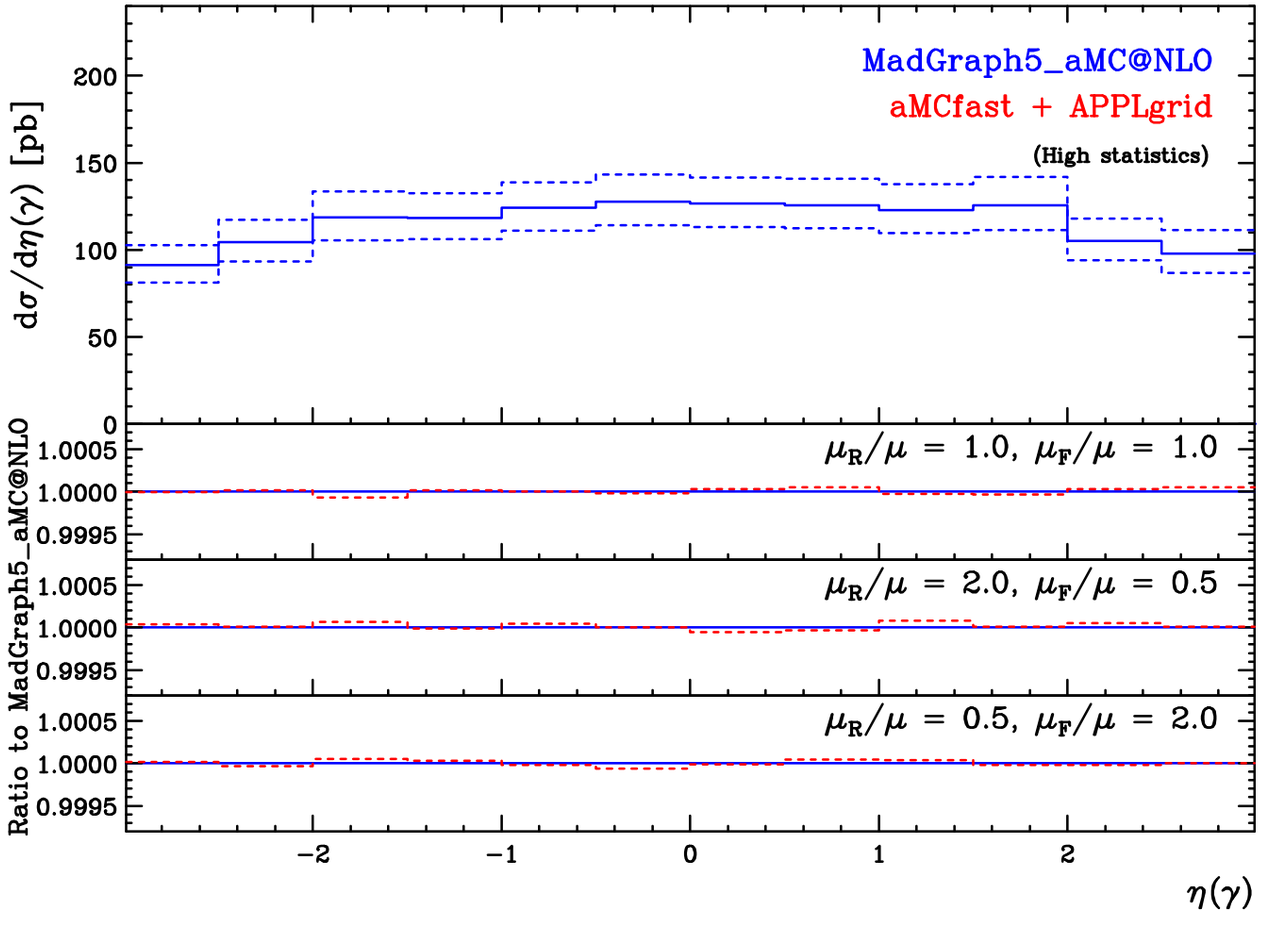
$$+ \quad \alpha_s^{b}(\mu_R^{(\alpha)}) W_B \delta_{\alpha S}$$

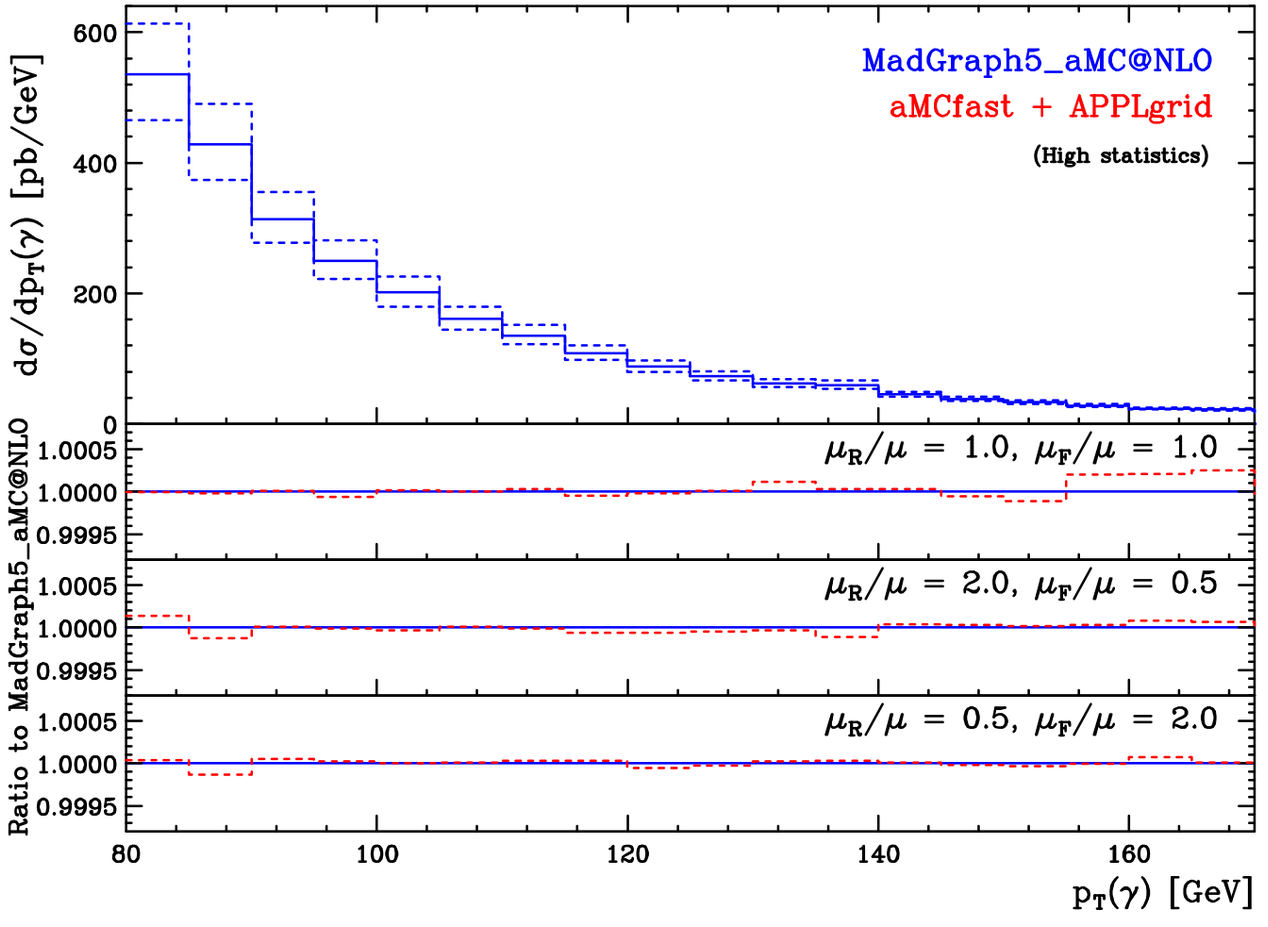
- 4 **slow functions** $\Rightarrow$ **4 interpolation grids**.
- The **fast functions** are functions of 4 independent variables (x_1,x_2,μ_F,μ_R) $\Rightarrow$ 4-dimentional interpolation grids needed.
- But assuming $\mu_F \propto \mu_R$ $\Rightarrow$ **3-dimentional** interpolation grids.

The aMCfast Interface

Validation: Photon Production with one Jet

- Important for the gluon in the region relevant for Higgs production in gluon fusion.
- We looked at the following observables:
  - the pseudo-rapidity distribution of the photon (left),
  - the tranverse momentum distribution of the photon (right).
- High statistics plots:

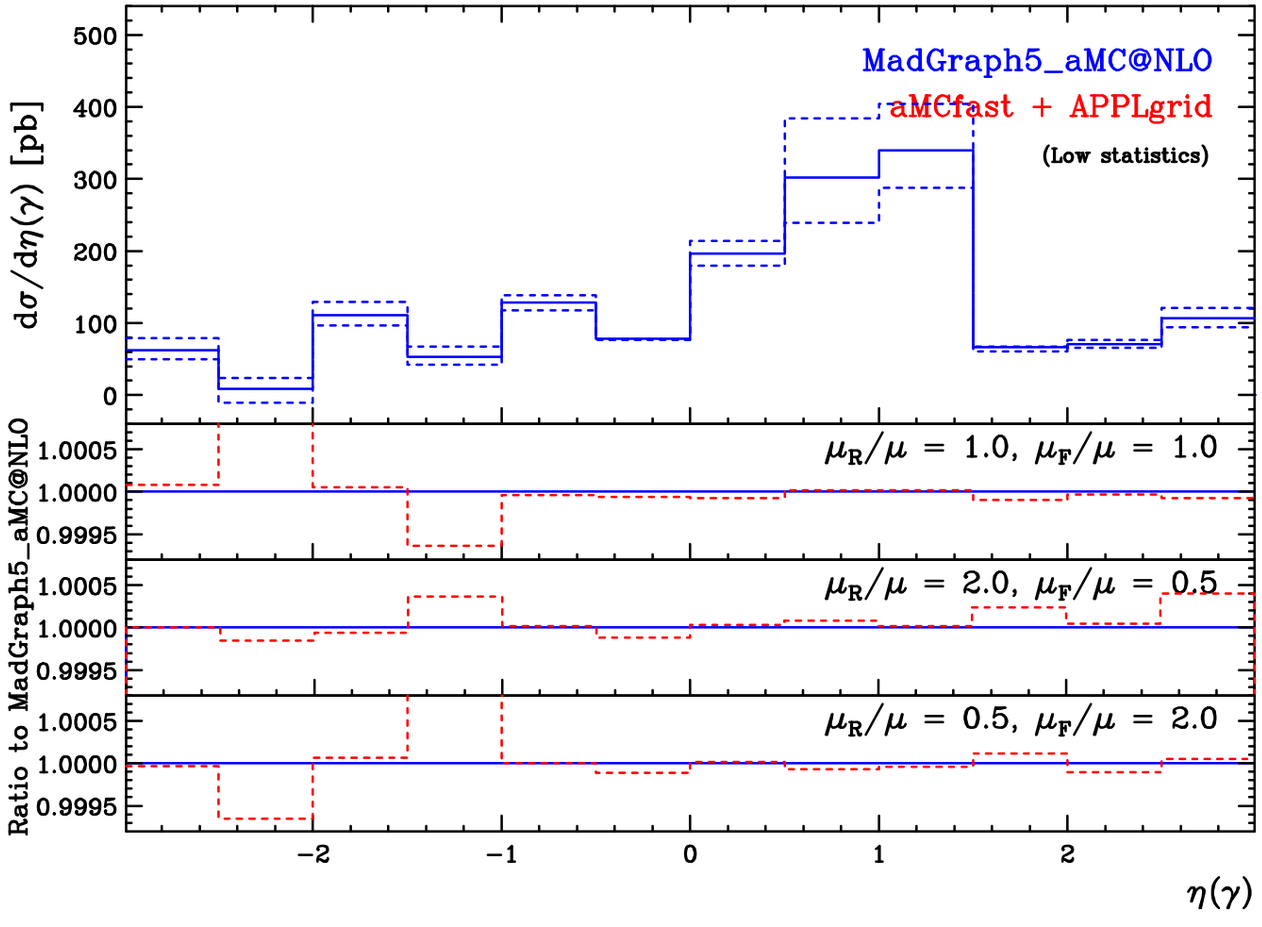


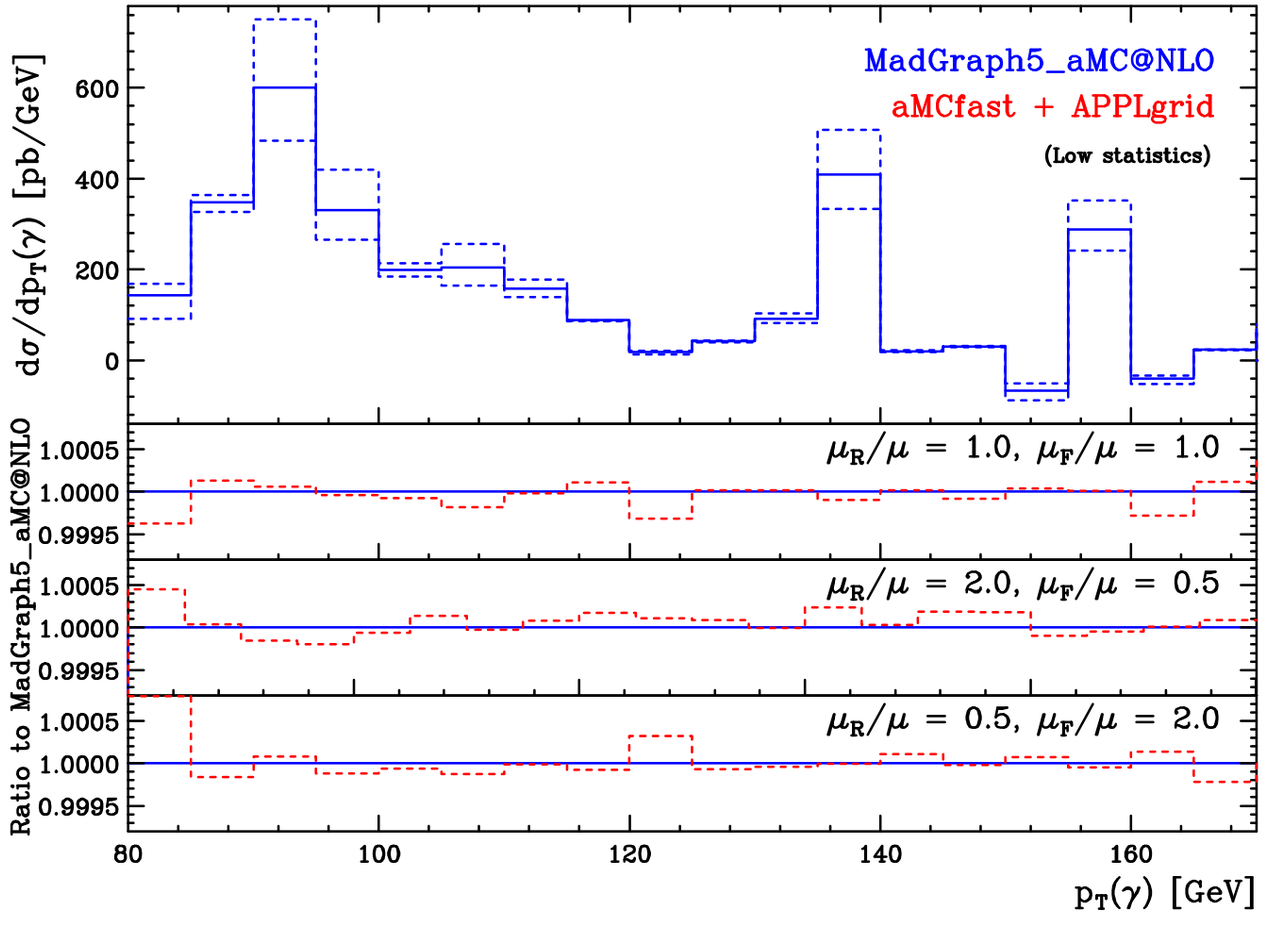


The aMCfast Interface

Validation: Photon Production with one Jet

- Important for the gluon in the region relevant for Higgs production in gluon fusion.
- We looked at the following observables:
  - the pseudo-rapidity distribution of the photon (left),
  - the tranverse momentum distribution of the photon (right).
- Low statistics plots:

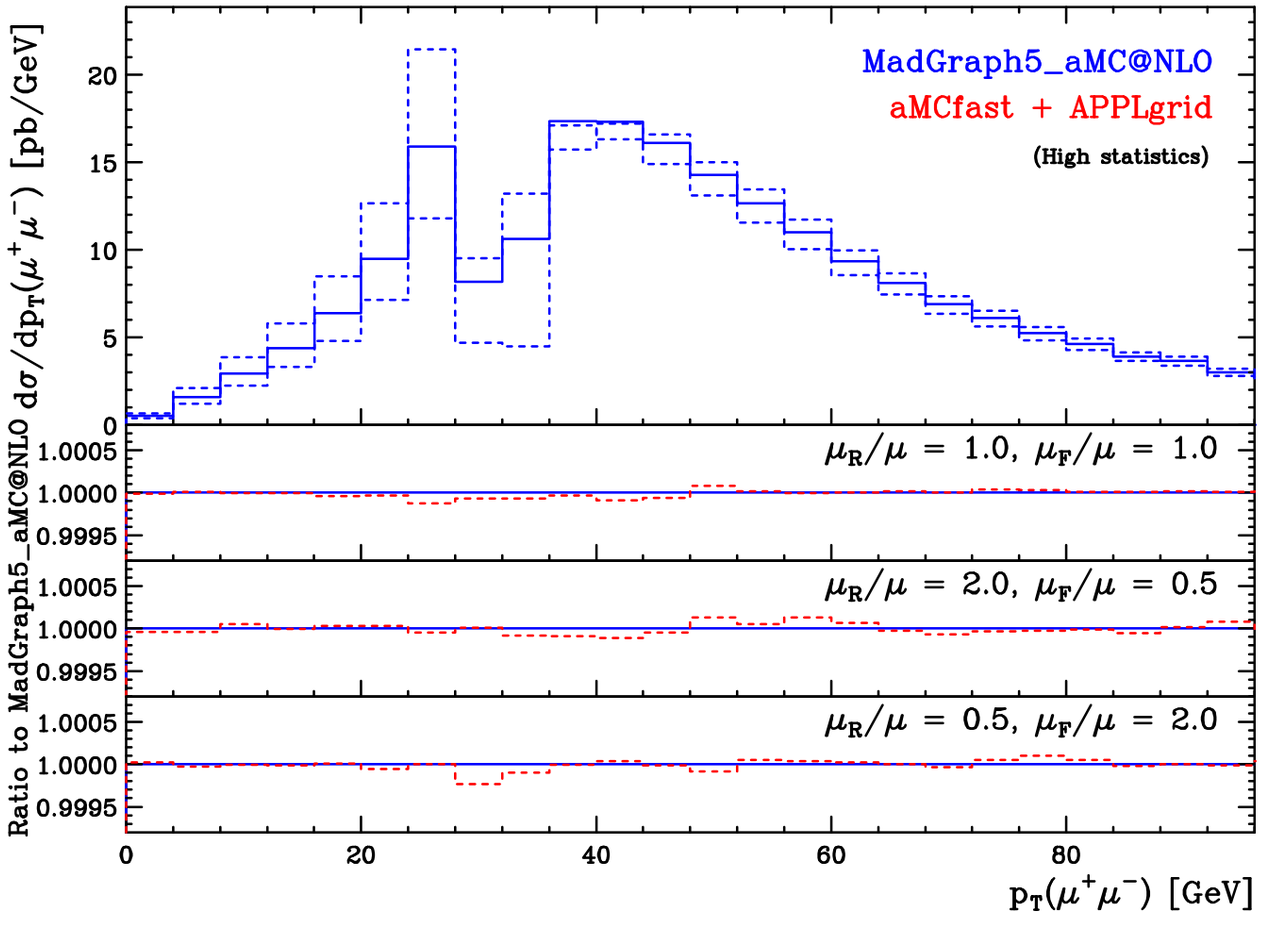


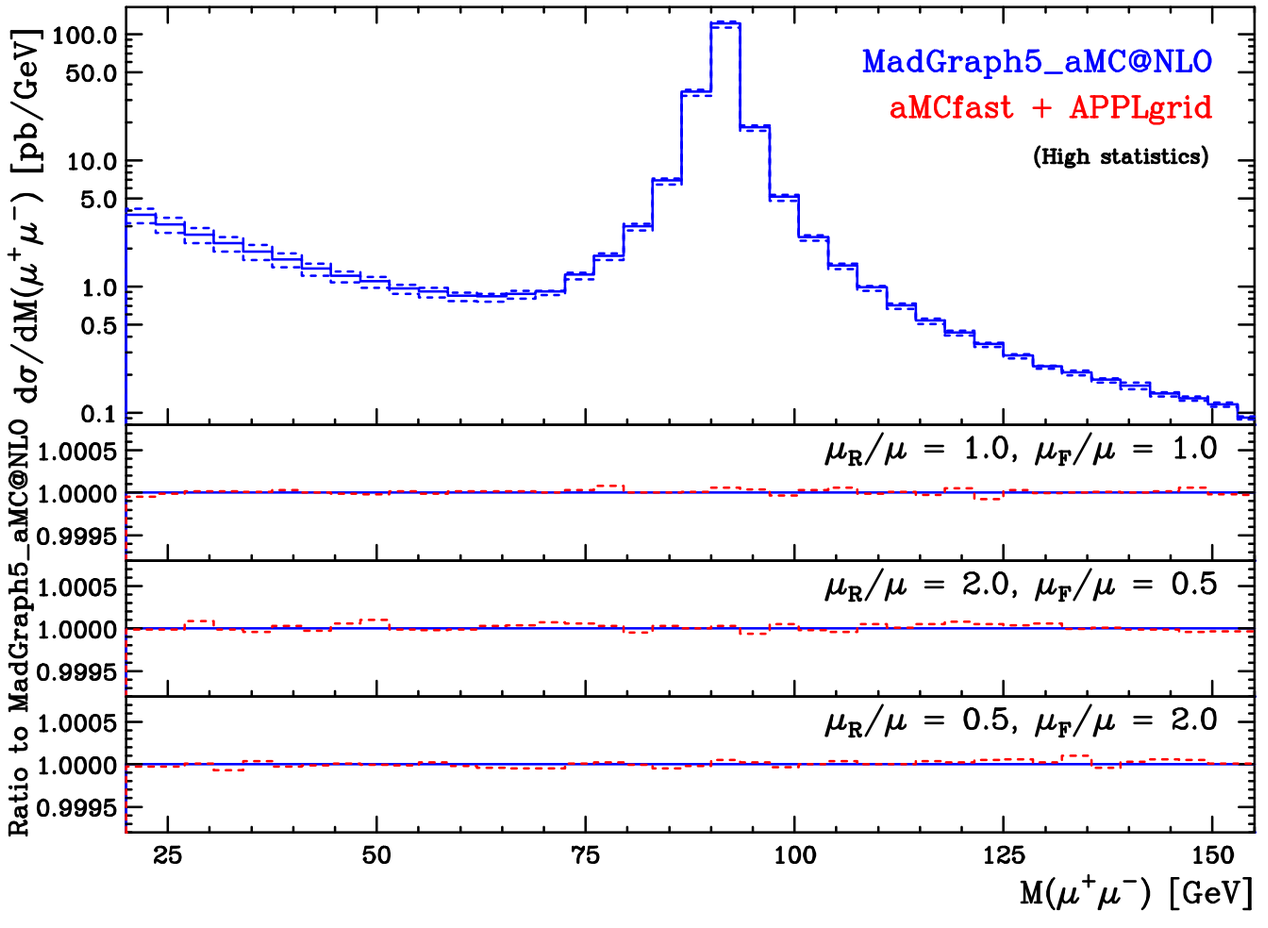


The aMCfast Interface

Validation: Dilepton Production with one Jet

- Relevant for quarks and antiquarks in the large-x region.
- We looked at the following observables:
  - the transverse momentum distribution of the lepton pair (left),
  - the invariant mass distribution of the lepton pair (right).
- High statistics plots:

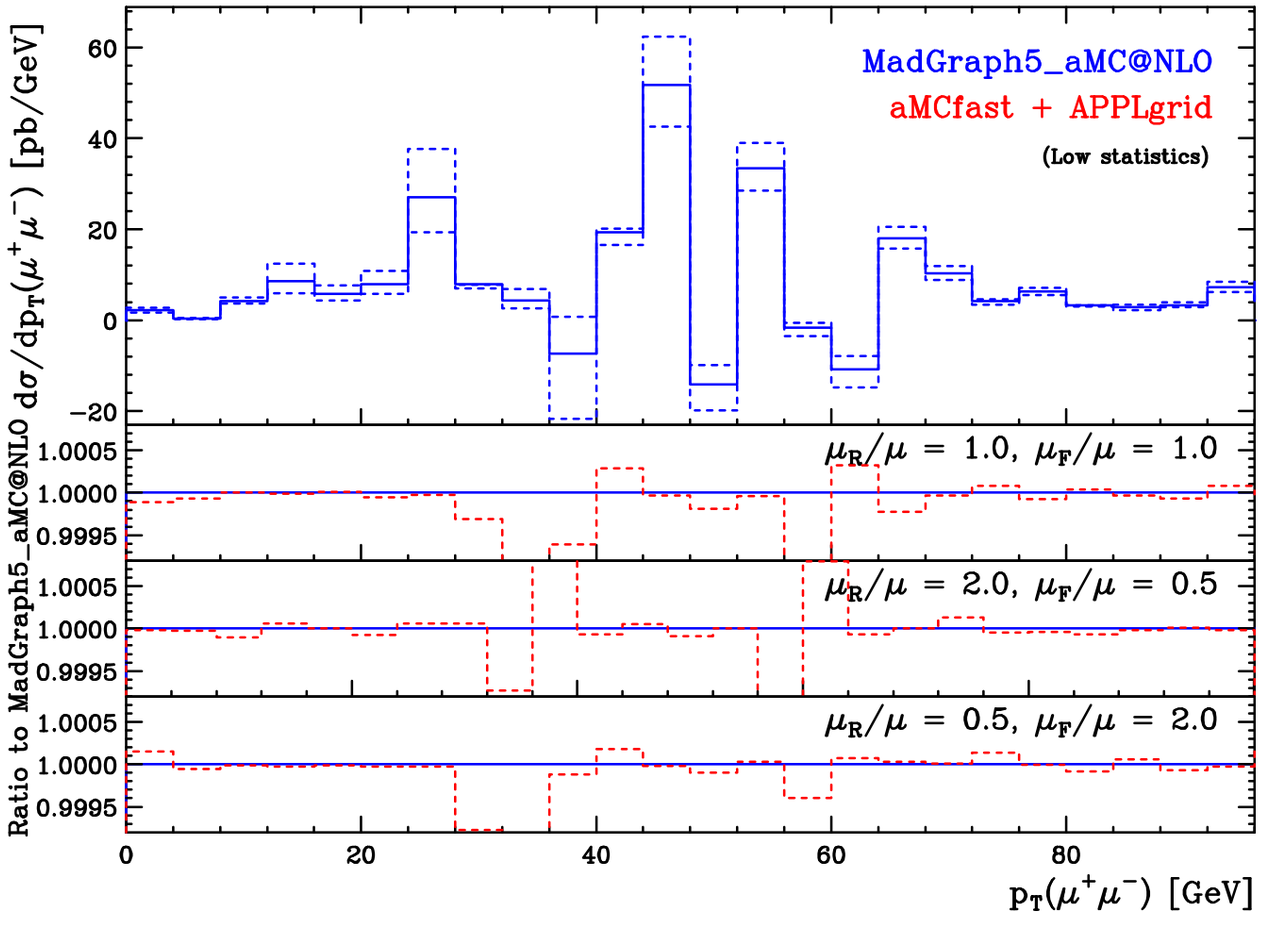


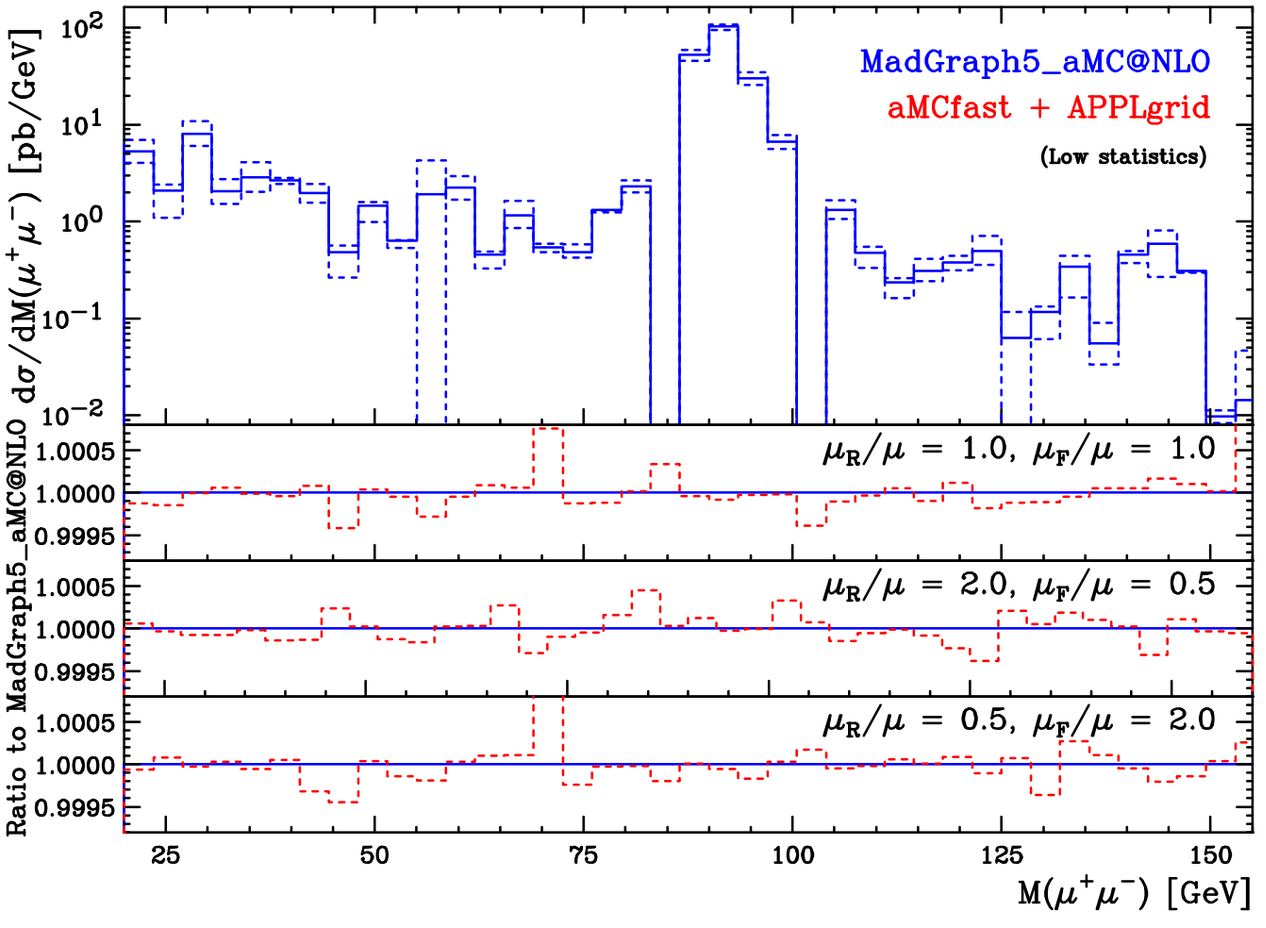


The aMCfast Interface

Validation: Dilepton Production with one Jet

- Relevant for quarks and antiquarks in the large-x region.
- We looked at the following observables:
  - the transverse momentum distribution of the lepton pair (left),
  - the invariant mass distribution of the lepton pair (right).
- Low statistics plots:

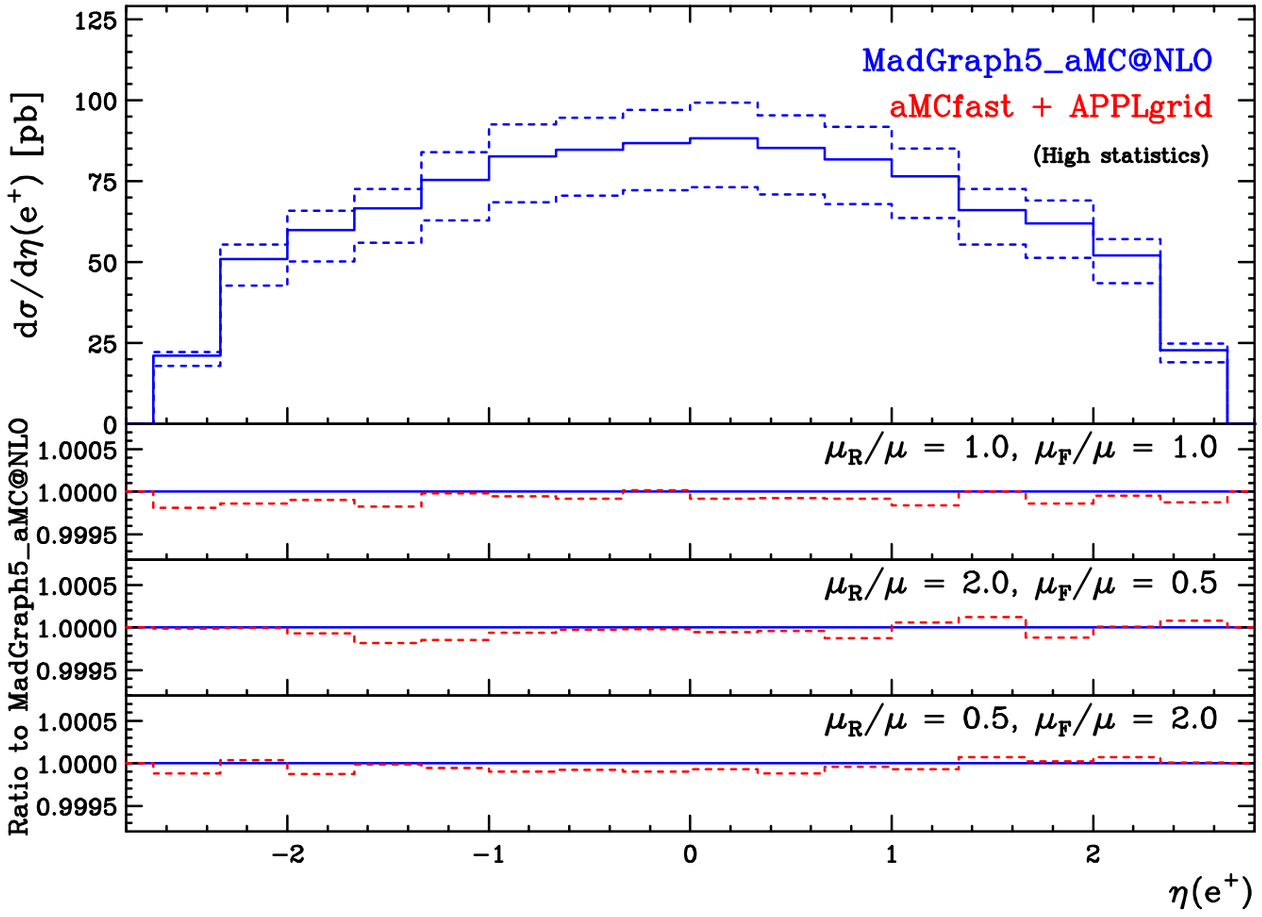


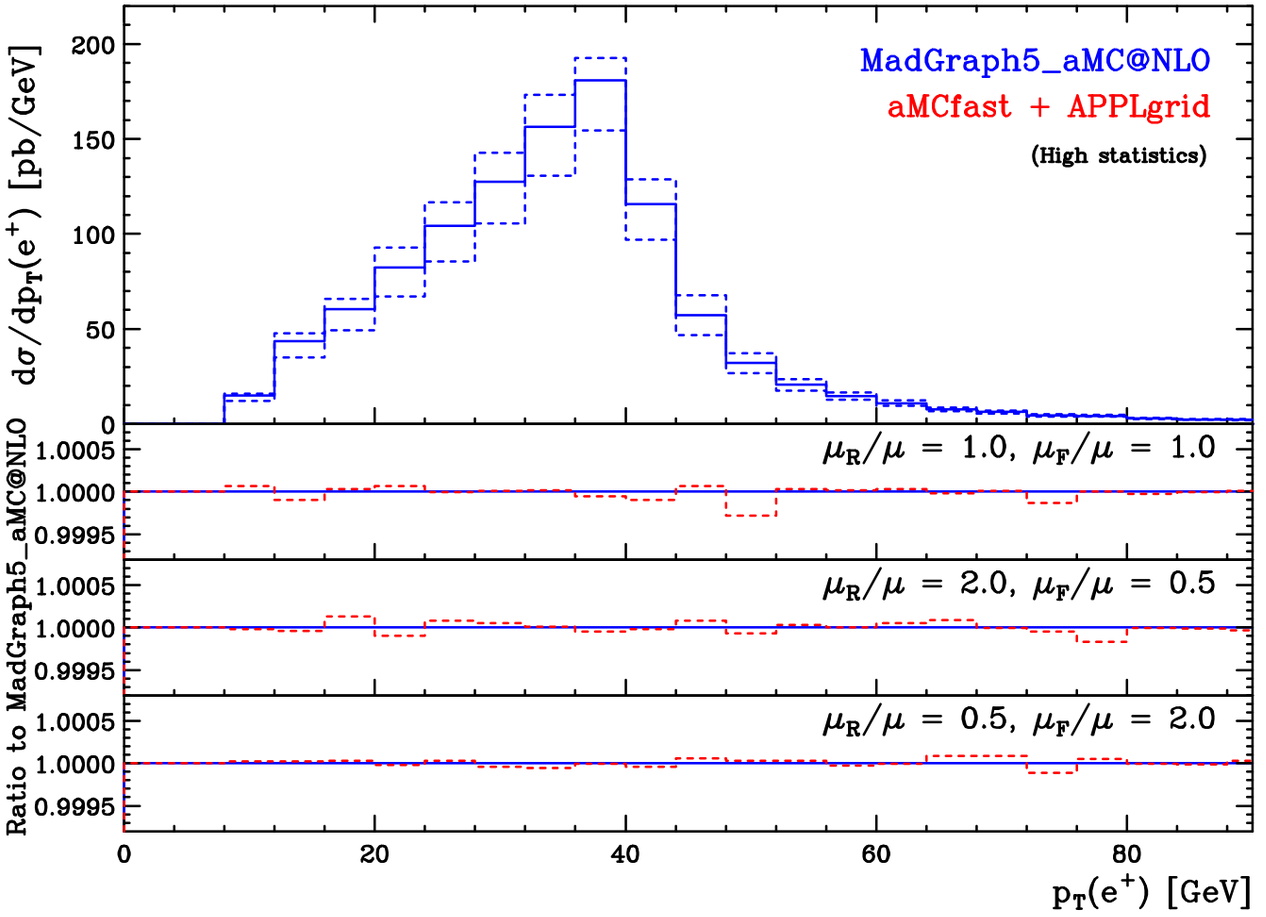


The aMCfast Interface

Validation: W + c Production

- Important for constraining the strange-quark PDFs.
- We looked at the following observables:
  - the pseudo-rapidity distribution of the lepton (left),
  - the transverse momentum distribution of the lepton (right).
- High statistics plots:

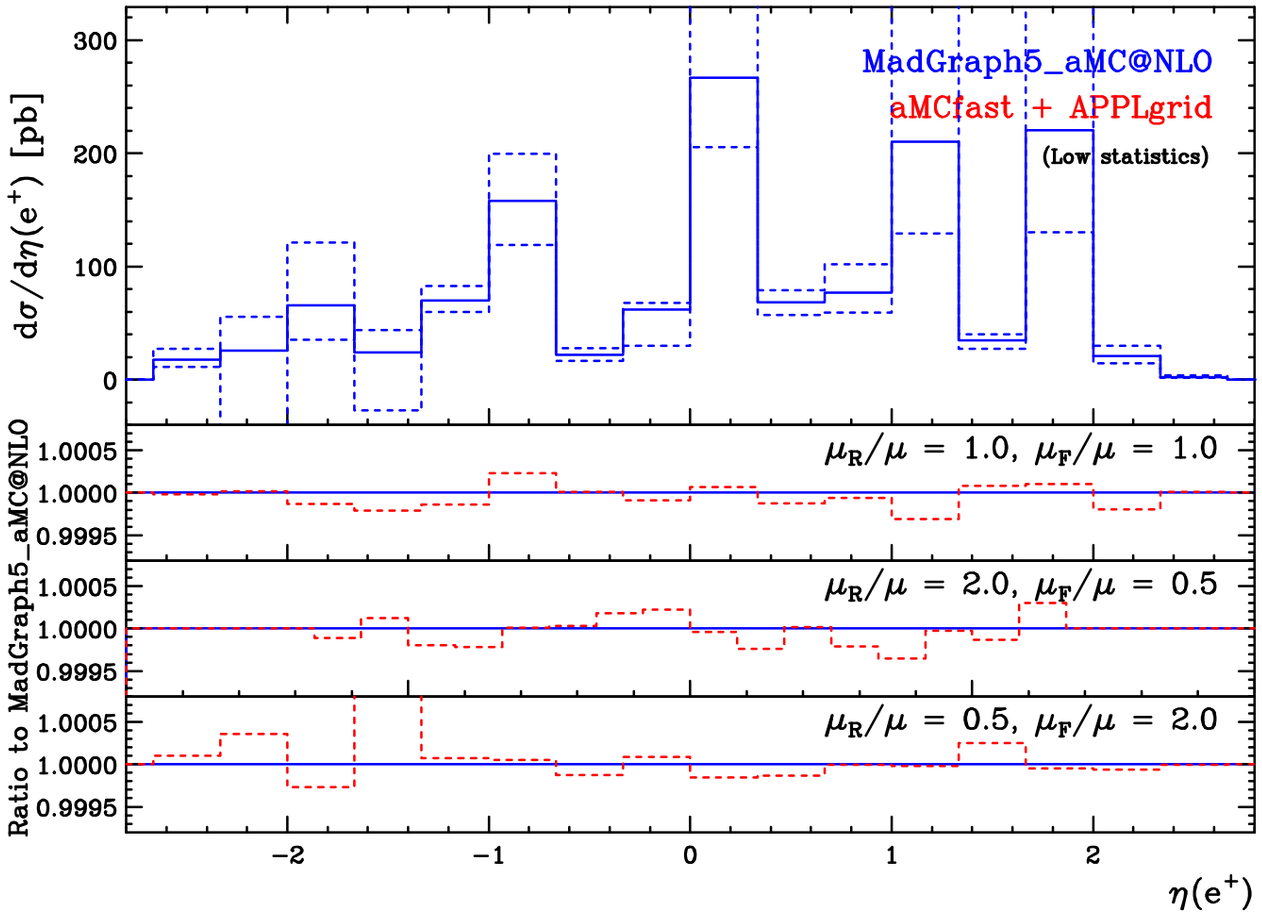


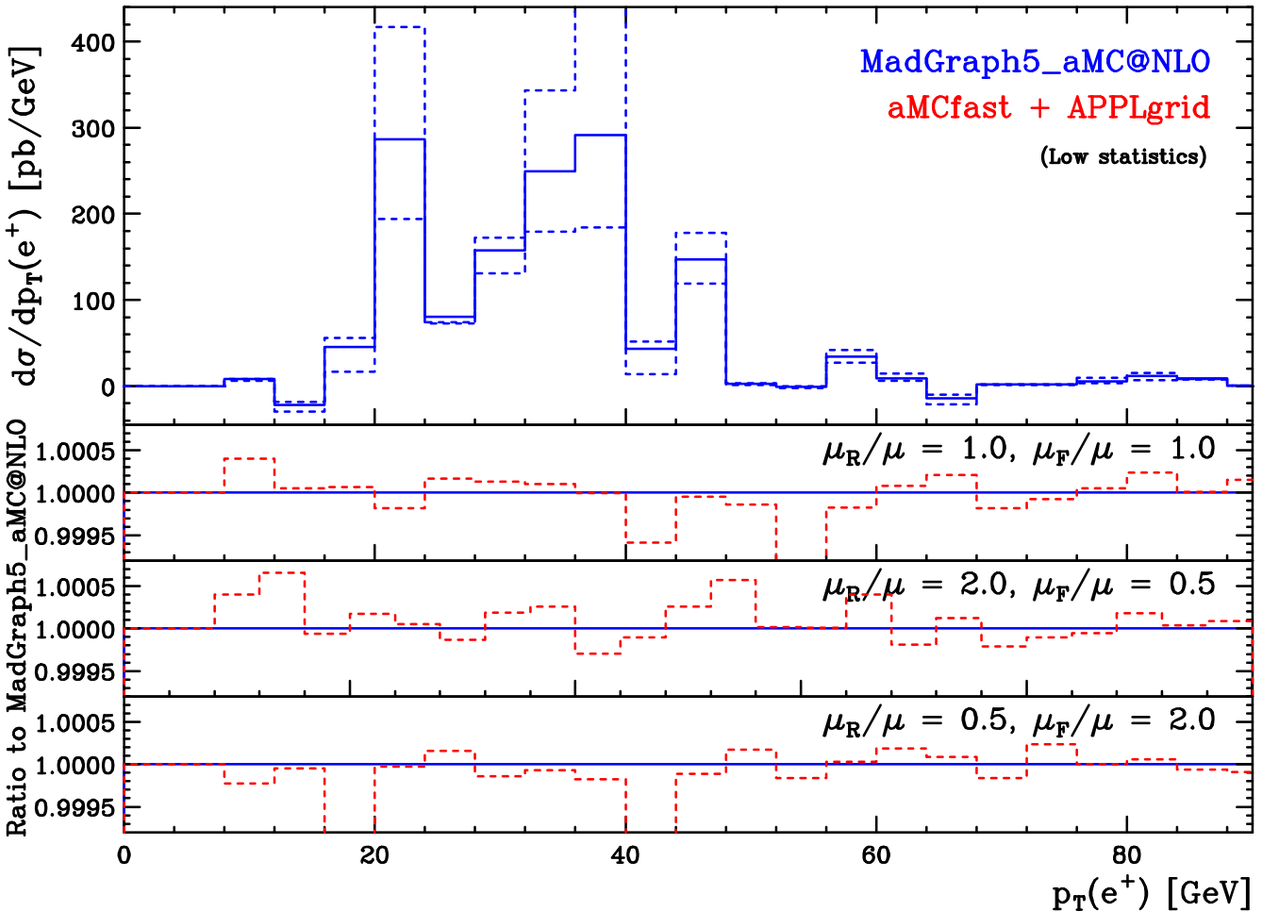


The aMCfast Interface

Validation: W + c Production

- Important for constraining the strange-quark PDFs.
- We looked at the following observables:
  - the pseudo-rapidity distribution of the lepton (left),
  - the transverse momentum distribution of the lepton (right).
- Low statistics plots:

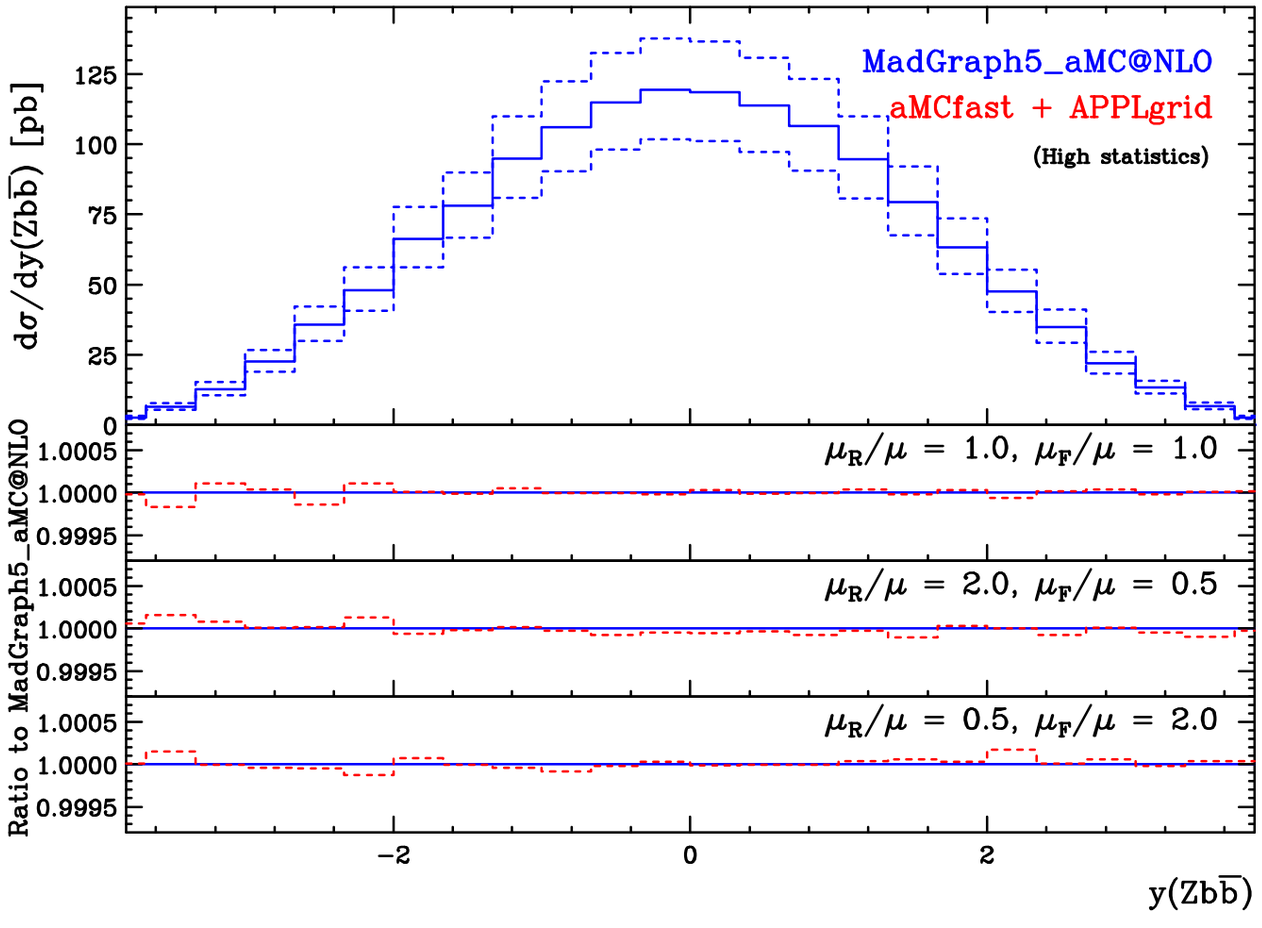


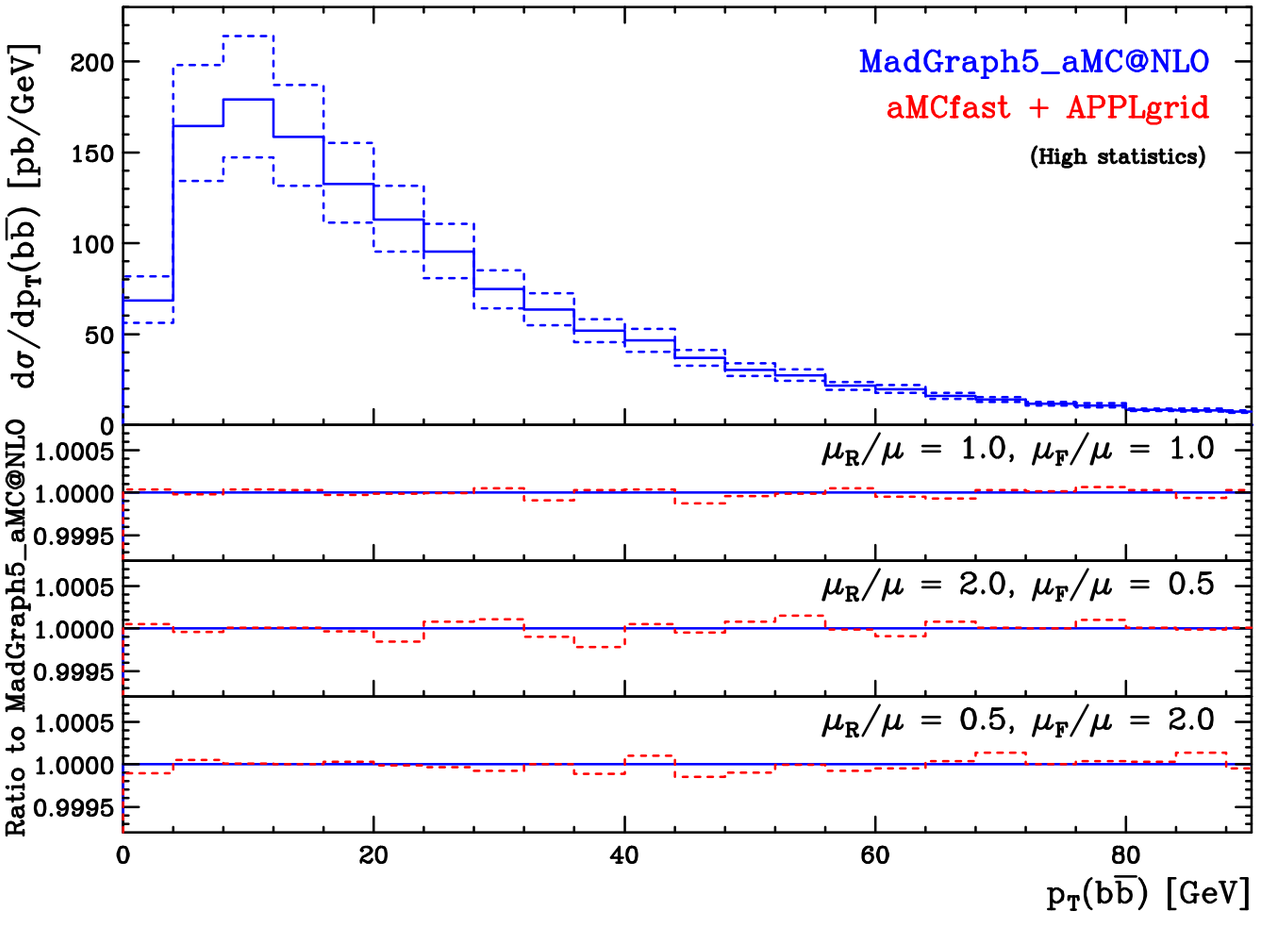


The aMCfast Interface

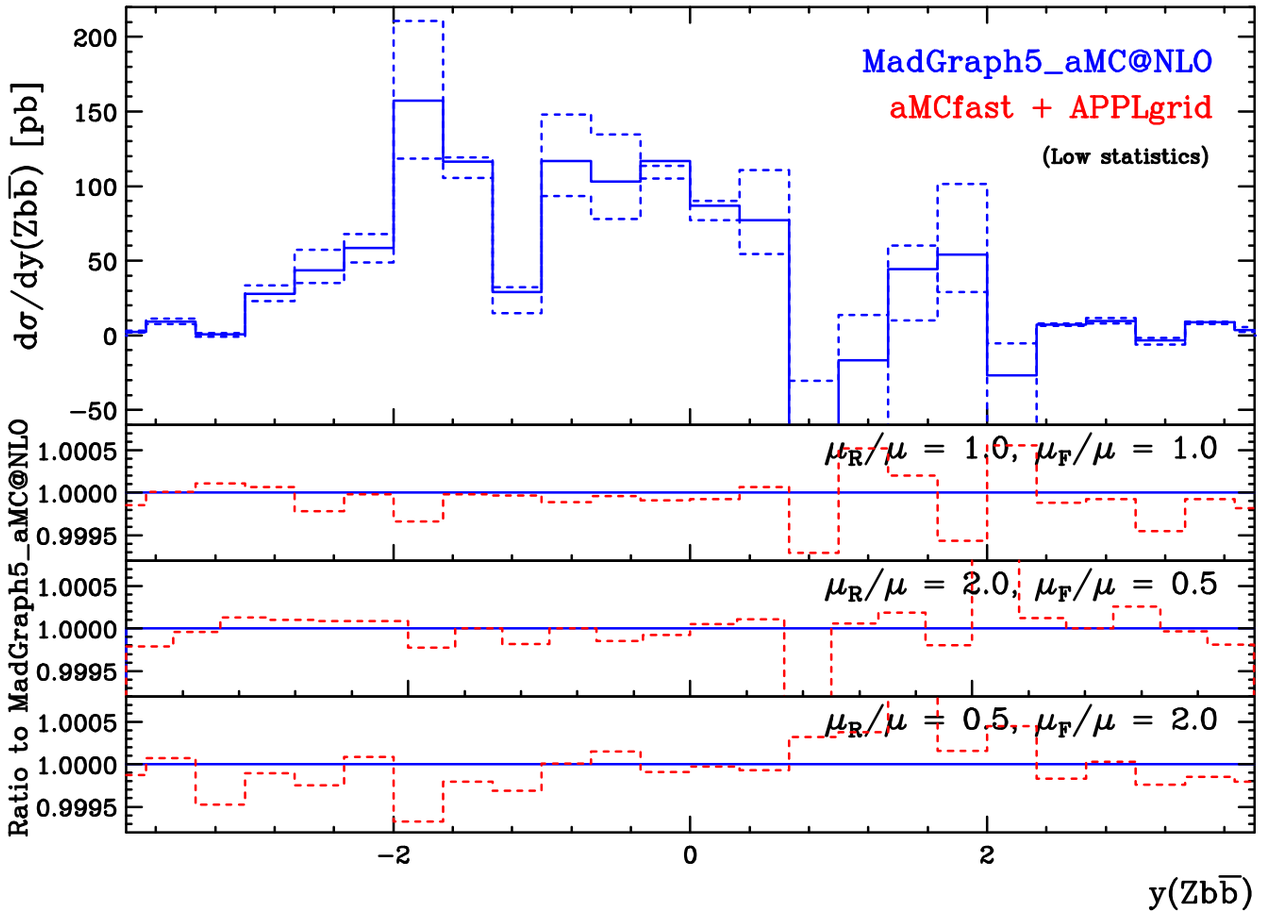
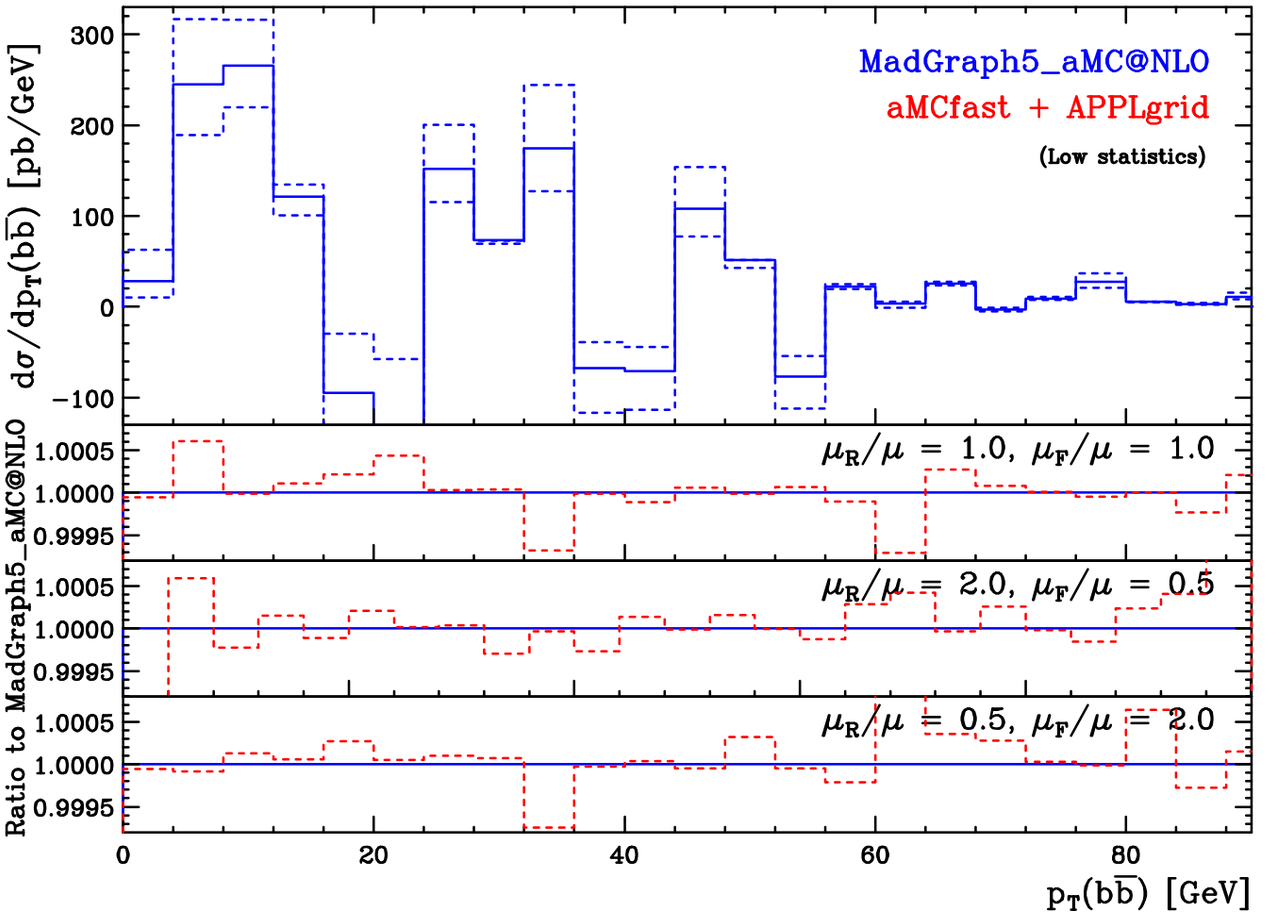
Validation: Z + bb Production

- This is just an example of complicated process.
- We looked at the following observables:
  - the rapidity distribution of the *Zbb* system (left),
  - the transverse momentum distribution of the *Zbb* system (right).
- High statistics plots:





The aMCfast Interface

Validation: Z + bb Production

- This is just an example of complicated process.
- We looked at the following observables:
  - the rapidity distribution of the Zbb system (left),
  - the transverse momentum distribution of the Zbb system (right).
- Low statistics plots: