Statistical tools for impact studies at EIC

Rabah Abdul Khalek, 21/05/2020

2nd EIC YR Workshop at Pavia University

Collaborating with:
Valerio Bertone
Timothy Hobbs
Daniel Pitonyak
Alexei Produkin
Nobuo Sato

On benchmarking

https://jeffersonlab.github.io/txgrids/_build/html/bench.html

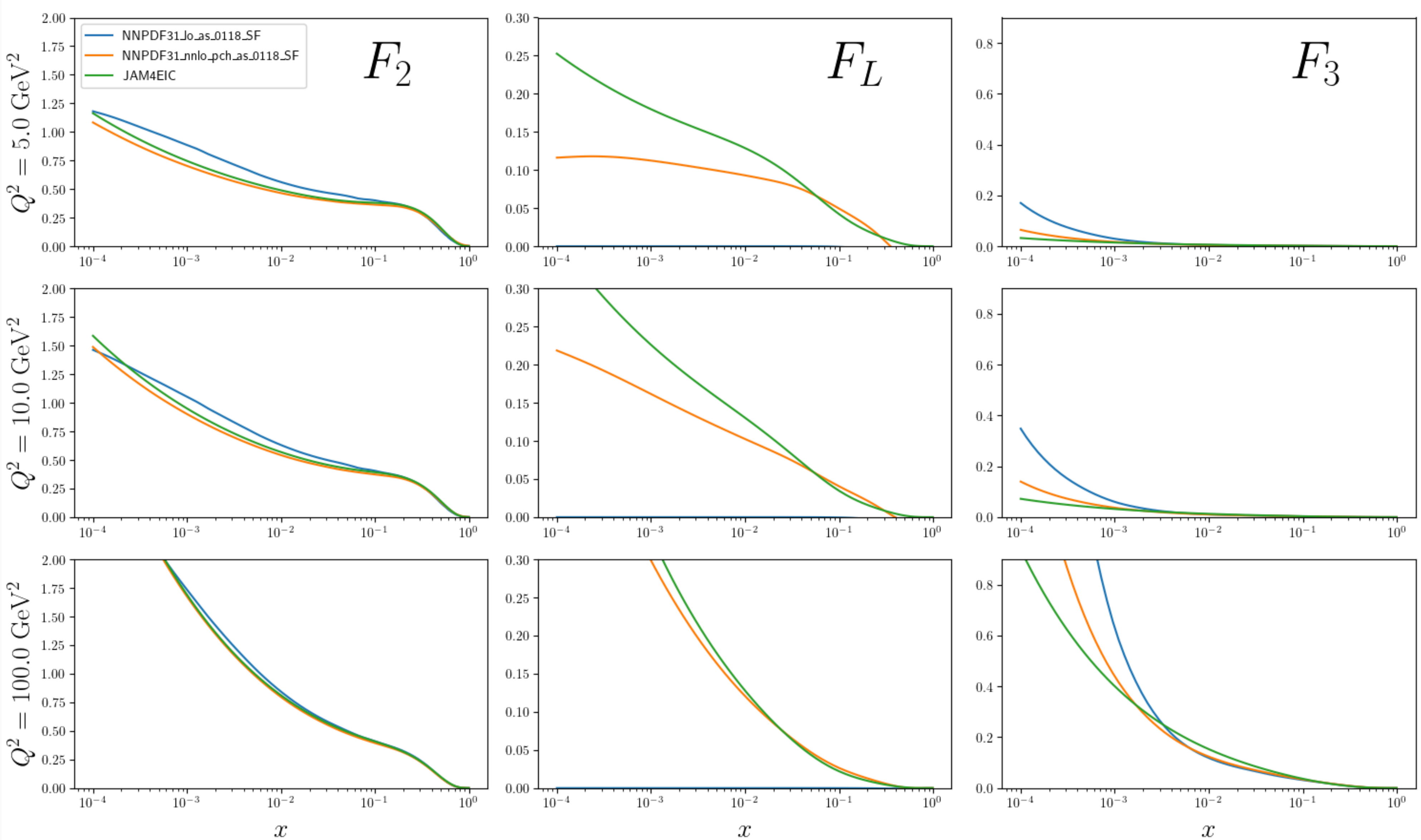
Benchmarks

Thanks to N. Sato

NC cross sections

name	values	theory	$\sqrt{s}$	kin. cuts
NNPDF31_lo_as_0118_SF	$9.1826 \times 10^{8} \pm 3.2447 \times 10^{3}$(fb)	LO	140.7GeV	$Q^2_{\rm min}$= 1.0 (GeV^2) $W^2_{\rm min}$= 10.0 (GeV^2)
NNPDF31_nnlo_pch_as_0118_SF	$7.8199 \times 10^{8} \pm 3.1779 \times 10^{3}$(fb)	NNLO	140.7GeV	$Q^2_{\rm min}$= 1.0 (GeV^2) $W^2_{\rm min}$= 10.0 (GeV^2)
JAM4EIC	$8.0504 \times 10^{8} \pm 3.2625 \times 10^{3}$(fb)	NLO	140.7GeV	$Q^2_{\rm min}$= 1.0 (GeV^2) $W^2_{\rm min}$= 10.0 (GeV^2)

Structure functions



Motivation

- **Comparing the impact of different EIC detector configurations on PDFs from vertex-level generated events**
- **Giving experimentalists a (hopefully reliable) tool to efficiently gauge the impact of different detector configurations.**
- **Providing an alternative for fitting PDFs with many configuration settings**

Work-flow

1. Consider a feature to study
e.g Impact on strangeness of
unpolarised PDFs

2. Define two extreme cases
Construct two PDF sets with two
Extreme cases of strange (max and min)
According to phenomenology

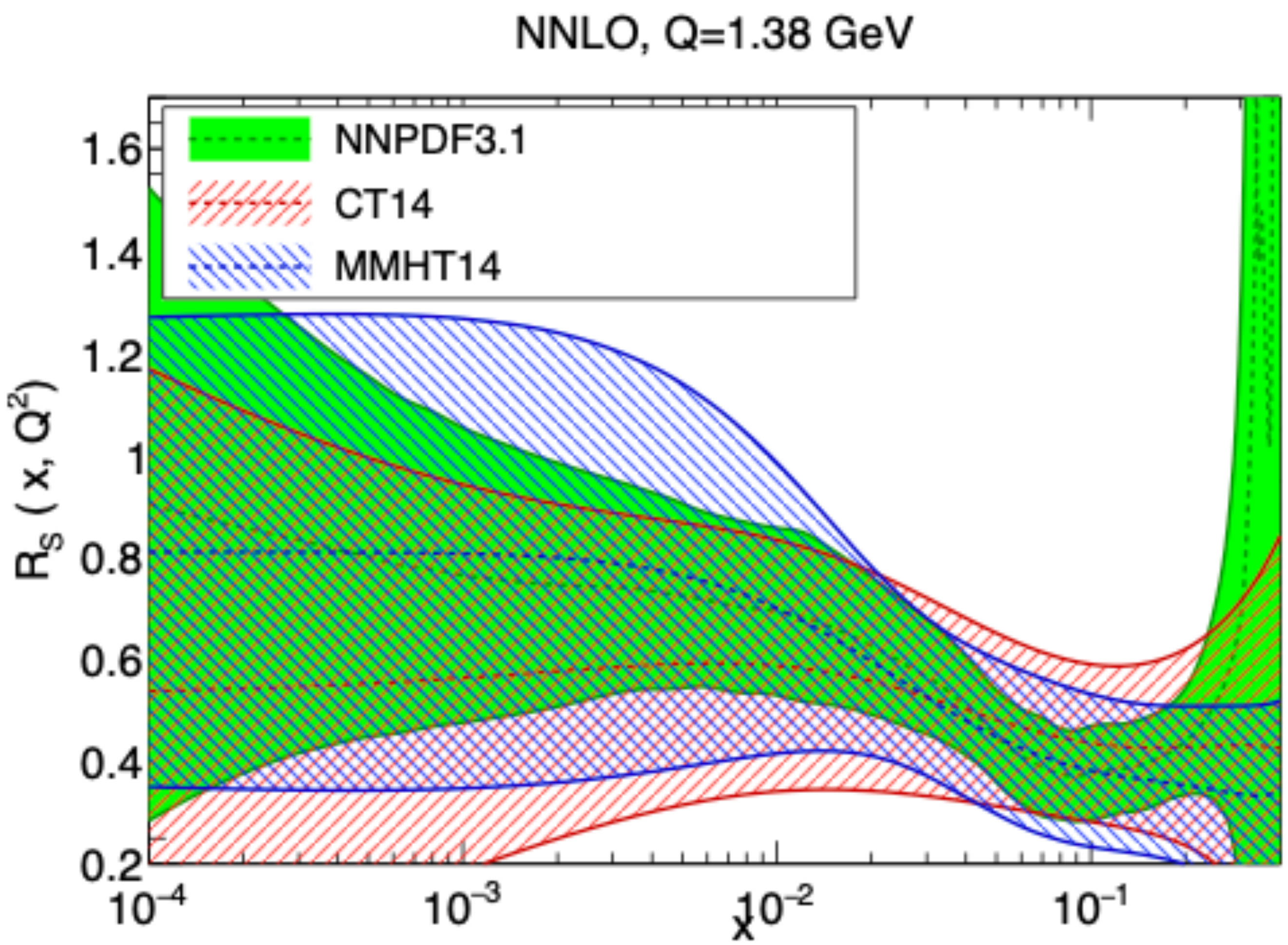
3. Compute SF from the extreme PDFs
in LHAPDF grids format

4. Generate 2 MC samples
reduced cross sections from max and min cases

5. Perform statistical test on the samples
to gauge the sigma-level significance of discrimination in bin of (x,Q2)

Work-flow

1. Consider a feature to study
e.g Impact on strangeness of unpolarised PDFs

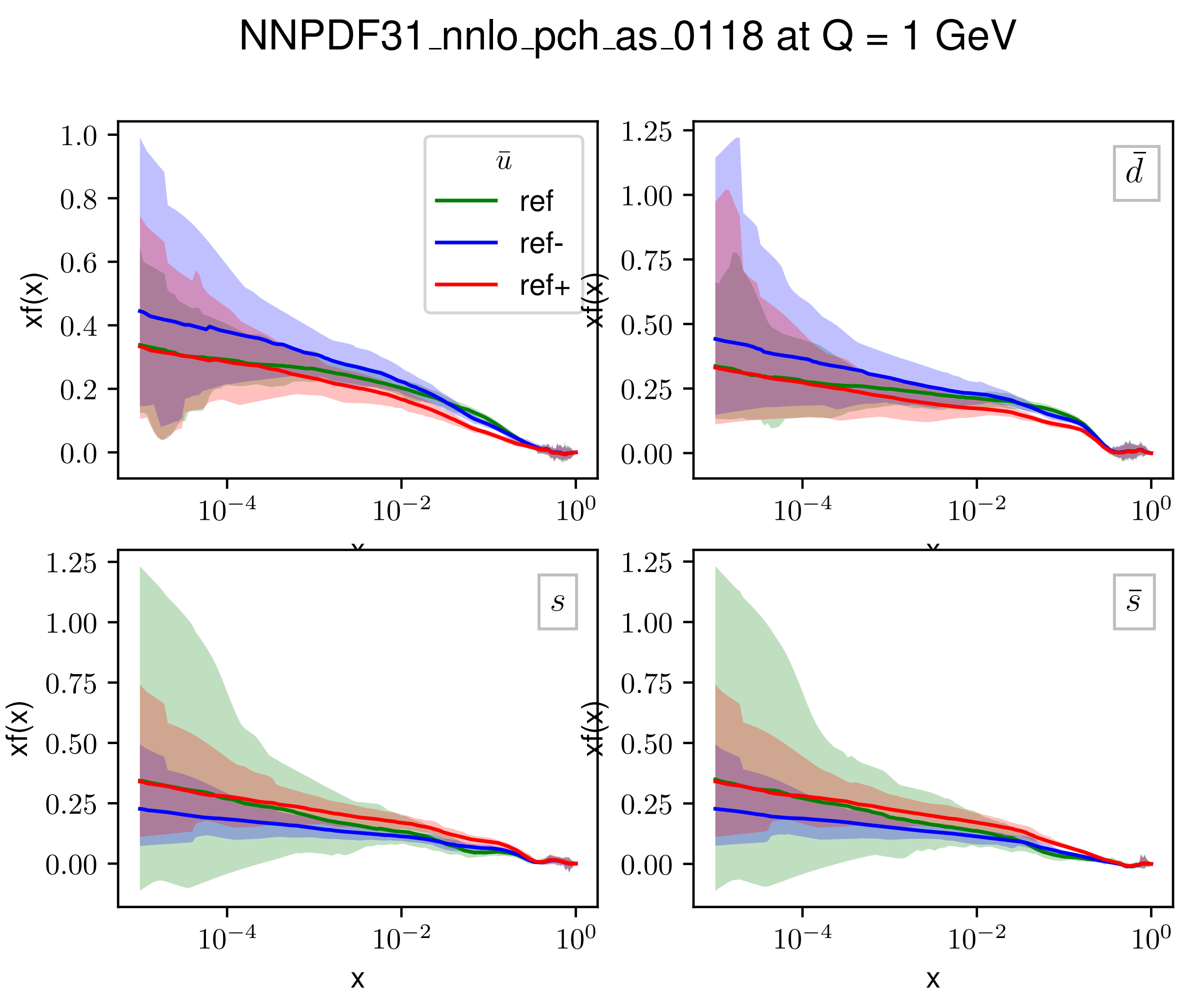


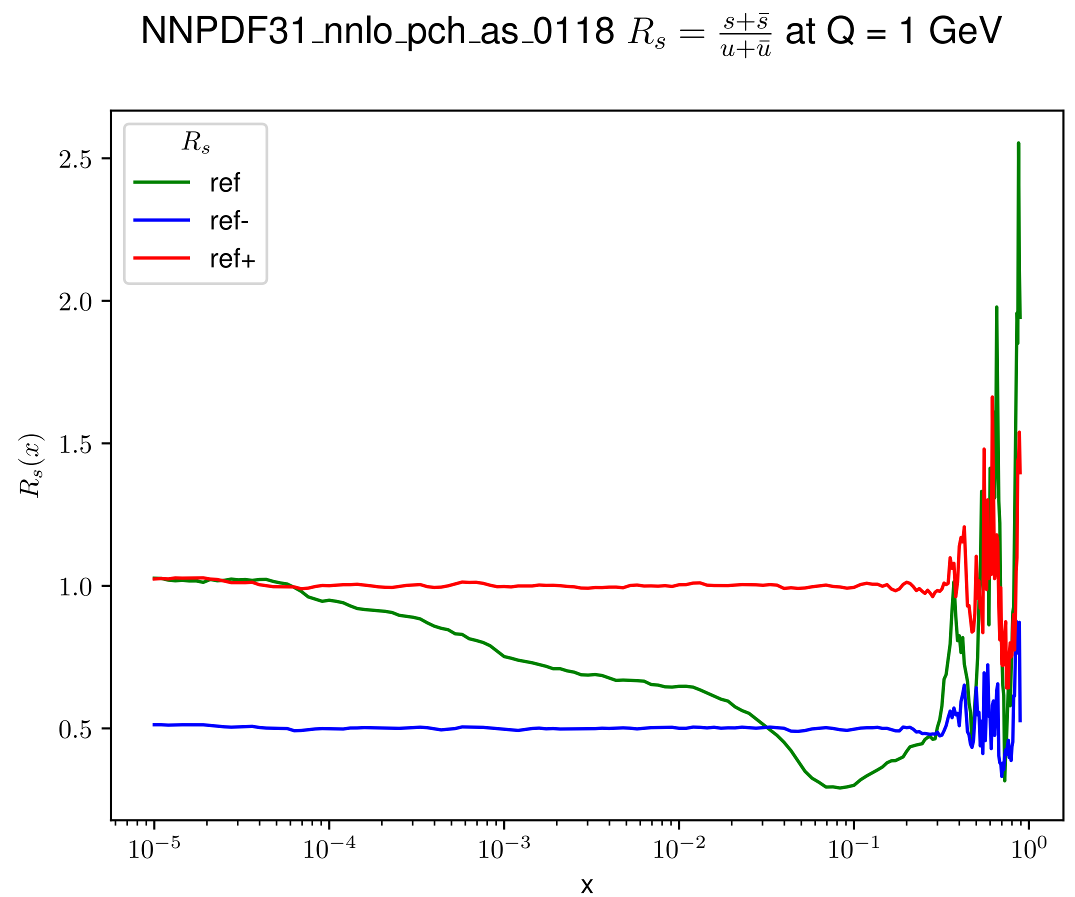
$$R_s = \frac{s + \bar{s}}{\bar{u} + \bar{d}} \simeq [0.5, 1] \, ,$$

Work-flow

2. Define two extreme cases
Construct two PDF sets with two
Extreme cases of strange (max and min)
According to phenomenology

Thanks to V. Bertone





$$s(x) \to s(x) + \kappa(x)\,, \quad \bar{s}(x) \to \bar{s}(x) + \kappa(x)\,,$$

and:

$$\bar{u}(x) \to \bar{u}(x) - \kappa(x)\,, \quad \bar{d}(x) \to \bar{d}(x) - \kappa(x)\,.$$

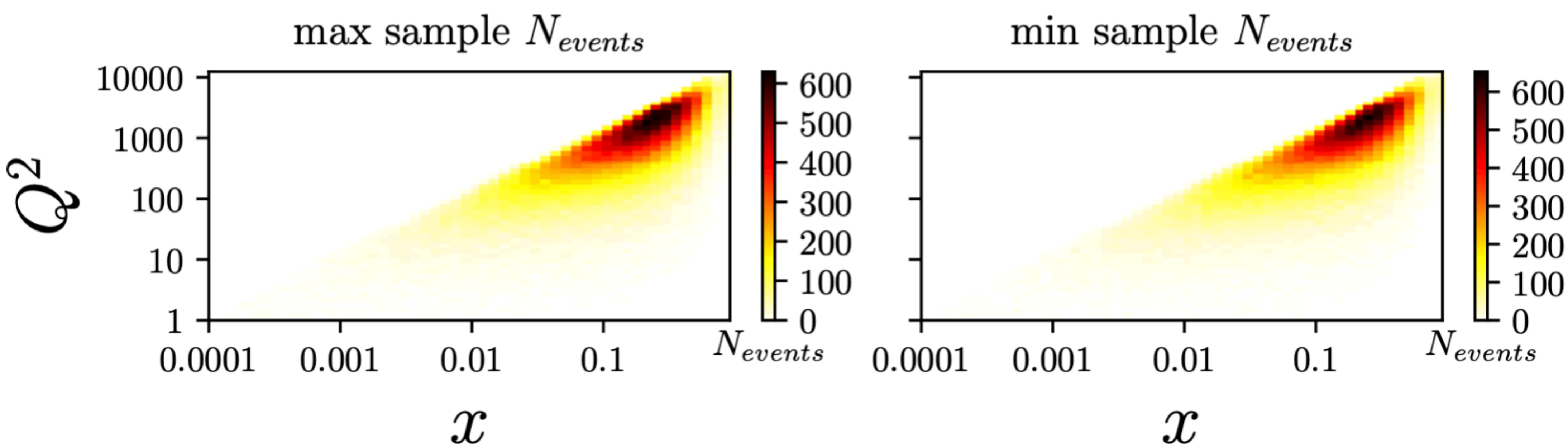
Such that

$$0.5 \leq R_s \leq 1$$

Work-flow

3. Compute SF from the extreme PDFs
in LHAPDF grids format

4. Generate 2 MC samples
reduced cross sections from
max and min cases



$N_x = 50,\ N_{Q^2} = 50$
$N_{events} = 100k$
$\sqrt{s} = 140.70$ GeV
$W^2 > 10\ ,\ Q^2 > 1$

Work-flow

5. Perform statistical test on the samples
to gauge the sigma-level significance of discrimination in bin of (x,Q2)

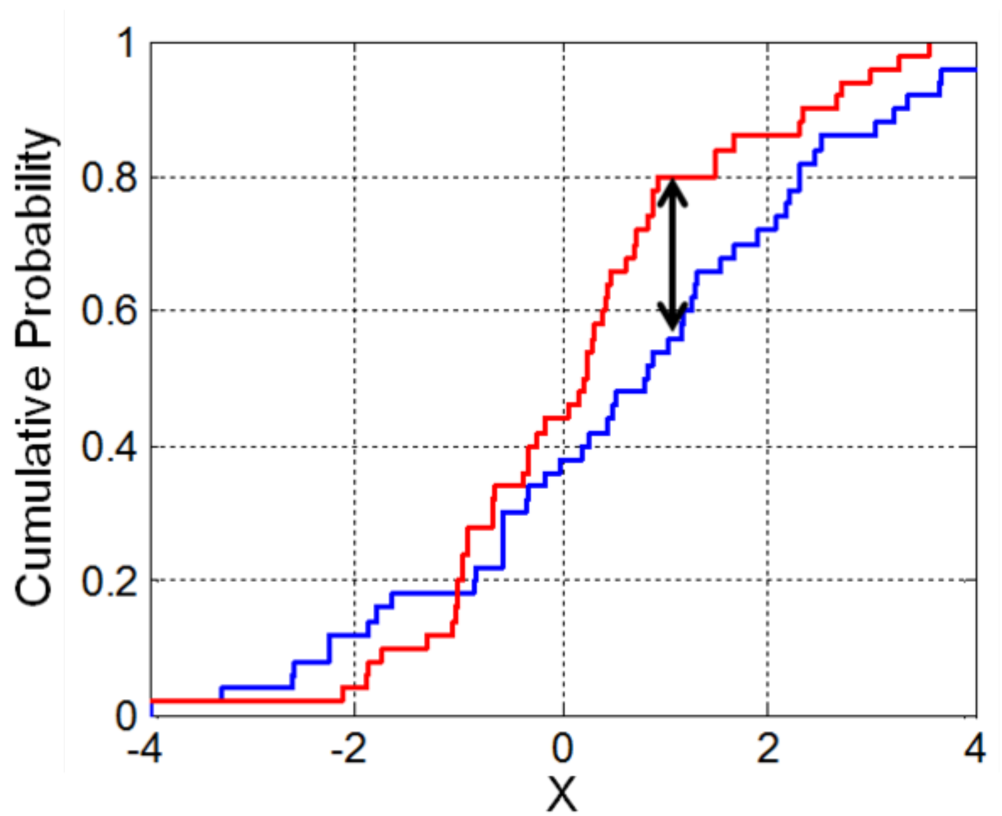
For example: **Kolmogorov-Smirnov Test**

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|$$

For large samples, the null hypothesis is rejected at level α if:

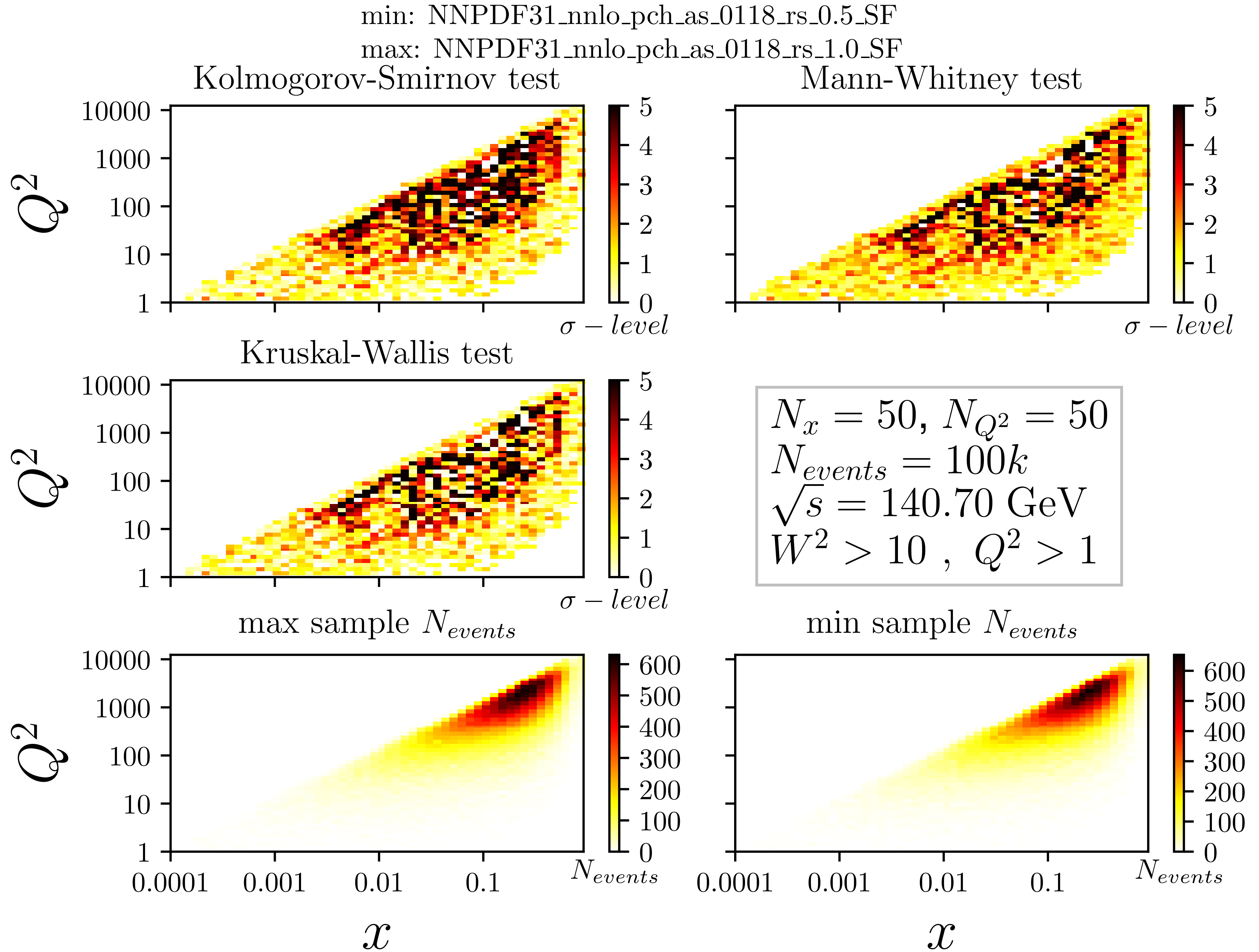
$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n \cdot m}}.$$

Where α (p-value) is the probability of detecting a difference under the assumptions of the null hypothesis (that the two samples are drawn from the same distribution).



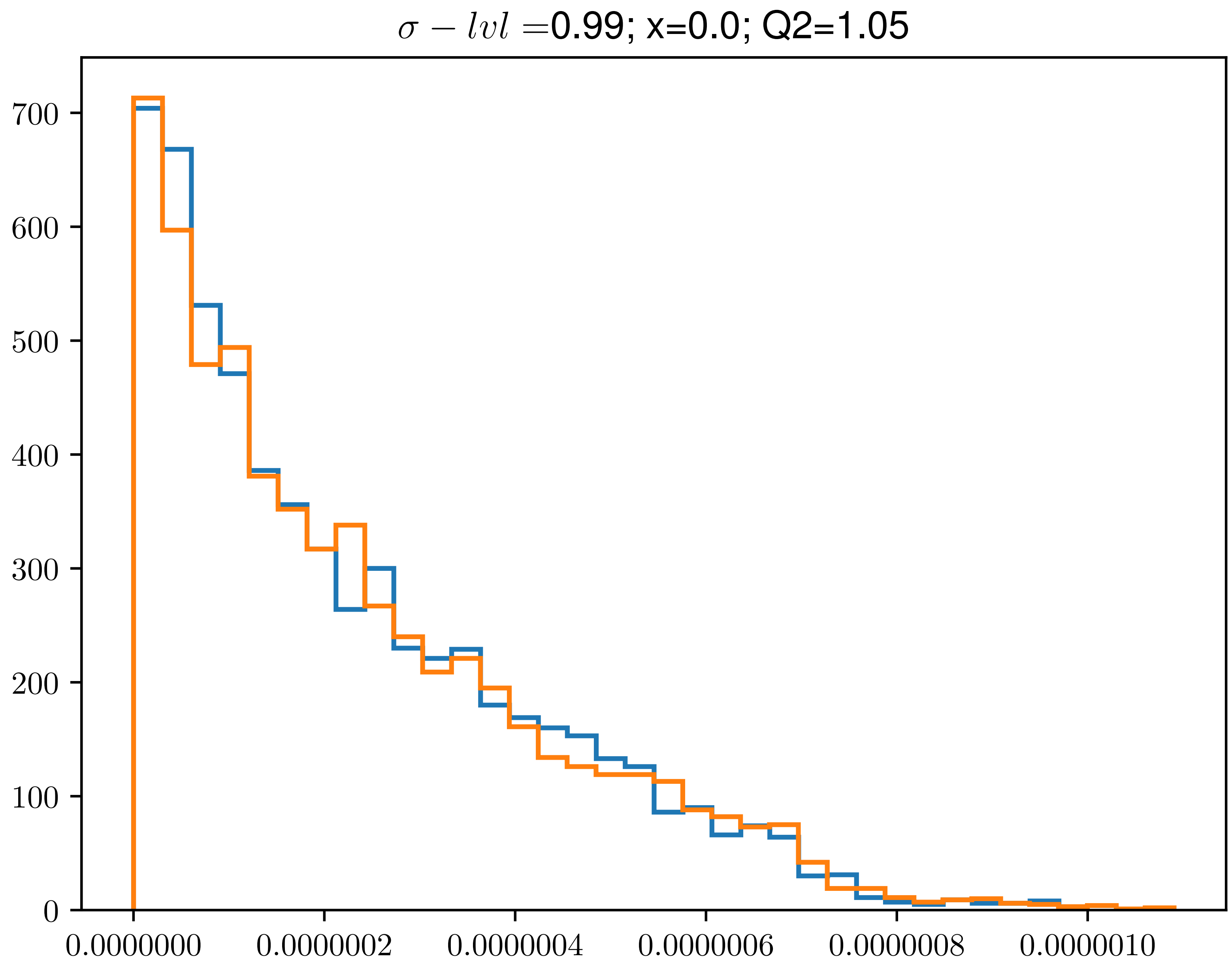
Work-flow

5. Perform statistical test on the samples
to gauge the sigma-level significance of discrimination in bin of (x,Q2)



Sensitivity of KS

2 samples from SAME underlying law,
Generated from DIFFERENT seeds



Summary

- **We need to assess how sensitive our tests are to statistical fluctuations (work-in-progress)**

- **We still need to perform a background events substraction, and add detector effects to our cross sections:**

  $$\sigma(x, Q^2 | H) = [\sigma_S(x, Q^2 | H) - \sigma_B(x, Q^2 | H)]\ \pi(x, Q^2 | H, \mathscr{L}, RC, detector)$$

  1. It is at this level that our exercise would start to be useful, comparing the change of discrimination before and after adding detector effects.

  2. At this point, we'll be also able to compare also between different configuration settings.

- **Currently working on implementing the t-test as performed in https://arxiv.org/pdf/1501.03156.pdf**